Income and Education of the States of the United States: 1840-2000

Scott Baier, Sean E. Mulholland, Robert Tamura and Chad Turner*

October, 2005

Abstract

This article introduces original annual average years of schooling measures for each state from 1840 to 2000. Our methodology results in state estimates similar to those reported in the United States Census from 2000 back to 1940 and national, turn of the century estimates strikingly close to those presented by Schultz (1961) and Fishlow (1966). These data show that the New England, Middle Atlantic, Pacific, East North Central and West North Central regions have been educational leaders over the entire time period. In contrast, the South Atlantic, East South Central and West South Central regions have been educational laggards. The Mountain region behaves differently than either of the aforementioned groups. To further determine the validity of our state schooling estimates, we first combine original data on real state per worker output with existing data to provide a more comprehensive series of real state output per worker from 1840 to 2000. We then estimate aggregate Mincerian earnings regressions and discover that the return to a year of schooling for the average individual in a state ranges from 10 percent to 13 percent. This range is robust to various time periods, various estimation methods and various assumptions about the endogeneity of schooling. These estimates are in line with the body of evidence from the labor literature and thus further support the validity of our constructed years of schooling measures.

Previous research on state-level output and school enrollment has focused on the relative rank ordering of states (Mulligan and Sala-i-Martin 1997, Goldin and Katz 2003). However, in order for economists, educators, and policy makers to empirically analyze historic educational choices, state-level educational stock measures are needed. Organizing these quantitative data over a long time period lends itself easily to many types of analyses, such as changes in state schooling policies, income growth, and changes in total factor productivity. To this end, this paper makes two contributions: (1) it introduces original annual measures of years of schooling and average years of experience in the labor force for each of the states of the United States, generally from 1840 through 2000, and (2) it constructs original real state per worker output estimates for 1850, 1860, 1870, 1890 and 1910, and combines them with existing data for 1840, 1880, 1900 and 1920 and 1929 through 2000. Furthermore, it captures the educational choices made by individuals (aggregated to the state level) over much of the history of the United States. We use data from the decennial censuses of the United States, Richard Easterlin's work on state income, *Historical Statistics of the United States:*

*Baier and Tamura are affiliated with Clemson University and the Atlanta Federal Reserve Bank. Mulholland is at Moravian College and Turner is affiliated with Nicholls State University. All views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. We thank the workshop participants at Clemson University, Texas A & M, the University of Kentucky, SUNY Buffalo, the University of Virginia, the University of South Carolina, and the 2004 Midwest Macroeconomic Meetings for helpful comments. We benefited from suggestions by Kevin Murphy and Casey Mulligan. All errors are the responsibility of the authors.

Colonial Times to 1970 as well as information contained in annual *Statistical Abstracts of the United States* to produce these estimates.[1] We admit that the accuracy of these enrollment data have been questioned by previous analyses.[2] We are comforted however, by Fishlow's 1966 conclusion that "for most purposes [the Census statistics] seem to suffice in their present form."

Though our paper presents original measures of schooling at the state level, we are not the first to attempt to measure historical levels of human capital in the economy. Theodore Schultz (1961), following the earlier work of Long (1958), used 1940 Census information (the first Census to report years of schooling) on schooling by age cohort to backward project the national stock of education for previous census years back to 1900. For 1900, Schultz estimated that the average years of schooling was 4.14 years.[3] Albert Fishlow (1966) used Census data before 1940 to calculate the national stock of education for both 1860 and 1900. For 1900, he determined the national average years of schooling was 4.96, and for 1860, 2.06. These estimates are within 2 percent of our labor force weighted national estimates of 4.91 for 1900 and 2.10 for 1860.

To further check the validity of our state-level estimates we estimate the relationship between the level of state education and income. We find that estimated return to a year of schooling for the average individual in a state ranges from 10 percent to 13 percent. This range is robust to various time periods and various estimation methods. Although not necessarily producing similar results, we view this work as complementary to the work of Mulligan and Sala-i-Martin (1997, 2000).[4] We also document the long-term enrollment trends in primary, secondary, and tertiary schooling as well as the patterns of income growth across census regions. We show both within region and across region convergence.

The remainder of the paper is organized as follows: The next section provides the accounting framework for calculating average years of schooling by state. We present in graphical and tabular form the results of these calculations by census region. Section III presents our measures of state output per worker. Section IV contains our estimates returns to schooling and returns to potential experience estimates. We find that OLS estimates are quite robust to alternative specifications, and that a year of schooling returns about 10 percent to an individual in additional productivity. Section V concludes and describes broader implications and future work.

II. EDUCATION IN THE STATES

We use a perpetual inventory method, employed by Barro and Lee (1993) and Baier, Dwyer and Tamura (2005), to construct average years of schooling in the labor force for each state. Because we are interested in the relationship between human capital and output per worker, it is more appropriate to calculate the average years of schooling in the labor force instead of the average years

[1] Data covering a large number of states (28) is first available in 1840. Before 1840, we are aware of enrollment data for nine states: Maine, New Hampshire, Connecticut, Rhode Island, Massachusetts, New York, South Carolina, Virginia, and Kentucky. For a greater discussion of schooling in the first half of the nineteenth century see Fishlow (1966).

[2] The American Statistical Association offered an official critique of the 1840 Census and found errors in the collection of common school data (Senate Document No. 5, 28th Congress, 2nd Session).

[3] Schultz results are reported in his Table 7 on page 68.

[4] Mulligan and Sala-i-Martin (1997,2000) construct two different state level human capital measures for the census years 1940-1990, inclusive. Our years of schooling human capital measure is highly correlated with theirs, averaging approximately 0.8. See Appendix D for more detail.

of schooling of all state residents.[5][6][7] Enrollment data from Censuses, Digests of Education Statistics and Statistical Abstracts of the United States present the number of students enrolled in one of three educational categories: primary, secondary, and college.[8] In order to calculate the average years of schooling in the work force, our methodology must account for:

1. the number of school age children;
2. the number of new labor force participants, I_t^i , and their education level;
3. the departure rate of workers from the workforce, δ_t^i ;
4. the interstate migration of students post education;
5. and the impact of foreign-educated immigrants.

We assume that there are four categories of workers: those with no schooling (none); those exposed to primary schooling and no more (primary); those exposed to secondary schooling and no more (secondary);and those with exposure to higher education (college). To calculate average years of schooling, we assignthe average years of schooling attained for each of these categories, with the uneducated group getting zero years of schooling. Suppressing the state subscript, H_t^i is the number of workers in the labor force in year t in education category i. The perpetual inventory method produces the following law of motion:

$$H_{t+1}^i = H_t^i\left(1-\delta_t^i\right) + I_t^i,\ i = \text{none, primary, secondary, college} \tag{1}$$

where δ_t^i is the departure rate from the labor force between year t and t+1 and I_t^i is the gross flow of new workers into the labor force from education category i.

In order to get estimates of the flows into each education category, we use the following information:

$$I_t^{\text{college}} = \frac{r_t^{\text{college}} lfpr^{\text{college}} \ell[18-24]_t}{7}\Theta \tag{2}$$

$$I_t^{\text{secondary}} = \frac{\left(r_t^{\text{secondary}} - r_t^{\text{college}}\Theta\right) lfpr^{\text{secondary}} \ell[14-17]_t}{4} \tag{3}$$

$$I_t^{\text{primary}} = \frac{\left(r_t^{\text{primary}} - r_t^{\text{secondary}}\right) lfpr^{\text{primary}} \ell[5-13]_t}{9} \tag{4}$$

$$I_t^{\text{none}} = \frac{\left(1 - r_t^{\text{primary}}\right) lfpr^{\text{primary}} \ell[5-13]_t}{9} \tag{5}$$

[5] Additional details on the derivation and the data sources are furnished in Appendix B.

[6] Ideally we would use information to produce average years of schooling for men and women separately in the labor force, however, enrollment information by sex is not consistently available. However Series H 433-441, page 370 of *Historical Statistics of the United States: Colonial Times to 1970*, indicates that there was little difference in enrollment rates of men and women:

sex	1850	1860	1870	1880
male	49.6	52.6	49.8	59.2
female	44.8	48.5	46.9	56.5

. From 1890 onward differences in enrollment rates were less than one percentage point. We acknowledge that our calculations implicitly assume the labor force participation rate is common across men and women.

[7] We are unable to account for changes in the labor force participation rates by educational category because we do not have data on labor force participation by education category prior to 1960.

[8] See Appendix A for additional information on enrollment rates by educational category.

where in year t r_t^i is the enrollment rate in education category i, $lfpr_t^i$ is the labor force participation rate for each educational category, and $\ell[i-j]_t$ is the population in age category $[i-j]$, inclusive.[9] We assume that population in each age category is uniformly distributed and that higher education enrollment rates are constant across ages within an education category. The constant Θ is an adjustment for the fact that, because there is high rate of attrition in the early part of higher education, assuming a uniform enrollment rate across ages will understate the true inflow into the higher educational category.[10]

Although values of are δ_i^t are not directly available, we are able to calculate three different departure rates: one for college workers, $\delta_t^{\text{colllege}}$,one for secondary workers, $\delta^{\text{secondary}}$, and one for all other workers, $\delta_t^{\text{primary}}$ using the following solution:[11][12]

First, we assume that workers with some college exposure do not disappear at a calculated rate, but only after 45 years of employment. Thus for college exposed workers, the law of motion becomes:

$$H_{t+1}^{\text{college}} = H_t^{\text{college}} - I_{t-45}^{\text{college}} + I_t^{\text{college}} \tag{6}$$

We let h_{t+1}^i represent the share of the labor force exposed to educational category i. Dividing the law of motion equation by the labor force in period t+1 for the higher educational category provides:

$$h_{t+1}^{\text{college}} = h_t^{\text{college}} \frac{L_t}{L_{t+1}} - \frac{I_{t-45}^{\text{college}}}{L_{t+1}} + \frac{I_t^{\text{college}}}{L_{t+1}} \tag{7}$$

For the very early years, $I_{t-45}^{\text{college}}$ is approximated using the first observed measure of higher education enrollment rates in t.[13] Once enough years have past, we use our own calculations for $I_{t-45}^{\text{college}}$.

Second, for workers exposed to secondary schooling, we choose $\delta^{\text{secondary}}$ by utilizing decennial census data on the share of workers exposed to secondary education from 1940-2000. Given the structure of our laws of motion and inflow calculations, we choose the value of $\delta^{\text{secondary}}$ that results

[9] For labor force participation rates by educational attainment we used data from the 1940-2000 censuses. We use .91, .82 and .60 for $lfpr^{\text{college}}$, $lfpr^{\text{secondary}}$, and $lfpr^i$, i =primary, none. We used these labor force participation rates for the entire 1840-2000 period. While it may seem strange to use a constant labor force participation rate, in 1840 the labor force participation rate for 14-65 year old individuals was 51 percent and in 1900 the labor force participation rate for this same category was 57 percent. Since the majority of our labor force is either without education or with only primary education in this period, we feel that holding labor force participation rates constant over time across education categories is reasonable.

[10] Since our calculations of the inflow to all categories are equal to the total enrollment across all ages in the category divided by the total population across all ages in the category, they implicitly assume the enrollment rate is constant across ages within each education category. To the extent that this assumption is erroneous, the true inflow in to the category will be understated. While this assumption is implicit in our calculations for inflows into all educational categories, it is most problematic where there is a high attrition rate between age. As attrition rates are between the first and second years of higher education, \we multiply the measured inflow into the higher education category by a factor denoted Θ . We allow Θ to vary across states, but assume it is time invariant. For additional details and for the values of Θ for each state, see Appendix B.

[11] We deliberately omit the time subscript on the departure rate for the secondary education category. Our reasoning is discussed in greater detail later in this section. Also, we use a common departure rate for the primary and none educational categories, which we denote $\delta_t^{\text{primary}}$.

[12] The creation of a separate departure rate for college workers is a motivated by the fact that a common departure rate for all education categories produces a share of workers exposed to higher education significantly below the value reported in the census in 2000. After making this adjustment, we further find that a departure rate common to the remaining classes (secondary, elementary and none) produces some states where the share of workers exposed to elementary schooling is less than zero. As a result, we also allow for a separate departure rate for those workers exposed to secondary schooling.

[13] This is not much of an issue in the early years because higher education enrollments are near zero. Further details are discussed in Appendix B.

in the closest match of the evolution of $h_t^{\text{secondary}}$ to that of the corresponding census data from 1940–2000.[14] We note that unlike the departure rates for other educational categories, $\delta^{\text{secondary}}$ is time invariant. For values, see Appendix B. The result is:

$$h_{t+1}^{\text{secondary}} = h_t^{\text{secondary}} \frac{L_t}{L_{t+1}} \left(1 - \delta^{\text{secondary}}\right) + I_t^{\text{secondary}} \tag{8}$$

Third, though we are unable to calculate the departure rate for the remaining educational classes directly, we can isolate $\delta_t^{\text{primary}}$ using the following identity:

$$L_{t+1} = H_{t+1}^{\text{college}} + H_{t+1}^{\text{secondary}} + H_{t+1}^{\text{primary}} + H_{t+1}^{\text{none}} \tag{9}$$

Dividing through by L_{t+1} and then substituting using (1) for the primary and none categories yields:

$$1 = \frac{H_{t+1}^{\text{college}}}{L_{t+1}} + \frac{H_{t+1}^{\text{secondary}}}{L_{t+1}} + \frac{H_t^{\text{primary}}\left(1 - \delta_t^{\text{primary}}\right) + I_t^{\text{primary}}}{L_{t+1}} + \frac{H_t^{\text{none}}\left(1 - \delta_t^{\text{primary}}\right) + I_t^{\text{none}}}{L_{t+1}} \tag{10}$$

$$1 - h_{t+1}^{\text{college}} - h_{t+1}^{\text{secondary}} = \left(h_t^{\text{primary}} + h_t^{\text{none}}\right) \frac{L_t}{L_{t+1}} \left(1 - \delta_t^{\text{primary}}\right) + \frac{I_t^{\text{primary}} + I_t^{\text{none}}}{L_{t+1}} \tag{11}$$

We then isolate our estimate of the $\frac{L_t}{L_{t+1}}\left(1 - \delta_t^{\text{primary}}\right)$ term:

$$\frac{L_t}{L_{t+1}} \left(1 - \delta_t^{\text{primary}}\right) = \frac{1 - h_{t+1}^{\text{college}} - h_{t+1}^{\text{secondary}} - \left(\frac{I_t^{\text{primary}} + I_t^{\text{none}}}{L_{t+1}}\right)}{\left(h_t^{\text{primary}} + h_t^{\text{none}}\right)}. \tag{12}$$

Thus for the share of labor force with primary schooling exposure we return to (1), divide by L_{t+1} and produce:

$$h_{t+1}^{\text{primary}} = h_t^{\text{primary}} \frac{L_t}{L_{t+1}} \left(1 - \delta_t^{\text{primary}}\right) + \frac{I_t^{\text{primary}}}{L_{t+1}}, \tag{13}$$

and then use the following adding up restriction for the share of the labor force with no educational exposure:

$$h_{t+1}^{\text{none}} = 1 - h_{t+1}^{\text{college}} - h_{t+1}^{\text{secondary}} - h_{t+1}^{\text{primary}}.\text{[15]} \tag{14}$$

We use information from the 1940-2000 Censuses to get estimates for expected number of years of schooling completed, conditional on being in each education category for each state. These expected years of schooling by category are represented by $yrs_{it}^{\text{college}}$, $yrs_{it}^{\text{secondary}}$ and $yrs_{it}^{\text{primary}}$. For the intervening years we log linearly interpolate. Initial values for $yrs_{it}^{\text{college}}$, $yrs_{it}^{\text{secondary}}$, and $yrs_{it}^{\text{primary}}$ are set at 4, 10 and 14 for primary, secondary and higher education, respectively, in the year that data becomes available for each state.[16] We then log linearly interpolate from these initial values to the 1940 value. Thus for state i we calculate average years of schooling in the labor force as:

$$\widehat{E}_{it} = h_{it}^{\text{college}} yrs_{it}^{\text{college}} + h_{it}^{\text{secondary}} yrs_{it}^{\text{secondary}} + h_{it}^{\text{primary}} yrs_{it}^{\text{primary}} \tag{15}$$

[14] We simply utilize our methodology for each value of $\delta^{\text{secondary}}$ across a grid in increments of 0.0001. We select the value of $\delta^{\text{secondary}}$ for each state that most closely matches our calculated data to the census data, giving equal weight to each decennial observation.

[15] There are occasions when $h_t^{none} < 0$. In these instances, we set $h_t^{none} = 0$ and renormalize the shares to sum to 1. These instances are rare and small in absolute value.

[16] See Appendix B for more details on the evolution of average years of schooling.

To account for interstate migration, we adjust our years of schooling measure by residents state of birth reported in the 1850 through 2000 Censuses. We assume that all education is undertaken in an individual's state of birth and that all current migrants are educationally representative of their birth state. Due to data limitations, we can not allow for selective migration. Let $\widehat{E}_{jt}$ be the years of schooling at time t for those born in state j. Our estimate of years of schooling in state i therefore is:

$$E_{it} = \sum_{j=1}^{52} S_{ijt} \widehat{E}_{jt} \tag{16}$$

where S_{ijt} is the share of state i residents in year t that were born and educated in state j.[17] There are 52 categories: 50 states, the District of Columbia, and the foreign born. For foreign born we assume that the individuals come from the k^{th} percentile of the primary, secondary and higher education distributions. We use the information from each of the 1940-2000 Censuses to determine the best fitting k^{th} percentile for each state and census year in order to match the state's average years of schooling. For years prior to 1940 we assume that foreign born workers have the average $\overline{k}^{th}$ percentile, where the average is for the 1940-2000 period, and is state specific.[18]

To illustrate our years of schooling measure, the next four figures display the average years of schooling in the labor force by census region. [19][20][21] While initial conditions certainly come into play in the first few years, within 20 years, the initial conditions have little impact. Thus New England, the Middle Atlantic and Pacific regions were clearly education leaders in the US. All three regions remain above the average years of schooling in the US throughout the entire 1840 to 2000 period. Figure 4 indicates that the East North Central and, by 1880, the West North Central were educational leaders as well. From 1880 to 2000 the labor forces of these five regions were better educated than the average person in the labor force in the US. In contrast, the South Atlantic, East South Central and West South Central regions were educational laggards. They start with less schooling than the average in the US and remain below average throughout the data. However by 2000, these three regions have closed the gap between themselves and the US. Figure 3 illustrates the different behavior of the Mountain region. Unlike the Pacific region which remained above the US average, the Mountain region initially lagged behind the US, and in fact lagged behind the southern states from roughly 1850 to 1870. However from 1920 to the present the Mountain region was either at or above the US average in schooling. These results are summarized in Table 1 below.

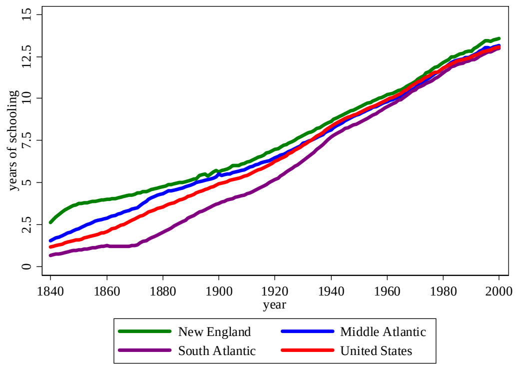
[17] In 2000, data availability is limited. The census reports the fraction of a state's residents that were born in that state, S_{ii}, and the fraction that is foreign born $S_{i,for}$. However, for those residents of a state who were not born in that state(S_{ij}, j≠i, j≠for), only the census region of birth is given. Conditioned on living in state i and being born in census region k, we assume the probability of having been born in state j is equal the population of state j divided by the population of region k. We make the necessary adjustment when the region of birth contains the state of residence. As data is not available for 1840, we assume the shares in 1840 are identical to the values in 1850. Also, data is not available for Alaska and Hawaii in 1940 and 1950. We assume these shares are identical to the values in 1960. For non-Census years, we linearly interpolate the shares born in state j residing in state in in year t.

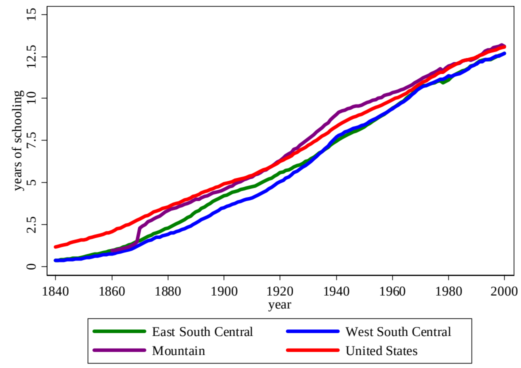
[18] Details are in Appendix B. For information on how well our measure matches the Census data from 1940 to 2000 see Appendix D.

[19] Rather than presenting graphs with 50 lines or tables with 50 rows, aggregation at the census region is a parsimonious manner to present the data. For empirical sections, we use state level data.

[20] For a listing of states within each region, see Appendix A.

[21] We do present information about maximum gaps between states in some of our tables.





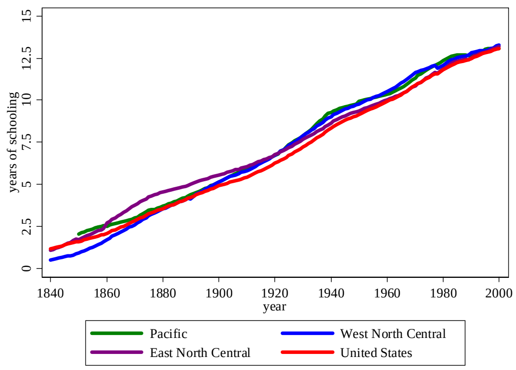


Table 1: Average Years of Schooling in the Labor Force

	1840	1860	1880	1900	1920	1940	1960	1980	2000
United States	1.14	2.10	3.56	4.91	6.24	8.34	9.94	11.8	13.1
New England	**2.61**	**3.98**	**4.77**	**5.64**	**6.95**	8.65	10.2	12.2	**13.6**
Middle Atlantic	1.54	2.88	4.36	5.50	6.47	8.15	9.83	11.8	13.2
South Atlantic	0.65	1.22	2.04	3.76	5.18	7.71	9.53	11.5	13.0
E. South Central	0.36	0.93	2.30	4.20	5.58	7.48	9.43	11.1	12.7
W. South Central	0.35	0.74	1.92	3.51	5.04	7.70	9.41	11.3	12.7
Mountain	-	0.93	3.33	4.58	6.32	9.06	10.4	11.9	13.1
Pacific	-	2.50	3.69	5.13	6.72	**9.27**	10.4	**12.4**	13.1
W. North Central	0.51	1.71	3.53	5.14	6.73	9.05	**10.5**	12.1	13.3
E. North Central	1.07	2.70	4.54	5.54	6.76	8.65	10.0	11.9	13.2
max. region gap	2.26	3.24	2.85	2.13	1.91	1.79	1.13	1.24	0.89
state max.	3.09	4.57	5.15	5.88	7.30	10.9	11.0	12.6	13.7
state min.	0.24	0.50	1.07	2.60	3.78	6.25	8.42	10.2	11.4

Table 1 contains the labor force weighted average years of schooling for each of the nine census regions and the average for the US for various years. For the US as a whole, the typical worker in 1940 had completed primary schooling and a third year of high school. By 1980 the typical worker was just about a high school graduate. In 2000 the labor forces in all regions have average schooling above 12 years. In 1880 the maximum gap between regions, 2.85 years, existed between the New England and West South Central regions. We pick 1880 as this is likely to be the first year in which initial conditions have no effect on the estimates. By 1900 the maximum gap between regions dropped to 2.13 years and existed between the East North Central and West South Central regions. From 1900 to 2000 the educational gap continues to narrow, reaching a nadir of 0.89 years in 2000.

In order to check the validity of our measures we compare our results to previous studies estimating the national level of educational stock. Schultz (1861), following the earlier work of Long (1958), used information in the1940 Census (the first to report years of schooling) on schooling by age cohort to backward project the national stock of education for previous census years back to 1900. For 1900 Schultz estimated that the average years of schooling was 4.14 years.[22] Our national estimate in 1900 of 4.91 is about .77 years greater than reported by Schultz. Looking at another national estimate, Fishlow (1966) used Census data before 1940 to calculate the national stock of education for both 1860 and 1900. For 1900 he determined the national average years of schooling was 4.96, just .05 years greater than our estimate of 4.91. Our 1900 national estimate lies between both Schultz and Fishlow. For 1860 Fishlow estimated a national educational stock of 2.06, just .04 years less than our estimate of 2.10. Our labor force weighted national estimates are within 2 percent of Fishlow's.

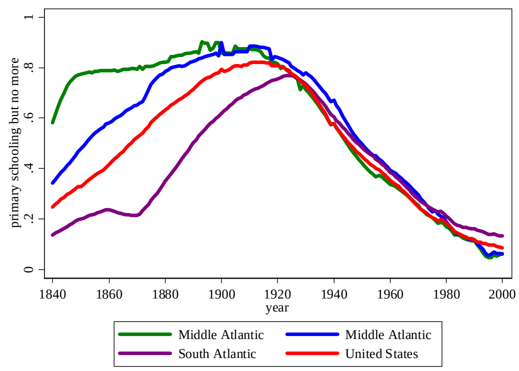
Table 2 presents the maximum gap between regions, in the row marked R, and states, in the row market S, at the decadal frequency, since 1890. Table 2 illustrates the clear convergence across regions, except for the very end. The evidence for the states is also compelling.

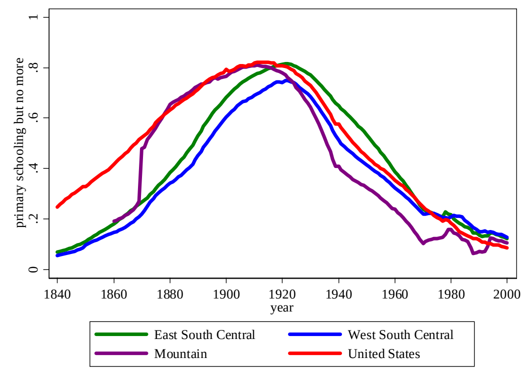
Table 2: Maximum Schooling Gaps between Regions and States

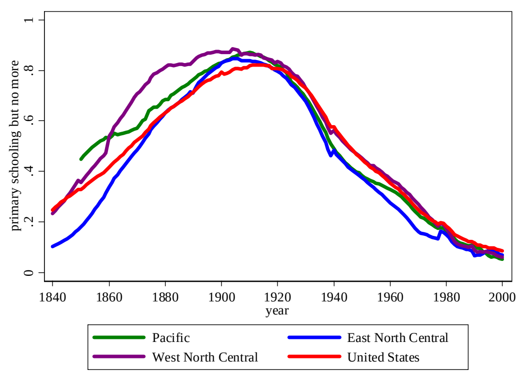
	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
R	2.53	2.13	2.10	1.91	1.76	1.79	1.62	1.13	1.09	1.24	0.74	0.89
S	3.57	3.28	3.40	3.52	4.11	4.60	3.72	2.59	1.97	2.41	2.04	2.31

The differences in average years of schooling between regions are the result of systematic differences in enrollment rates across regions. The New England, Middle Atlantic, Pacific, East North Central and, with a short lag, West North Central regions led the nation in educational attainment. These regions were the first to provide universal primary schooling, universal secondary schooling, and near universal higher education. In contrast, the South Atlantic, East South Central and West South Central regions lagged behind the country in each of these education categories. Finally the Mountain region is in between these two extreme groups. The next three figures illustrate the average fraction of the labor force that has been exposed to primary school, but no more.

[22] Schultz (1961) reports these results in Table 7 on page 68.



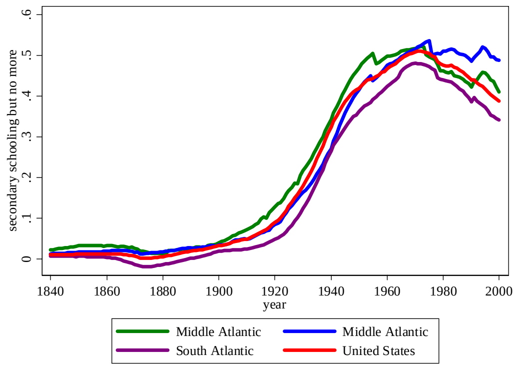


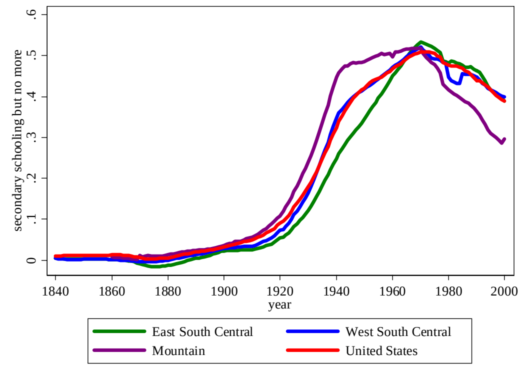


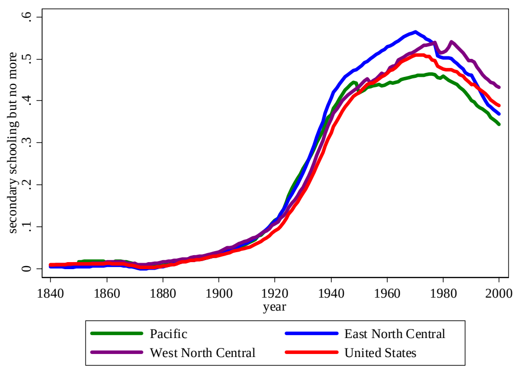
As the previous three figures illustrate, the South Atlantic, East South Central and West South Central regions display the lowest education exposure. From 1840 until about 1910 each of the three regions had a lower share of the labor force with elementary schooling exposure, and, as will be shown below, a lower fraction with secondary schooling exposure and higher education exposure as well. From 1920 to 2000 two of these regions have a greater share of the labor force with no more than an elementary schooling, and all three are higher after 1970.[23] The New England, Middle Atlantic, East North Central and to a slightly lesser degree the West North Central have a higher share of the labor force with elementary schooling exposure than the national average from 1840 (roughly 1870 for the West North Central) until the early part of the 20th century, between 1900 and 1920.

The next three figures illustrate the evidence of some secondary schooling exposure but no more.

[23] In early periods, regions with large shares of the labor force exposed to elementary educations are educational leaders. However, as these states are the first to have a significant fraction of their labor force exposed to secondary education, having a *smaller* fraction of the labor force exposed to elementary school only is evidence of educational leadership.

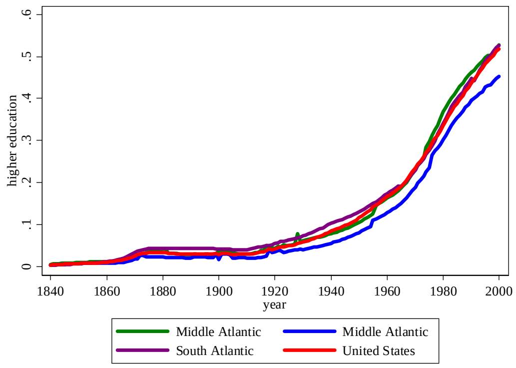


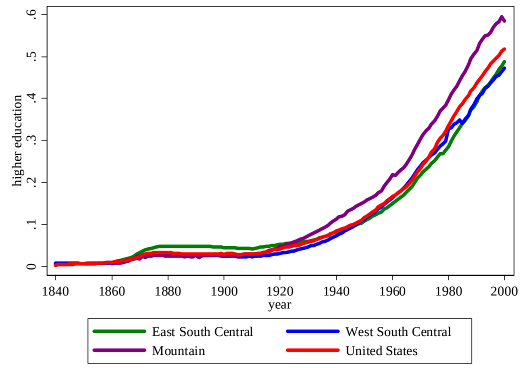


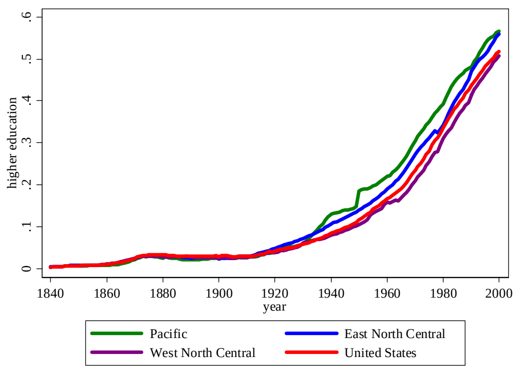


For secondary schooling exposure and no more, the nine census regions behave much like they did in elementary schooling exposure. From 1840 to 1940 the Pacific, and for 1840 to 1960, the Middle Atlantic, East North Central and West North Central regions display higher than average shares exposed to secondary schooling. As Goldin (1999) and Goldin and Katz (2000) have shown, these were the leaders of the high school movement in the US as well as the world. The South Atlantic, East South Central, West South Central regions all lagged behind the average for the US from 1840 to the present. Combining these exposure rates with the primary exposure rates shows that the South Atlantic, East South Central, and West South Central clearly have the smallest portion of their labor force exposed to higher education.

The next three graphs present this the evidence for higher education. The regions with higher share of the labor force exposed to higher education are New England, West North Central, Mountain and Pacific. The South Atlantic, East South Central and West South Central regions remain below average throughout the entire time period. The Middle Atlantic and East North Central regions seem to almost mimic the national average.







III. STATE PER WORKER OUTPUT

This section presents both original and existing data on state per worker output converted into real 2000 dollars.[24] In addition to the work of Easterlin (1960a,b), who provides per capita income in 1840, 1880, 1900, and 1919-1921 (1920), and government data from 1929-2000, we add our original estimates of state per capita income for 1850, 1860, 1870, 1890, and 1910. Our work uses government sources to produce estimates of real agricultural output, manufacturing output and mining output per state for these years. In combination with our measures of the labor force and the sectoral allocation of the labor force, we construct estimates of the non-agricultural, non-manufacturing non-mining output. With these estimates we create output per worker by state. The details of these calculations are in Appendix C. We note that the data from 1850-1920 are for state output per worker. For the period 1929-2000, the data are for state income per worker.

The next set of figures displays the average output per worker in each census region and the

[24] We convert all nominal values into real 2000 dollars, using the GDP deflator data from Gordon (1999) for years 1870-2000. For values between 1840-1869 we use the wholesale price index from the *Historical Statistics of the United States: Colonial Times to 1970* to compute inflation rates over this period. We then use the calculated wholesale price inflation to create a GDP deflator for the 1840-1869 period. To account for regional price differences, we use Berry, Fording, and Hanson (2000), Mitchener and McLean (1997), and Williamson and Linder (1980). The first deflators provide measures of output or income in constant national dollars and the regional price corrections adjust for regional price variation. For the 1840-1880 period we extrapolated the trend in relative price levels for the Mountain and Pacific region. Thus the output measures are best thought of as real income per worker. More details on price level are available in Appendix B.

national average output per worker. As with the educational measures, we present the data in regional aggregates in order to easily facilitate data presentation. The real income per worker series has many similarities with the educational attainment data. The Middle Atlantic and Pacific regions are consistently more productive than the US from 1840-2000, and the South Atlantic, East South Central and West South Central regions are consistently less productive than the US from 1840-2000. The remaining three regions, Mountain, West North Central and East North Central are essentially as productive as the US from 1840-2000.

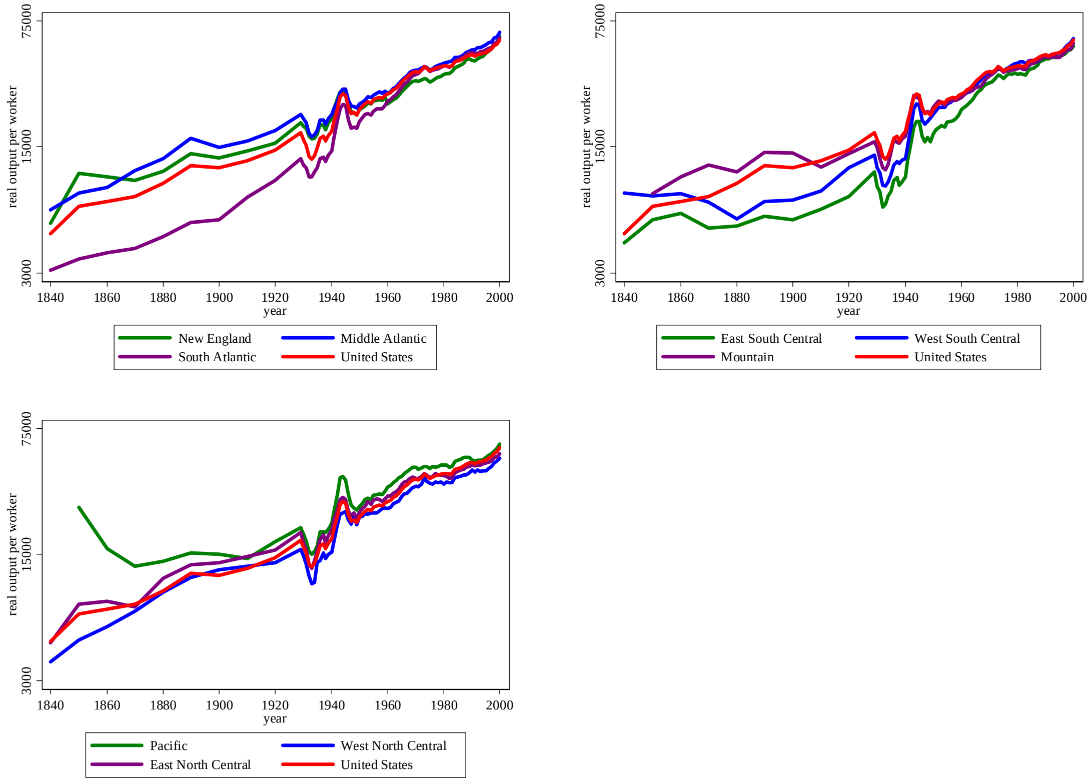


Table 3: Real Output per Worker
(regional leaders in bold)

	1840	1860	1880	1900	1920	1940	1960	1980	2000
United States	4950	7490	9448	11477	14430	18328	29514	42083	58791
New England	5640	10216	10998	13073	15706	21518	26042	38074	61426
Middle Atlantic	6709	8952	12954	14947	**18469**	**22639**	29854	43667	**64758**
South Atlantic	3089	3882	4751	5929	9770	14278	26982	42058	60216
E. S. Central	4391	6442	5447	5900	7947	10240	24092	37899	54134
W. S. Central	**8363**	8209	5971	7641	11512	12993	28521	43845	59833
Mountain	-	10236	10913	13838	13823	17247	28272	40690	56277
Pacific	-	**16167**	**13787**	**14992**	17607	22302	**35638**	**47185**	61374
W. N. Central	3825	5945	9248	12395	13497	15515	26991	36952	51527
E. N. Central	4867	8265	11147	13440	15841	20512	31641	40972	54162
region $\frac{\max}{\min}$	2.71	4.16	2.90	2.54	2.32	2.21	1.48	1.28	1.26
state max.	9218	16672	18972	17088	20492	28797	38531	62117	82438
state min.	2660	3144	3297	3678	6019	7135	20032	31558	41653
state $\frac{\max}{\min}$	3.47	5.30	5.75	4.65	3.40	4.04	1.92	1.97	1.98

As apparent in the figures as well as Table 3, real output per worker has increased substantially in the US, and across all regions. Consistent with evidence for the US from Baier, Dwyer and Tamura (2005), real output per worker grew at an annual rate of 1.6 percent per year. The nine census regions had annual real output per worker growth rates of 1.5 (New England), 1.4 (Middle Atlantic), 1.9 (South Atlantic), 1.6 (East South Central), 1.2 (West South Central), 1.2 (Mountain), 1.0 (Pacific), 1.6 (West North Central) and 1.5 (East North Central). The surprising values come from the West South Central, Mountain and Pacific.

In the case of the West South Central, the high value in 1840 comes from Louisiana, with real output per worker of 9218 dollars. Workers in the only other state in this region for 1840, Arkansas, realized a real output per worker of 5313 dollars. From 1860 to 2000, the West South Central saw real output per worker grow at 1.4 percent per year.

For the Mountain region in 1860, only New Mexico and Utah are in the data. Each has worker productivity in excess of 9800 dollars, well above the US value of 7500 dollars. In 1870 Colorado, Montana and Nevada enter the data. Montana, New Mexico and Utah all have worker productivity of about 5900 dollars, however Colorado and Nevada are very productive mining states, eaching having worker productivity in excess of 20,000 dollars. In 1880, Arizona and Idaho arrive in the data; all but New Mexico and Utah have worker productivity in excess of the US average, 9447 dollars.

In the case of the Pacific region, California, Oregon and Washington all have real output per worker values in excess of 10,000 dollars. These states were likely very high cost of living states as many manufactured goods would have to be imported from the rest of the US or abroad. Real output per worker for the Pacific region grows at an annual rate of 1.3 percent from 1880-2000 and 1.4 percent from 1900-2000, and 1.5 percent per year from 1920-2000.

Our results are also consistent with Goldin and Margo (1992a) who find falling real wages for artisans, laborers and clerical workers between 1840-1856. Between 1840 and 1860 we find that nonagricultural workers real output falls from 8009 to 6986, or a decline of 0.7 percent per year.[25] While agricultural workers saw rising output per worker from 1840 to 1860, 3984 dollars to 7937 dollars, their share of the labor force fell from 76 percent to 53 percent.

The effects of the Civil War are quite prominent in the figures, and are evident in Table 3. The

[25] Between 1840 and 1856 Goldin and Margo (1992a) present annualized real wage growth rates for artisans, laborers and clerical workers as: -0.7, 0.4 and 0. These average figures are obtained by equally weighting each of their four geographic regions.

states of the old Confederacy, South Atlantic, East South Central and West South Central clearly have lower growth rates. Between 1860 and 1880, these three regions experienced real annual income per worker growth of 1.0 percent, -0.8 percent and -1.6 percent, respectively. For the South Atlantic and East South Central, these understate the magnitude of the reduction in output per worker since 1870 is the nadir for these regions. The annual growth rates of income per worker from 1860 to 1870 for these three regions are 0.6 percent, -1.9 percent and -1.0 percent, respectively. In 1860 their relative worker productivity values were 52, 86, and 100 percent of the natinal average, while in 1880 their relative productivity have fallen to 50, 58, and 63 percent respectively. By 2000 only the East South Central remains below the national average.

The final four rows of Table 3 present evidence on regional output per worker convergence. These contain the ratio of the maximum regional income per worker to minimum regional income per worker, the maximum and minimum state per worker income, and the ratio of the maximum state income per worker to minimum state income per worker. Inequality in 1870 and 1880 are certainly higher than in the pre Civil War period. Inequality in output per worker is reduced throughout the next century. By 1980 the relative region gap is about one third of its value in 1880, and the relative state gap is less than a third of its 1880 value. Though the relative state gap has increased somewhat in 2000 compared to its 1980 value, the relative region gap has fallen since 1880.[26]

IV. ROBUSTNESS CHECK: RETURNS TO SCHOOLING

Though our estimated years of schooling appear similar to national estimates by Schultz and Fishlow, we also estimate returns to state-level measures of schooling to determine if our measures exhibit reasonable returns. Before we present evidence on the rate of return to schooling, it is necessary to deal with missing data on other inputs. Consider a model with two factors of production, human capital and all other inputs which we call physical capital. We assume production of a single final output is Cobb-Douglas. We assume perfect competition in factor markets and free mobility of capital. Output per worker in state i is given by:

$$y_{it} = A_{it} k_{it}^{\alpha} \left(human\ capital\right)_{it}^{1-\alpha} \tag{17}$$

where k_{it} is physical capital per worker and $human\ capital_{it}$ is human capital per worker. Under perfect competition in the output market, with final output as numeraire, the representative firm solves:

$$\max \left\{ A_{it} k_{it}^{\alpha} \left(human\ capital\right)_{it}^{1-\alpha} - r_t k_{it} - w_t human\ capital_{it} \right\} \tag{18}$$

where r_t and w_t are the rental rate per unit of physical capital and human capital, respectively. Under competition firms choose physical capital in proportion to the human capital in the workforce:

$$k_{it} = \left(\frac{w_t}{r_t}\right) \left(\frac{\alpha}{1-\alpha}\right) human\ capital_{it} \tag{19}$$

Therefore substituting this back into the output equation produces:

$$y_{it} = A_{it} \left(\frac{w_t}{r_t} \left(\frac{\alpha}{1-\alpha}\right)\right)^{\alpha} human\ capital_{it} \tag{20}$$

[26] These results are consistent with those found using state income per capita from 1880, 1900, 1920 and 1930-1990 at the decadal frequency in Tamura (2001).

We assume that *human capital*$_{it}$ can be specified in a Mincerian fashion:

$$\textit{human capital}_{it} = \exp\left(\beta E_{it} + \gamma x_{it}\right) \tag{21}$$

where E_{it} is years of schooling in state i in year t, and x_{it} is experience in state i in year t.[27] In order to construct average experience by state, we calculated average age in the state not enrolled in school and under the age of 65. From average age we subtract the sum of our average years of schooling measure in the labor force and the 6 years before individuals typically begin school enrollment. With this definition of *human capital*$_{it}$ the "earnings regression" is:

$$\ln y_{it} = \ln A_{it} + \alpha \ln\left(\frac{w_t}{r_t}\left(\frac{\alpha}{1-\alpha}\right)\right) + \beta E_{it} + \gamma x_{it} \tag{22}$$

Identification of β in (21) requires assumptions on the nature of state specific levels of Total Factor Productivity, A_{it}, as well as the national wage-rental ratio. If we assume that each state has a common level of TFP, and that labor and physical capital are perfectly mobile, then we can estimate (21) using time dummies in a pooled time series cross section. The coefficient on years of schooling identifies β.[28]

We first estimate (22) on each year, because of the possibility of technological progress. If the returns to schooling and experience are constant over time, but Total Factor Productivity rises over time, i.e. rising A_{it}, then any inability to properly control for the rising level of TFP will induce an upward bias on our estimates to schooling. The following figure presents the annual variation of the returns to schooling with one standard error bands.[29] With only four exceptions the estimates are always positive, and with very few exceptions the estimates are at least two standard errors away from zero. It is clear from the figure that the rate of return to schooling fell dramatically during the Depression; from 1929-1936 our estimates do not differ statistically from 0. However from 1937-1959 rates of return to schooling exceed 8 percent and for 1942-1959 an additional year of schooling returns in excess of 11 percent in every year. These high rates of return correspond to the

[27] Those familiar with the standard Mincer earnings regression may wonder why we exclude the quadratic term in experience. This is because of aggregation bias. While one can construct a model in which the linear terms in education and experience are identified by state variation, the quadratic term is not identified upon aggregation. When we experimented with identification, the results confirmed the bias in estimation, and hence we ignore the diminishing returns to work experience. The results indicate that experience returns are significantly below that from additional schooling and hence suggest that ignoring the quadratic term is not problematic.

[28] If we drop the assumption that labor and physical capital are perfectly mobile across state boundaries, (21) becomes:

$$\begin{aligned}\ln y_{it} &= \ln A_{it} + \alpha \ln\left(\frac{k_{it}}{\textit{human capital}_{it}}\left(\frac{\alpha}{1-\alpha}\right)\right) + \beta E_{it} + \gamma x_{it} \\ &= \ln A_{it} + \alpha \ln k_{it} + \alpha \ln\left(\frac{\alpha}{1-\alpha}\right) + \beta(1-\alpha)E_{it} + \gamma(1-\alpha)x_{it}\end{aligned}$$

We can estimate the above equation with state specific time trends and time dummies. Our estimate on years of schooling will be a combination of both the rate of return to schooling and labor's share of income. Therefore we need an estimate of the share of output that labor receives, $(1-\alpha)$. Table 4 in Appendix D presents evidence on labor's share, it varies between $\frac{2}{3}$ and $\frac{4}{5}$ throughout the 100 years of observations. Our estimates using this methodology are generally higher than those presented in the paper. We do not report them for brevity, they are available on request from the authors.

[29] The results come from annual weighted regressions of log state output per worker on years of schooling and experience, where the weights are the labor force of each state. Over the entire period, 1840-2000, the mean return, inclusive of physical capital's return, to a year of schooling is .0949 with a mean standard error of .0385. From 1880-2000, the mean return, inclusive of physical capital's return, to a year of schooling is .1012 with a mean standard error of .0347.

diminishing dispersion of education across states, as well as rising levels of schooling. Together these help to explain the Great Compression in the middle of the 20th century as identified by Goldin and Margo (1992b). The falling returns to an additional year of schooling from the 1950s through the 1970s is consistent with the work of Freeman (1976). Although muted due to aggregation, the rising returns to schooling in the latter half of the 1980s and the recovery from the 1990-1991 recession are consistent with those found in Murphy and Welch (1992).

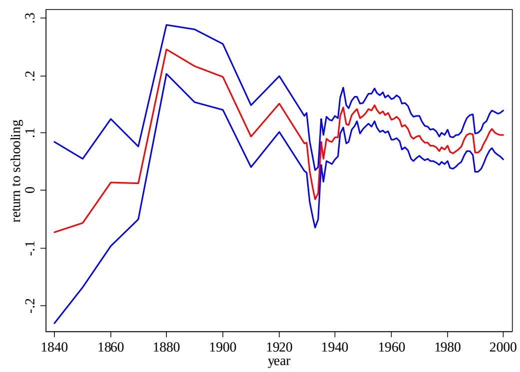


Table 5 contains the results of real per worker output regressed on years of schooling. The first three columns include year dummies to allow for more variation in technological change than a deterministic trend. The second column allows for a different return to schooling for Alaska. The third column allows for a different return to schooling and a different return to experience in Alaska. Under the hypothesis that TFP does not differ across states, i.e., $A_{it} = A_t$ for all i, differencing each state's log output per worker from the labor force weighted log US output per worker, years of schooling, and average experience from the labor force weighted US averages allows for the estimation of (Eq. 21) without any time controls. These differenced regressions are reported in the final three columns of Table 5.

Table 5: Earnings Regressions: Annual Data (standard errors)

E	.1330	.1324	.1324	.1330	.1329	.1326
	(.0044)	(.0044)	(.0044)	(.0044)	(.0044)	(.0043)
exp.	.0460	.0477	.0477	.0460	.0461	.0470
	(.0017)	(.0017)	(.0017)	(.0017)	(.0017)	(.0017)
N	4000	4000	4000	4000	4000	4000
$\overline{R}^2$	.9137	.9148	.9148	.3316	.3318	.3368
year dummies	yes	yes	yes	no	no	no
ak E	no	yes	yes	no	yes	yes
ak exp.	no	no	yes	no	no	yes
differenced	no	no	no	yes	yes	yes

The results in Table 5 indicate an overall return to schooling, including the implied physical capital return, of 13 percent per year of schooling. These results are consistent with the evidence presented in Angrist and Krueger (1991), Staiger and Stock (1997), and Card (1995). The returns to experience, reflecting on-the-job training or learning by doing, are similar across all four columns. A one year increase in average experience raises worker productivity by about five percent.

Failing to account for the rising female labor force participation rate present over this period may result in poor estimates. To control for this we correct for the share of the labor force that is female (male) and interact these shares with average years of experience. This allowed us to separately measure the rate of return to experience for each sex. The results of these are contained in Table 6. The first three columns report the average return to schooling and average estimated returns to experience by sex with varying controls for Alaska. The remaining three columns are the differenced regressions, as in Table 5. The rows marked F and Prob $> F$ contain the F statistic on the test of equality of returns to experience between men and women, and the p value of the statistic.

Table 6: Earnings Regressions: Annual Data (standard errors)

E	.1269	.1264	.1265	.1268	.1269	.1266
	(.0045)	(.0045)	(.0045)	(.0045)	(.0045)	(.0047)
exp male	.0604	.0619	.0622	.0605	.0604	.0614
	(.0030)	(.0030)	(.0030)	(.0030)	(.0030)	(.0030)
exp female	.0295	.0313	.0318	.0301	.0301	.0309
	(.0033)	(.0033)	(.0033)	(.0033)	(.0033)	(.0033)
N	4000	4000	4000	4000	4000	4000
$\overline{R}^2$	.9144	.9156	.9158	.3369	.3379	.3430
year dummies	yes	yes	yes	no	no	no
F	33.94	33.93	33.52	35.82	33.59	34.05
Prob $> F$	.0000	.0000	.0000	.0000	.0000	.0000
ak E	no	yes	yes	no	yes	yes
ak exp (male & female)	no	no	yes	no	no	yes
differenced	no	no	no	yes	yes	yes

The results of Table 6 indicate that the estimated returns to schooling are robust to the possible differences in returns to experience between men and women. It is reasonable to state that an additional year of schooling in a randomly chosen state returns 13 percent. Rates of returns to experience for men and women are different. In all six regressions we reject the null that they are identical. The typical male worker becomes about 6 percent more productive at the individual level per additional year of experience, while the typical female worker becomes about 3 percent more productive at the individual level per additional year of experience.

One might be concerned that our estimates of the return to schooling may be biased because we assume a common intercept for all states in any time period. To address this concern, one way to correct for this is to allow for state specific effects. To help guide our think about alternative specifications that would correct for this potential bias, we return to equation (21)

$$\ln y_{it} = \ln A_{it} + \alpha \ln \left(\frac{w_t}{r_t} \left(\frac{\alpha}{1-\alpha} \right) \right) + \beta E_{it} + \gamma x_{it} \tag{23}$$

One way to rewrite the above specification in a form that allows for different intercept is to assume that the state specific technology is constant over time. The regression specification inferred from equation (21) is, therefore, given by:

$$\ln y_{it} = c_i + b_t + \beta E_{it} + \gamma x_{it} + u_{it} \tag{24}$$

where c_i is the state specific fixed effects and b_t is a time specific effect common to all states; in the context of equation (21) we are assuming that technology is given by $\ln(A_{it}) = c_i + u_{it}$ and that there are national labor and capital markets so that $\frac{w_t}{r_t} = b_t$. To correct for the state specific effects, there are two standard approaches to adjust for these effects: (1) fixed effects regressions or (2) OLS on first differenced data. In both cases, it is required that there are no feedback effects from innovations in income to future levels of educational attainment. If this is the case then standard fixed effects regressions or first-differencing will lead to inconsistent estimates of the return to schooling. The first column of Table 7, reports the results of standard fixed effects regression. To see if feedback effects are present, we follow Wooldridge (2002) and run a fixed effects regression with a lead of educational attainment in the specification. If the coefficient on educational attainment is statistically different from zero, then we will take this as evidence that contemporaneous innovations in income lead to future educational attainment. These results are reported in column 2 of Table 7.

Table 7: Fixed Effects with Leads of Education

	(1)	(2)
E	0.114	0.063
	(.0150)	(0.027)
$E(t+1)$	0.054	
	(.028)	
exp	0.0420	0.038
	(.0060)	0.007
N	663	663
Decade Dum.	yes	yes

With time dummies, the return to schooling form the fixed effects regression is roughly 11 percent. When we add a lead of education to the fixed effects regression, the coefficient has a p-value of 0.057. Thus, at the five percent level we would not reject the null hypothesis of no feedback effects. However, this p-value is sufficiently low that we would like allow for feedback effects from current income to future education and experience. If feedback effects are present, the standard approaches to correct for state effects will lead to inconsistent estimates. To correct for the possibility of state specific effects, we difference the data in equation (21) to get

$$\Delta \ln y_{it} = \Delta b_t + \beta \Delta E_{it} + \gamma \Delta x_{it} + \Delta u_{it} \tag{25}$$

Feedback effects from innovations in income to education implies $E\left(\Delta E_{it} \Delta u_{it}\right) = E[(E_{it} - E_{it-1})(u_{it} - u_{it-1})] \neq 0$. To consistently estimate the above equation we must find instruments for ΔE_{it} that satisfy the standard instrumental variable assumptions; that is, (1) the instruments should be correlated with ΔE_{it} and (2) the instruments are uncorrelated with the error term. Following Arrelano and Bond (1991), we use lags of educational attainment and experience. As additional instruments, we experimented with lags of the difference between state i's average educational attainment and the average educational attainment of the other states in the region – this variable may capture the changes in educational attainment related to regional convergence. More specifically, we create the variable

$$E_{it}^{c} = \left[E_{it} - \frac{1}{N^{R}-1} \sum_{j \neq i}^{N^{R}} E_{jt} \right] \tag{26}$$

where N^R is the number of states in region R. Thus, the variable E_{it}^c measures how far ahead or behind state i is relative to the rest of the states in the region. We use different lag levels of state i's educational attainment, experience, and E_{it}^c as instruments for ΔE_{it} and exp in a two-step GMM estimator In particular we rewrite equation (23) as

$$\Delta \ln y_{it} = \Delta x_{it}^{*} \beta^{*} + \Delta u_{it}$$

$\Delta$where $\Delta x_{it}^{*} = (\Delta b_t, \Delta E_{it}, \Delta x_{it})$ and $\beta^{*} = (\beta_t, \beta, \gamma)$. The instrument matrix for state is given by

$$Z_i = \begin{bmatrix} z_i^0 & 0 & 0 & 0 \\ 0 & z_i^1 & 0 & 0 \\ 0 & 0 & ... & 0 \\ 0 & 0 & 0 & z_i^{t-1} \end{bmatrix}$$

where z_i^j contains two to four lags of education and experience. The two step GMM procedure minimizes

$$J_N = \left(\frac{1}{N} \sum_{\iota=1}^{N} Z_i \Delta u_i \right)' W_N \left(\frac{1}{N} \sum_{\iota=1}^{N} Z_i \Delta u_i \right)$$

where the weighting matrix is given by

$$W_N = \left(\frac{1}{N} \sum_{\iota=1}^{N} Z_i' \Delta \widetilde{u}_i \Delta \widetilde{u}_i' Z_i \right)$$

with $\Delta u_i = \Delta y_i - \Delta x_i \widehat{\beta}$ where $\widehat{\beta}$ is from the first stage regression when W_N is the identity matrix and is adjusted for small sample bias (Windmeijer 1995). Table 8 present the results when we use two to four lags of educational attainment and experience. When we use just lagged levels of the state's educational attainment and experience, we find the average return to an additional year of schooling is roughly 14 to 16 percent. We we include the state's deviation from the regional average, the average return to an additional year of schooling ranges between 10 to 14 percent. Thus the results obtained using the dynamic panel where we allow for feedback effects from income to education yield estimates that are similar to standard Mincerian regressions

	IV Educ	IV Educ	IV Educ	IV Educ	IV Educ	IV Educ
E	0.162	0.148	0.145	0.148	0.120	0.106
	(.037)	(.035)	(.033)	(.030)	(.042)	(.050)
exp.	0.028	0.038	0.043	0.037	0.027	0.046
	(.015)	(.0140)	(.013)	(.015)	(.020)	(.020)
N	663	663	663	663	663	663
Instruments	2 lags ed,exp	3 lags ed,exp	4 lags ed,exp	2 lags ed,exp	2 lags ed,exp	2 lags ed,exp
Decade Dum.	yes	yes	yes	yes	yes	yes

V. CONCLUSION AND EXTENSIONS

This paper employs historic state enrollment and population data to produce original average years of schooling measures for each state from 1840 to 2000. These measures will benefit any economics, social science, education, or history researcher searching for more consistent historic schooling measures for empirical studies. We show that there has been tremendous increases in schooling in the US over the 1840-2000 period, with average years of schooling rising from 1 year to over 13 years. In addition there has been a reduction in the variance across states. We also construct original estimates for state per worker output for the census years 1850, 1860, 1870, 1890 and 1910. Coupling our constructed data with previous work by Easterlin and government data, we produce state per worker income measures for 1840 through 1920 at the decadal frequency and 1929 through 2000 at the annual frequency. We then estimate aggregate Mincerian earnings regressions and discover that the return to a year of schooling for the average individual in a state ranges from 10 percent to 13 percent. This range is robust to various time periods, various estimation methods and various assumptions about the endogeneity of schooling.

Though many of the cross-country analyses have increased our knowledge of the importance of TFP and TFP growth in determining both the level differences in income as well as the growth rate of income and its variation, many economists, as listed in Temple (1999), object to the empirical work on growth. One objection is the inability to account for large heterogeneity in social, religious, and institutional characteristics. Another criticism is the small time frame over which cross-country inputs, income, and TFP are estimated. This work is part of a larger research agenda to construct a systematic analysis that overcomes both of these criticisms by analyzing cross-state income variation in the United States from 1840 to 2000. By creating and analyzing new state measures of human capital, physical capital, and income of the United States for 160 years, we inted to reduce both the possible problems associated with the social, religious, and institutional heterogeneity and the errors that can be induced by business cycles when comparing cross-sectional TFP over shorter periods of time. Therefore we expect to obtain a more precise measure of technology growth, and with it, a more comprehensive explanation of why income variation occurs across developing counties such as the United States in the 1800s.

Following the cross-country work of Klenow and Rodriguez-Clare (1997) and Easterly and Levine (2001), we also envision determining if the variance in the growth rate of TFP may account for the majority of the variance in the growth rate of output. Three possible sources of regional TFP variation include: variation in educational attainment by race; variation in educational quality; and variation in sectoral allocation of labor. We intend to merge additional data on demographics, educational quality, and labor allocation by sector to determine the impact that variation within a region has on the growth of TFP.

REFERENCES

Angrist, Joshua D. and Krueger, Alan B. "Does Compulsory School Attendance Affect Schooling and Earnings?", *Quarterly Journal of Economics*, 106, 1991: 979-1014.

Arellano, M., and Bond, S. “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” *Review of Economic Studies* 58, 1991: 277-297.

Baier, Scott, Dwyer, Gerald, and Tamura, Robert. “How Important Are Capital and Total Factor Productivity for Economic Growth?” , *Economic Inquiry*, 2005.

Barro, Robert and Lee, Jong-Wha. "International Comparisons of Educational Attainment," *Journal of Monetary Economics* 32, 1993: 363-394.

Berry, William D., Fording, Richard C., and Hanson, Russell L. "An Annual Cost of Living Index for the American States, 1960-95," *Journal of Politics*, 60 (May), 2000: 550-67.

Bills, Mark and Klenow, Peter J. "Does Schooling Cause Growth?" *Amercian Economic Review*. 90, December, 2000: 1160-1183.

Card, David. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling", in Louis N. Christofides, E. Kenneth Grant and Robert Swidinsky, eds. *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*. University of Toronto Press, Toronto, Canada. 1995. 201-222.

Card, David, and Krueger, Alan. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy* 100, 1992: 1-40.

Denison, Edward. *The Sources of Economic Growth in the United States and the Alternatives Before Us*, Committee for Economic Development: New York, 1962.

Easterlin, Richard. "Regional Growth of Income: Long Term Tendencies, 1880-1950." In *Population Redistribution and Economic Growth, United States, 1870-1950*, vol. 1, *Analyses of Economic Change*, edited by Simon Kuznets, Ann Ratner Miller, and Richard Easterlin. Philadelphia: American Philosophical Society, 1960a.

Easterlin, Richard. "Interregional Differences in Per Capita Income, Population, and Total Income, 1840-1950," in *Trends in the American Economy in the Nineteenth Century*, ed. William N. Parker, Princeton University Press: Princeton, 1960b.

Easterly, William and Levine, Ross. ""It's Not Factor Accumulaton: Stylized Facts and Growth Models", *World Bank Economics Review* 15, 2001, 177-219.

Fishlow, Albert. "The Common School Revival: Fact or Fancy?" *In Industrialization in Two Systems* (Essays in honor of Alexander Gerschenkron), Henry Rosovsky, ed. (John Wiley & Sons) 1966: 40-67.

Freeman, Richard. *The Overeducated American*, Academic Press: New York, 1976.

Goldin, Claudia. "America's Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century," *Journal of Economic History* 58, 1999: 345-374.

Goldin, Claudia, and Margo, Robert A. "Wages, Prices, and Labor Markets before the Civil War," in *Strategic Factors in Nineteenth Century American Economic History: A Volume to Honor Robert W. Fogel* (eds.) Claudia Goldin and Hugh Rockoff, NBER, University of Chicago Press: Chicago, 1992a.

Goldin, Claudia, and Margo, Robert A. "The Great Compression: The Wage Structure in the United States at Mid-Century," *Quarterly Journal of Economics* 107, 1992b: 1-34.

Goldin, Claudia, and Katz, Lawrence. "Education and Income in the Early 20^{th} Century: Evidence from the Prairies," *Journal of Economic History* 60, 2000: 782-818.

Gordon, Robert. *Macroeconomics*, Addison-Wesley: New York, 1999.

Klenow, Pater J., and Rodriguez-Clare, Andres. "The NeoClassical Revival in Growth Economics: Has It Gone Too Far?", *NBER Macroeconomics Annual*, 1997: 73-114.

Kuznets, Simon. *National Income: A Summary of Findings*, National Bureau of Economic Research: New York, 1946.

Lazear, Edward. "Educational Production" *Quarterly Journal of Economics.* 116, 2001:777-803

Long, Clarence D. *The Labor Force under Changing Income and Employment.* Princeton, Princeton University Press. 1958

Mitchener, Kris J. and McLean, Ian W. "U.S. Regional Growth and Convergence 1880 -1980." *Journal of Economic History*, 59, 1997: 1016-1042.

Mulligan, Casey, and Sala-i-Martin, Xavier. "A Labor-Income Based Measure of the Aggregate Value of Human Capital," *Journal of Japan and the World Economy* 9, 1997: 159-191.

Mulligan, Casey, and Sala-i-Martin, Xavier. "Measuring Aggregate Human Capital," *Journal of Economic Growth* 5, 2000: 215-252.

Murphy, Kevin M., and Welch, Finis. "The Structure of Wages," *Quarterly Journal of Economics* 107, 1992: 285-326.

National Catholic Education Association. *United States Catholic elementary and secondary schools*, National Catholic Education Association: Washington, D.C., various years.

Schultz, Theodore. "Education and Economic Growth," in *Social Forces Influencing American Education*, N.B. Henry, ed., Chicago: University of Chicago Press, 1961.

Snyder, Thomas, Hoffman, Leff, and Geddes, Claire. *State Comparisons of Education Statistics: 1969-70 to 1996-97*, National Center for Education Statistics: U.S. Department of Education: Washington, D.C., 1998.

Staiger, Douglas and Stock, James H. "Instrumental Variables Regression with Weak Instruments", *Econometrica* 65, 1997: 557-586.

Tamura, Robert. "Teachers, Growth and Convergence," *Journal of Political Economy* 109, 2001: 1021-1059.

Temple, Jonathon. "The New Growth Evidence," Journal of Economic Literature 37, 1999: 112-156.

United States Census Bureau. *Statistical Abstracts of the United States*, U.S. Government Printing Office: Washington, D.C., various years.

United States. Congress. Senate. By American Statistical Association. 28th Cong., 2nd sess. Senate. Document No. 5. Washington: GPO, 1845.

United States Department of Commerce. *Historical Statistics of the United States: Colonial Times to 1970*, U.S. Government Printing Office: Washington, D.C., 1975.

United States Department of Education. *Digest of Education Statistics*, U.S. Government Printing Office: Washington, D.C., various years.

United States Department of Health, Education and Welfare. *Projections of Educational Statistics to ...*, U.S. Government Printing Office: Washington, D.C., various years.

Welch, Finis. "Measurement of the Quality of Schooling," *American Economic Review* 56, 1966: 379-392.

Williamson, Jeffrey G. and Linder, Peter H. *American Inequality*, Academic Press 1980: 97-132.

Windmeijer, ...

Wooldridge, Jeffrey. *Econometric Analysis of Cross Section and Panel Data*, M.I.T. Press: Cambridge, MA 2002.

APPENDIX A

There are nine census regions in the US. The following Table provides the regional groups.

New England	*Middle Atlantic*	*South Atlantic*	*E. South Central*	*W. South Central*
Connecticut	New Jersey	Delaware	Alabama	Arkansas
Maine	New York	D.C.	Kentucky	Louisiana
Massachusetts	Pennsylvania	Florida	Mississippi	Oklahoma
New Hampshire		Georgia	Tennessee	Texas
Rhode Island		Maryland		
Vermont		North Carolina		
		South Carolina		
		Virginia		
		West Virginia		
Mountain	*Pacific*	*W. North Central*	*E. North Central*	
Arizona	Alaska	Iowa	Illinois	
Colorado	California	Kansas	Indiana	
Idaho	Hawaii	Minnesota	Michigan	
Montana	Oregon	Missouri	Ohio	
Nevada	Washington	Nebraska	Wisconsin	
New Mexico		North Dakota		
Utah		South Dakota		
Wyoming				

Tables A1-AC contain the time series of average elementary school enrollment rates, secondary school enrollment rates, and higher education enrollment rates by census region as well as for the US as a whole from 1840-2000. We note that the elementary enrollment rates are often over 100 percent. In the early periods, higher elementary enrollment rates are due to two factors: older aged first time enrollment and less social promotion. The methodology we present addresses a portion of these sources.

Table A1: Average Elementary Enrollment Rates

	1840	1860	1880	1900	1920	1940	1960	1980	2000
US	48.1	74.7	95.9	106.8	108.4	104.4	100.8	100.6	105.1
New England	129.1	118.9	118.8	116.9	109.4	106.5	101.7	102.1	106.3
Mid Atlantic	75.4	94.0	114.7	108.2	103.4	106.6	104.4	98.8	105.0
So. Atlantic	13.7	28.1	73.4	95.4	106.6	101.7	98.1	101.7	106.7
E. So. Central	13.3	42.3	75.3	104.2	114.2	109.1	98.9	100.4	108.9
W. So. Central	8.2	24.9	46.4	82.9	104.4	101.1	97.1	102.5	108.2
Mountain	-	19.1	82.7	109.4	114.4	101.5	100.5	100.7	100.6
Pacific	-	70.4	108.6	121.0	122.1	106.6	99.9	100.3	103.3
W. N. Central	18.2	77.9	105.1	119.9	112.7	105.6	102.5	99.5	103.1
E. N. Central	45.6	111.8	113.9	113.6	107.2	102.7	102.0	100.0	104.3

Table A2: Average Secondary Enrollment Rates

	1840	1860	1880	1900	1920	1940	1960	1980	2000
US	2.0	3.1	4.0	10.3	28.0	72.4	84.9	89.2	92.7
New England	4.7	4.2	4.2	20.5	40.4	78.4	88.6	90.5	94.7
Mid Atlantic	2.9	3.8	4.7	12.5	27.9	83.0	90.0	94.9	97.2
So. Atlantic	0.8	1.5	3.6	5.1	14.6	56.8	76.8	85.2	91.0
E. So. Central	0.8	2.0	3.5	4.7	12.1	44.2	74.7	83.8	88.9
W. So. Central	0.5	1.5	2.9	4.8	18.7	63.8	81.7	84.6	89.7
Mountain	-	0.9	5.0	10.5	40.4	76.9	86.3	87.6	88.6
Pacific	-	3.3	4.0	12.9	56.4	89.2	86.1	90.0	98.3
W. N. Central	0.8	3.2	3.9	11.7	37.2	78.2	90.0	91.6	93.7
E. N. Central	1.7	4.3	4.2	13.4	34.4	80.5	88.5	90.9	90.1

Table A3: Average Higher Education Enrollment Rates

	1840	1860	1880	1900	1920	1940	1960	1980	2000
US	0.7	1.4	0.9	1.5	6.2	8.4	22.2	40.4	57.0
New England	0.9	0.9	1.2	1.8	8.6	8.2	26.6	47.2	71.9
Mid Atlantic	0.6	0.7	0.7	1.2	5.9	7.7	22.5	40.1	59.0
So. Atlantic	0.6	1.7	1.2	1.3	4.8	6.4	16.1	35.6	53.9
E. So. Central	0.7	1.6	1.7	1.6	2.9	5.6	16.4	31.8	48.0
W. So. Central	1.7	1.7	0.6	0.9	3.7	8.3	20.2	33.9	47.7
Mountain	-	1.4	1.3	2.1	5.9	10.5	25.6	42.0	61.5
Pacific	-	1.5	1.1	2.2	10.2	12.9	29.5	54.1	59.8
W. N. Central	0.9	2.1	0.7	1.8	8.0	9.6	24.4	38.3	62.6
E. N. Central	0.7	1.7	0.5	1.4	7.4	9.2	22.5	39.3	58.0

APPENDIX B

In this Appendix we provide more details on the calculations of years of schooling.

I. Data description
- A. Public Elementary / Secondary Enrollment
- B. Private Elementary / Secondary Enrollment
- C. Higher Educational Enrollment
- D. Population: 5-13, 14-17, 18-24
- E. Labor Force
- F. Price levels
- G. Expected years

II. Describe calculation of
- A. Enrollment rates
- B. Shares (primary, secondary, college)
    - 1. Initial conditions
    - 2. Higher education inflow constant
    - 3. Secondary departure rates
- C. Shares for foreign born

III. Idiosyncrasies
- A. DC / MD / VA

B. AK / HA
C. ND / SD / Dakota
D. OK / Indian Territory
IV. Table listing first year of data availability
V. References

Data Description

Public Enrollment Data.—

Public Enrollment, 1840-1916 Data for total (elementary and secondary) public enrollment are available from decennial census data, by state, in 1840, 1850, 1860, 1870. Total public enrollment data are available in *Statistical Abstracts of the United States* for the years 1872, 1877, 1879-1887, 1889-1891, and 1893-1916.

Data for total public enrollment for non-decennial years between 1840 and 1870 was geometrically interpolated. Data for the years 1871, 1873-1876, 1878, 1888, and 1892 was also geometrically interpolated.

We do not observe the fraction of total public enrollment that is elementary versus secondary until the year 1899. However, we do have national aggregates that make this breakdown in 1870, 1880, and 1890-1898.

Letting $pub.enroll_{it}^{\text{primary}}$ designate the public primary enrollment level in state i for time period t, and $pub.enroll_{it}^{\text{total}}$ refer to the total (primary and secondary) enrollment level, we assign:

$$pub.enroll_{it}^{\text{primary}} = pub.enroll_{it}^{\text{total}} \frac{\sum_j pub.enroll_{j,1870}^{\text{primary}}}{\sum_j pub.enroll_{j,1870}^{\text{total}}}, \ t \leq 1870 \tag{27}$$

$$pub.enroll_{it}^{\text{primary}} = pub.enroll_{it}^{\text{total}} \frac{\sum_j pub.enroll_{j,1880}^{\text{primary}}}{\sum_j pub.enroll_{j,1880}^{\text{total}}}, \ 1871 \leq t \leq 1880 \tag{28}$$

$$pub.enroll_{it}^{\text{primary}} = pub.enroll_{it}^{\text{total}} \frac{\sum_j pub.enroll_{j,1890}^{\text{primary}}}{\sum_j pub.enroll_{j,1890}^{\text{total}}}, \ 1881 \leq t \leq 1890 \tag{29}$$

$$pub.enroll_{it}^{\text{primary}} = pub.enroll_{it}^{\text{total}} \frac{\sum_j pub.enroll_{jt}^{\text{primary}}}{\sum_j pub.enroll_{jt}^{\text{total}}}, \ 1891 \leq t \leq 1898 \tag{30}$$

$$pub.enroll_{it}^{\text{secondary}} = pub.enroll_{it}^{\text{total}} - pub.enroll_{it}^{\text{primary}} \tag{31}$$

Beginning in 1899, we observe both $pub.enroll_{it}^{\text{total}}$ and , $pub.enroll_{it}^{\text{secondary}}$ so we can simply calculate $pub.enroll_{it}^{\text{primary}}$.

Public Enrollment, 1918 - 1968 Data for public secondary enrollment and for total public enrollment are available biennially in the *Statistical Abstract of the United States* (even numbered years) from 1918 – 1968. In addition, data is also available in 1925, 1945, 1947, and 1949, 1955, and 1959. We geometrically interpolate any missing values from 1918 – 1968.

Public Enrollment, 1969 - 2000 Data from 1969 to 2000 are annual, and come from NCES, *State Comparisons of Education Statistics: 1969-70 to 1996-97*, as well as updates available from the NCES website.

Private Enrollment Data.—

Private Enrollment, 1840 - 1916 Data for total private enrollments are available from various censuses, by state in 1840, 1850, 1860, 1870, 1890, 1910, and 1920. We geometrically interpolate between the decennial values listed above for any non-decennial years.

Data for total private secondary enrollments are available on an annual basis from 1899 to 1916 from the *Statistical Abstracts of the United States*. For these years, we are able to take the measure of total private enrollment above and subtract secondary enrollment to arrive at private elementary enrollment.

Prior to 1899, we observe total private enrollment, but do not observe the breakdown into elementary and secondary. However, we do observe national aggregates in 1890. Proceeding as we did above in the public case, we calculate:

$$pri.enroll_{it}^{\text{primary}} = pri.enroll_{it}^{\text{total}} \frac{\sum_j pri.enroll_{j,1890}^{\text{primary}}}{\sum_j pri.enroll_{j,1890}^{\text{total}}}, \; t \leq 1890 \tag{32}$$

$$pri.enroll_{it}^{\text{secondary}} = pri.enroll_{it}^{\text{total}} - pri.enroll_{it}^{\text{primary}} \tag{33}$$

We also geometrically interpolate the secondary enrollment figures for 1891-1898 using the 1890 value (calculated directly above), and the 1899 figures.

Private Enrollment, 1918 - 1968 Data for private secondary enrollment and total private enrollment are available biennially in *Statistical Abstracts of the United States* (even numbered years) from 1918–1940 and 1948–1968. Data is also available in 1925, 1947, and 1949, 1955, and 1959. We geometrically interpolate any missing values from 1918 – 1968.

Private Enrollment, 1969 - 2000 For the years 1968 – 1980, 1991, 1993, 1995, 1997, and 1999, we observe private elementary and secondary enrollment figures from the *Digest of Education Statistic*s. We geometrically interpolate the 1992, 1994, 1996, and 1998 values.

For the years between 1980 through 1991, we are unable to obtain private elementary and private secondary enrollment figures by state directly. However we are able to obtain annual estimates of the national private elementary and private secondary totals from Projections of Education Statistics, various issues, as well as state level data on Catholic elementary and Catholic secondary enrollment figures in 1985, 1988, and 1990 – 1999 from the *National Catholic Education Association*, various issues. We assume that the distribution of total private elementary and total private secondary enrollment figures across states is identical to the distribution of Catholic elementary and Catholic secondary enrollment figures across states. We inflate the Catholic state level data enrollment data to correspond to the national totals for 1985, 1988, and 1990. We geometrically interpolate values for years 1981-1984, 1986-1987, and 1988.

Higher Education Enrollment.—

1840 – 1899

Data for states are available from decennial census data in 1840, 1850, 1860, and 1870. In 1886, 1890, and 1891 data are available, typically subdivided into Medical, Theological, Law, and Liberal Arts enrollments. Data for non-census years between 1840 and 1870, as well as 1871-1885, 1887-1889, and 1892-1898 are geometrically interpolated.

1899 – 1920

Data are reported annually in *Statistical Abstracts* under a variety of titles and formats. Total higher education enrollment is the sum of sources below, except where enrollment figures are included in more than one source.

1. Schools of Technology and Institutions conferring only the B.S. degree (1899-1905)
2. Colleges and Seminaries for Women which confer degrees (1899-1910)
3. Coeducational Colleges and Universities and Colleges for men only (1899-1916, 1918)
4. Undergraduate Students in Univ., Colleges, and Schools of Tech. (1911 – 1916, 1918, 1920)
5. Professional Schools (1899-1916)
6. Public and Private Normal Schools (1899-1916, 1918, 1920)
7. Training Schools for Nurses, Comm. Schools, Manual and Industrial Training Schools (1910-1916, 1918, 1920)

1922 – 1946

Data is reported biennially in the *Statistical Abstracts* from 1922-1940, various issues, as Enrollment in Universities, Colleges, and Preparatory Schools. Similar data is also reported as Higher Education Enrollment in 1942, 1944, and 1946. Non-biennial years are geometrically interpolated.

1947 – 1968

Data is reported annually in *Statistical Abstracts*, various issues, as Institutions of Higher Educational, Fall Enrollment.

1969 – 2000

Data is reported in *State Comparisons of Education Statistics*. Higher educational enrollment is the sum of 2-year private, 2-year public, 4-year private, and 4-year public higher educational enrollment.

Population.—

We generally observe the age distribution of population in decennial years, beginning in 1840. In most cases, we are given data with 5-year population distributions. The usual structure is

$<$5, 5-9, 10-14, 15-19, 20-24. . . 55-59, 60-64, 65-69, 70-74. . . .

With the exception of calculating the average age of the population in a state, we are ultimately interested in the age groups: 5-13, 14-17, 18-24, 16-65. In order to calculate the number of persons in each group, we assume a uniform distribution of population across the age groups.

In 1840, the white age distribution is reported, but only broad categories of the black age distribution are available. In order to allocate the total black distribution amongst the various age groups, we assume the fraction of total black population in each age group is identical to the fraction in the 1850 black distribution.

Labor Force.—

All labor force data prior to 1970 is decennial data. For non-decennial years prior to 1970, data is geometrically interpolated. Labor force data for 1840 – 1860 is decennial census data. Data for 1870 – 1940 is gainful workers, 10 years old and over, and is taken from *Historical Statistics of the United States: Colonial Times to 1970*, pp. 129–131. Data for 1950 and 1960 is decennial Census of

Population data, and includes persons aged 14 and over. Data from 1970 – 2000 is Civilian Labor Force, 16 years and older, and is taken from the Bureau of Labor Statistics website.

Price Levels.—

National price level data from 1875-1999 is the GDP deflator, as reported in Gordon, *Macroeconomics*, 7th edition, pp. A1–A3. National price level data prior to 1875 is the wholesale price index (all commodities) from Warren and Pearson, printed in *Historical Statistics of the United States: Colonial Times to 1970*, pp. 201-202. Data from 1840-1875 are normalized to correspond to the price level given by Gordon in 1875.

In addition, we use three sources of information on relative price levels across regions. Mitchener and McLean (1999) and Willamson and Linder (1980) provide regional price levels for census regions which we use from 1840-1960. Data from the two sources is primarily non-overlapping. Where we have data from both sources, we take the arithmetic average of the relative price level in each region. Prior to 1880 these sources does not include relative price levels for the Pacific and Mountain region. For data prior to 1880 in each of these two regions, we extrapolate the relative regional price level using the trend observed from 1880 to 1920. Berry, Fording and Hanson (2000) display price levels for each state on an annual basis from 1960-2000. To maintain consistency, we aggregate these state level estimates into census regions. In non-decennial years, we interpolate relative price levels. We normalize regional price levels in all years to the national price level figures given in Gordon (and Warren and Pearson). All income measures are reported in 2000 dollars.

Expected Years.—

The portion of the population, 25 years old and over that has completed various levels of school is given in the Census of the Population in 1940 – 2000. From this information, we calculate the expected number of years of school completed, conditional on being in either the primary, secondary, or higher educational group. The values for yrs_t^{college}, $yrs_t^{\text{secondary}}$, and yrs_t^{primary} were obtained from decennial census data. Let $N(i-j)$ be the number of people who have completed between i and j years of schooling, inclusive.

$$yrs^{\text{primary}}_{1940} = \frac{2.5N(1-4)+5.5N(5-6)+7.5N(7-8)}{N(1-4)+N(5-6)+N(7-8)} \tag{34}$$

$$yrs^{\text{primary}}_{1950,1960,1970,1980} = \frac{2.5N(1-4)+5.5N(5-6)+7N(7)+8N(8)}{N(1-4)+N(5-6)+N(7)+N(8)} \tag{35}$$

(36)

$$yrs^{\text{primary}}_{1990} = \frac{2.5N(1-4)+7.23N(5-8)}{N(1-4)+N(5-8)} \tag{37}$$

(38)

$$yrs^{\text{primary}}_{2000} = \frac{6.42N(0-8)}{N(0-8)} \tag{39}$$

(40)

$$yrs^{\text{secondary}}_{1940} = \frac{10N(9-11)+12N(12)}{N(9-11)+N(12)} \tag{41}$$

$$yrs^{\text{secondary}}_{1950,1960,1970} = \frac{10N(9-11)+12N(12)}{N(9-11)+N(12)} \tag{42}$$

(43)

$$yrs^{\text{secondary}}_{1980} = \frac{9N(9)+10N(10)+11N(11)+12N(12)}{N(9)+N(10)+N(11)+N(12)} \tag{44}$$

(45)

$$yrs^{\text{secondary}}_{1990,2000} = \frac{10.5N(9-12)+12N(12)}{N(9-12)+N(12)} \tag{46}$$

(47)

$$yrs^{\text{college}}_{1940,1950,1960} = \frac{14N(13-15)+17N(16^{+})}{N(13-15)+N(16^{+})} \tag{48}$$

$$yrs^{\text{college}}_{1970} = \frac{14N(13-15)+16N(16)+18N(17^{+})}{N(13-15)+N(16)+N(17^{+})} \tag{49}$$

(50)

$$yrs^{\text{college}}_{1980} = \frac{\begin{array}{c}13N(13)+14N(14)+15N(15)+16N(16)\\+17.5N(17-18)+20N(19^{+})\end{array}}{N(13)+N(14)+N(15)+N(16)+N(17-18)+N(19^{+})} \tag{51}$$

$$yrs^{\text{college}}_{1990} = \frac{\begin{array}{c}14N(scn+a)+16N(b)\\+18N(ma)+19.75N(pr)+20N(d)\end{array}}{N(scn)+N(a)+N(b)+N(ma)+N(pr)+N(d)} \tag{52}$$

(53)

$$yrs^{\text{college}}_{2000} = \frac{14N(sc)+14N(a)+16N(b)+18N(ma)+19.75N(prg)}{N(sc)+N(a)+N(b)+N(ma)+N(prg)} \tag{54}$$

$9-12 =$ 9th to 12th grade, no diploma

sc = some college
scn = some college no degree
a = Associate degree
b = Bachelor's degree
ma = Master's degree
prg = Professional. or Graduate degree
pr = Professional school degree
d = Doctorate degree

In 1990, data are not reported as finely for those who have completed between 5 and 8 years of schooling. We need to assign a number of years of schooling to give to the group $N(5-8)$, but this distribution is highly skewed. We calculate the conditional distribution in the years 1960, 1970, and 1980. We assign 7.23 years in 1990.

$$yrs^{5\text{-}8}_{1960} = 5.5N(5-6)_{1960} + 7N(7)_{1970} + 8N(8)_{1960} = 7.22 \quad (55)$$
$$yrs^{5\text{-}8}_{1970} = 5.5N(5-6)_{1970} + 7N(7)_{1970} + 8N(8)_{1970} = 7.23 \quad (56)$$
$$yrs^{5\text{-}8}_{1980} = 5.5N(5-6)_{1980} + 7N(7)_{1980} + 8N(8)_{1980} = 7.24 \quad (57)$$
$$yrs^{5\text{-}8}_{1990} = 7.23 \quad (58)$$

In 2000, we need to assign a number of years of schooling to give to the group $N(0-8)$, whose distribution is highly skewed. We use March 2000 CPS data for the population of people age 15 or over, which gives us data that is less aggregated than the census data. We assign 7.74 years to $N(7-8)$, which is the average value from the 1960 (7.73), 1970 (7.75), and 1980 (7.75) $yrs^{5\text{-}8}$. Thus the calculated value for $yrs^{0\text{-}8}_{2000}$ is 6.42:

$$yrs^{0\text{-}8}_{2000} = 2.5N(1-4)_{2000} + 5.5N(5-6)_{2000} + 7.74N(7-8)_{2000} = 6.42 \quad (59)$$

Values for yrs^{i}_t for periods prior to 1940 were calculated by geometrically interpolating from an initial value for the year in which the state first has adequate data available (see Table A1) to the 1940 value. Initial values are 4, 10, and 14 for primary, secondary, and higher education, respectively.

All values for non-census years between 1940 and 2000 were geometrically interpolated. We do not include those persons for whom the educational attainment level is not reported.

Description of Calculations

General Enrollment Rates.—

Enrollment figures for public and private school are summed to obtain a total primary enrollment rate, total secondary enrollment rate, and total higher educational enrollment rate. From enrollment data, enrollment rates are calculated as below:

$$tot.enroll_t^{\text{primary}} = pub.enroll_t^{\text{primary}} + pri.enroll_t^{\text{primary}} \tag{60}$$

$$tot.enroll_t^{\text{secondary}} = pub.enroll_t^{\text{secondary}} + pri.enroll_t^{\text{secondary}} \tag{61}$$

$$tot.enroll_t^{\text{college}} = pub.enroll_t^{\text{college}} + pri.enroll_t^{\text{college}} \tag{62}$$

$$r_t^{\text{primary}} = \frac{tot.enroll_t^{\text{primary}}}{\ell[5-13]_t} \tag{63}$$

$$r_t^{\text{secondary}} = \frac{tot.enroll_t^{\text{secondary}}}{\ell[14-17]_t} \tag{64}$$

$$r_t^{\text{college}} = \frac{tot.enroll_t^{\text{college}}}{\ell[18-24]_t} \tag{65}$$

General educational exposure shares.—

To calculate the stock of human capital of each type, primary school stock, secondary school stock and higher education stock, we used a perpetual inventory method. The following will illustrate the nature of our calculations. We ignore state subscripts without loss of information. In period t+1, the stock of adults, with exposure to education level i, i=primary, secondary, and higher, but no more is given by:

$$H_{t+1}^i = H_t^i(1-\delta_t^i) + I_t^i \tag{66}$$

where δ_t^i is the departure rate from the labor force and I_t^i is the flow of new adults with exposure to education level i and no more. We first illustrate the general methodology where we assume a common departure rate for all education categories. We then estimate the departure rate separately for the secondary and higher educational classes.

It is useful to put the human capital measure as a fraction of the labor force. Thus, we normalize and produce

$$\frac{H_{t+1}^i}{L_{t+1}} = \frac{H_t^i}{L_t}\frac{L_t}{L_{t+1}}(1-\delta_t) + \frac{I_t^i}{L_{t+1}} \tag{67}$$

$$h_{t+1}^i = h_t^i\frac{L_t}{L_{t+1}}(1-\delta_t) + \frac{I_t^i}{L_{t+1}} \tag{68}$$

where h_t^i measures the share of the labor force exposed to education level i, and no more in year t. The flows into education categories are given by:

$$I_t^{\text{college}} = \frac{r_t^{\text{college}} lfpr_t^{\text{college}} \ell[18-24]_t}{7}\Theta \tag{69}$$

$$I_t^{\text{secondary}} = \frac{\left(r_t^{\text{secondary}} - r_t^{\text{college}}\Theta\right) lfpr_t^{\text{secondary}} \ell[14-17]_t}{4} \tag{70}$$

$$I_t^{\text{primary}} = \frac{\left(r_t^{\text{primary}} - r_t^{\text{secondary}}\right) lfpr_t^{\text{primary}} \ell[5-13]_t}{9} \tag{71}$$

$$I_t^{\text{none}} = \frac{\left(1 - r_t^{\text{primary}}\right) lfpr_t^{\text{none}} \ell[5-13]_t}{9} \tag{72}$$

where r_t^{i} i=college, secondary and primary are the respective enrollment rates, $lfpr_t^{\text{i}}$ are the labor force participation rates for education category i, $\ell[i-j]$ is the number of people between the ages of i and j, inclusive, and Θ is the constant to adjust the inflow into the higher educational category, described below.

In order to proceed we need a measure of δ_t^i , the death rate of adults. As $L_{t+1} = L_t(1-\delta_t) + I_t^{\text{college}} + I_t^{\text{secondary}} + I_t^{\text{primary}} + I_t^{\text{none}}$, dividing through by L_{t+1} and then using definitions above, allows for the calculation of $\frac{L_t}{L_{t+1}}(1-\delta_t)$:

$$\frac{L_t}{L_{t+1}}(1-\delta_t) = 1 - \frac{\frac{r_t^{\text{college}} lfpr_t^{\text{college}} \ell[18-24]_t}{7}\Theta + \frac{\left(r_t^{\text{secondary}} - r_t^{\text{college}}\Theta\right) lfpr_t^{\text{secondary}} \ell[14-17]_t}{4} + \frac{\left(r_t^{\text{primary}} - r_t^{\text{secondary}}\right) lfpr_t^{\text{primary}} \ell[5-13]_t}{9} + \frac{\left(1 - r_t^{\text{primary}}\right) lfpr_t^{\text{none}} \ell[5-13]_t}{9}}{L_{t+1}} \tag{73}$$

With this information, we can calculate each of the shares of the labor force in each schooling category.

Using this method produced a much smaller share of the labor force exposed to higher education than the census figures. Thus we estimate the death rate of those exposed to higher education independently. We assumed that there is no death, just retirement from the labor force after 45 years of work. The stock of adults exposed to higher education is then given as:

$$H_{t+1}^{\text{college}} = H_t^{\text{college}} - I_{t-45}^{\text{college}} + I_t^{\text{college}} \tag{74}$$

$$\frac{H_{t+1}^{\text{college}}}{L_{t+1}} = \frac{H_t^{\text{college}}}{L_t}\frac{L_t}{L_{t+1}} - \frac{I_{t-45}^{\text{college}}}{L_{t-45}}\frac{L_{t-45}}{L_{t+1}} + \frac{I_t^{\text{college}}}{L_{t+1}} \tag{75}$$

Thus, to calculate the higher education share in period t, we must measure $\frac{I_{t-45}^{\text{college}}}{L_{t-45}}$, which requires higher education enrollment data in period t-45. For the earlier portion of our sample, we do not observe enrollment rates early enough to make this calculation. Where necessary, we linearly interpolate between the 0 and the value of the higher education enrollment rate the first time it is observed. See Table B.2 for the years in which each state is first calculated, and for the first time we observe higher educational enrollment figures. Unfortunately we do not observe L_{t-45} until we have 45 years of state data. We assume a constant labor force participation rate and use additional population data to calculate L_{t-45} .

There is an additional complication concerning the higher educational category. Since our calculations of the inflow to all categories are equal to the total enrollment across all ages in the category divided by the total population across all age in the category, they implicitly assume the enrollment rate is constant across ages within each education category. We are assuming that enrollment rate of 12-year olds is the same as the enrollment rate of 13-year olds, and more problematically, that the enrollment rate of 18-year olds is identical to the enrollment rate of 19-year olds.

To the extent that this assumption is erroneous, such that enrollment rates decrease with age within a category, the true the inflow in to the category will be understated. For an illustration, consider an extreme case. Suppose there are 700 students whose age distribution is uniform across ages $18-24$. Suppose that 70 of the 100 persons aged 18 are enrolled in higher education, while no one above age 18 is enrolled (a 100 percent attrition rate between age 18 and age 19). As enrollment data is reported to us aggregated across ages, the data we would observe would be a higher educational enrollment rate of 10 percent (70 enrolled students and 700 college-aged students). This would seem to imply that only 10 percent of college-aged students were being exposed to some college. In fact, in this case 70 percent of all college aged students are being exposed to some college.

While this assumption is implicit in our calculations for the inflow to all of the educational categories, it is most troublesome where there is a high attrition rate between ages. While the attrition rate between 11th and 12th grade is greater than zero, it certainly the case that the attrition rate between the first and second year of college is significant. As a result, we feel it is necessary to increase the inflow into the higher education category to address this issue, and as such we multiply the measured inflow by a factor we denote Θ.We allow Θ to vary across states, but assume it is time invariant.

We next describe the methodology to obtain the value of Θ for each state. Recall that the equations for the law of motion for the higher education category and the inflow into the higher education category are:

$$H_{t+1}^{\text{college}} = H_t^{\text{college}} - I_{t-45}^{\text{college}} + I_t^{\text{college}} \tag{76}$$

$$I_t^{\text{college}} = \frac{r_t^{\text{college}} lfpr_t^{\text{college}} \ell[18-24]_t}{7}\Theta \tag{77}$$

We observe the time path of the enrollment rate, labor force participation rate, college-aged population and the labor force. By iteratively substituting in the law of motion equation, one could solve for $h_{2000}^{\text{college}}$as a function of the initial condition h_t^{college}, and the time path of the other observables. The result would be a polynomial with order 2000 – t. Therefore, if we knew the initial and terminal condition of h_t^{college}, we could solve for the value of Θ.

The decennial censuses report the fraction of the labor force that has been exposed to higher education at the decadal frequency from 1940 to 2000, which we denote $\widetilde{h}_t^{\text{college}}$. Essentially, we proxy for the initial condition with $\widetilde{h}_{1940}^{\text{college}}$and the terminal condition with $\widetilde{h}_{2000}^{\text{college}}$and solve for Θ for each state. The value of Θ for each state is reported in Table B1.

After making the adjustments for higher educated category described above, we then utilized a common departure rate for the remaining educational categories (secondary, primary, and none). However, we found that this resulted in calculated shares exposed to elementary education that were less than zero in some states. As a result, we utilize a separate departure rate for the secondary category, $\delta^{\text{secondary}}$, and a departure rate for the remaining elementary and none categories, $\delta_t^{\text{primary}}$.

To determine the value of $\delta^{\text{secondary}}$ for each state we utilize a simple grid search. In each state we began with a value of $\delta^{\text{secondary}} = 0.0001$ and then increase the value in increments of 0.0001. For each incremental value of $\delta^{\text{secondary}}$, we use the methodology described above to calculate the time path of the fraction of the labor force exposed to secondary education, $h_t^{\text{secondary}}$. Independent data on $h_t^{\text{secondary}}$is available from decennial census reports from 1940 – 2000, providing us with 7 observations. Letting $\widetilde{h}_t^{\text{secondary}}$represent the fraction of the labor force exposed to secondary schooling reported in the census, we select the value of that minimizes:

$$\sum_t \left(h_t^{\text{secondary}} - \widetilde{h}_t^{\text{secondary}} \right), \text{where } t = 1940, 1950, 1960, 1970, 1980, 1990, 2000 \tag{78}$$

The values of are state specific, but time invariant, and are provided in Table B1.

Having selected the value of $\delta^{\text{secondary}}$, we can calculate the share of workers exposed to secondary education using the following equation:

$$h_{t+1}^{\text{secondary}} = h_t^{\text{secondary}} \frac{L_t}{L_{t+1}}(1 - \delta^{\text{secondary}}) + \frac{I_t^{\text{secondary}}}{L_{t+1}} \tag{79}$$

Given that we have calculated for h_t^{college} and $h_t^{\text{secondary}}$ in all periods, we can proceed to calculate the shares for primary and no schooling. The next set of equations shows how we can identify the term $\frac{L_t}{L_{t+1}}\left(1-\delta_t^{\text{primary}}\right)$.

$$L_{t+1} = H_{t+1}^{\text{college}} + H_{t+1}^{\text{secondary}} + H_{t+1}^{\text{primary}} + H_{t+1}^{\text{none}} \tag{80}$$

$$\begin{aligned} L_{t+1} &= H_{t+1}^{\text{college}} + H_{t+1}^{\text{secondary}} + \left(H_t^{\text{primary}} + H_t^{\text{none}}\right)\left(1-\delta_t^{\text{primary}}\right) \\ &\quad + \left(I_t^{\text{primary}} + I_t^{\text{none}}\right) \end{aligned} \tag{81}$$

$$1 - h_{t+1}^{\text{college}} - h_{t+1}^{\text{secondary}} = \left(h_t^{\text{primary}} + h_t^{\text{none}}\right)\frac{L_t}{L_{t+1}}\left(1-\delta_t^{\text{primary}}\right) + \frac{I_t^{\text{primary}} + I_t^{\text{none}}}{L_{t+1}}$$

$$\frac{L_t}{L_{t+1}}\left(1-\delta_t^{\text{primary}}\right) = \frac{1 - h_{t+1}^{\text{college}} - h_{t+1}^{\text{secondary}} - \left(\frac{I_t^{\text{primary}}}{L_{t+1}} + \frac{I_t^{\text{none}}}{L_{t+1}}\right)}{\left(h_t^{\text{primary}} + h_t^{\text{none}}\right)} \tag{82}$$

$$\frac{L_t}{L_{t+1}}\left(1-\delta_t^{\text{primary}}\right) = \frac{1 - h_{t+1}^{\text{college}} - h_{t+1}^{\text{secondary}} - \left(\begin{array}{c} \frac{\left(r_t^{\text{primary}} - r_t^{\text{secondary}}\right) lfpr_t^{\text{primary}}\,\ell[5-13]_t}{9} \\ + \frac{\left(1-r_t^{\text{primary}}\right) lfpr_t^{\text{none}}\,\ell[5-13]_t}{9} \end{array}\right)}{\left(h_t^{\text{primary}} + h_t^{\text{none}}\right)} \tag{83}$$

We occasionally measure primary and secondary enrollment rates that are larger than unity. There are a couple of reasons why this occurs. The data contains individuals that were held back in school, and also there are people that receive education for the first time starting at an unusual age. Since we have very limited information on repeaters as well as unusual starters, we treat all cases as the latter.

Initial Conditions The initial condition for h_t^i, i = college, secondary and primary were the respective enrollment rate of each class divided by two.

Higher Ed Inflow Adjustment & Secondary Departure Rates

Table B1. Values of Θ and $\delta^{\text{secondary}}$

New England	Θ	δ^{sec}	E. South Central	Θ	δ^{sec}	W. North Central	Θ	δ^{sec}
Connecticut	1.34	.9871	Alabama	1.22	.9810	Iowa	1.13	.9738
Maine	1.61	.9803	Kentucky	1.20	.9796	Kansas	1.21	.9817
Massachusetts	1.01	.9787	Mississippi	1.17	.9734	Minnesota	1.59	.9870
New Hampshire	1.83	.9999	Tennessee	1.38	.9903	Missouri	1.23	.9846
Rhode Island	0.80	.9800				Nebraska	1.18	.9776
Vermont	1.26	.9848				North Dakota	1.07	.9577
						South Dakota	1.25	.9710
Middle Atlantic			W. South Central			E. North Central		
New Jersey	1.57	.9873	Arkansas	1.41	.9782	Illinois	1.17	.9822
New York	0.97	.9743	Louisiana	1.08	.9759	Indiana	1.18	.9847
Pennsylvania	1.02	.9754	Oklahoma	1.04	.9755	Michigan	1.14	.9830
			Texas	1.60	.9980	Ohio	1.19	.9827
						Wisconsin	1.21	.9847
South Atlantic			Mountain			Pacific		
Delaware	1.31	.9931	Arizona	1.46	.9999	Alaska	1.81	.9999
D.C.	0.40	.9584	Colorado	1.79	.9999	California	1.26	.9999
Florida	1.96	.9999	Idaho	1.76	.9896	Hawaii	1.43	.9837
Georgia	2.25	.9999	Montana	1.55	.9803	Oregon	1.56	.9997
Maryland	1.50	.9999	Nevada	2.78	.9999	Washington	1.68	.9999
North Carolina	1.49	.9905	New Mexico	1.34	.9868			
South Carolina	1.57	.9883	Utah	1.41	.9986			
Virginia	1.51	.9962	Wyoming	1.33	.9829			
West Virginia	0.70	.9575						

Foreign Shares.—

In the calculation of our measure of years of schooling in state i, recall that we multiply the fraction of state i's residents that were born in state j by the years of schooling in state j (assuming no mobility):

$$E_{it} = \sum_{j \neq for} S_{ijt} \widehat{E}_{jt} \tag{84}$$

We derived our measure of $\widehat{E}_{jt}$ from observing the enrollment rates in state j and using the perpetual inventory methodology described above. Because a fraction of the residents of state i's residents are foreign born, we require a measure of $\widehat{E}_{for,t}$, the average years of schooling for the foreign born. If we could observe the share of the foreign born in each education category, we would simply calculate:

$$\widehat{E}_{for,t} = h_{for,t}^{\text{primary}} yrs_{for,t}^{\text{primary}} + h_{for,t}^{\text{secondary}} yrs_{for,t}^{\text{secondary}} + h_{for,t}^{\text{college}} yrs_{for,t}^{\text{college}} \tag{85}$$

However, this data is not available, and thus we cannot calculate the corresponding measures of $h_{for,t}^{\text{primary}}$, $h_{for,t}^{\text{secondary}}$ and $h_{for,t}^{\text{college}}$.

We use two different adjustment algorithms. We initially calculate the average years of schooling excluding the contributions made by the foreign born, which we denote $\widetilde{E}_{it}$:

$$\widetilde{E}_{it} = \sum_{j \neq for} S_{ijt} \widehat{E}_{jt} \tag{86}$$

We then assign the number of years of schooling to the foreign born $\widehat{E}_{for,t}$ so that our overall years of schooling measure, E_{it} equals the years of schooling reported by the census, yrscen_{it} :

$$\widehat{E}_{for,t} = \frac{\left(\text{yrscen}_{it} - \widetilde{E}_{it}\right)}{S_{i,for,t}} \tag{87}$$

We then place a lower and upper bound on average years of schooling assigned to foreigners by:

$$\widehat{E}_{for,t} \in \left[1, yrs_{it}^{\text{college}}\right] \tag{88}$$

We allocate the shares among the educational categories such that:

$$\widehat{E}_{for,t} = \widehat{h}_{for,t}^{\text{primary}} yrs_{it}^{\text{primary}} + \widehat{h}_{for,t}^{\text{secondary}} yrs_{it}^{\text{secondary}} + \widehat{h}_{for,t}^{\text{college}} yrs_{it}^{\text{college}} \tag{89}$$

Although there is no unique allocation, we assigned the shares using the following algorithm, in order to preserve the equality of (87):

If $\widehat{E}_{for,t} < yrs_{it}^{\text{primary}}$, we allocate between the none and primary categories, assigning zero for the secondary and college. In this case, $\widehat{E}_{for,t} = \frac{yrs_{it}^{\text{primary}}}{S_{i,for,t}}$and $\widehat{h}_{for,t}^{\text{none}} = \left(1 - \widehat{h}_{for,t}^{\text{primary}}\right)$. If $yrs_{it}^{\text{primary}} < \widehat{E}_{for,t} < yrs_{it}^{\text{secondary}}$, we assign zero for the none and college categories and allocate between the primary and secondary categories. If $yrs_{it}^{\text{secondary}} < \widehat{E}_{for,t} < yrs_{it}^{\text{college}}$, we assign zero for the none and primary categories and allocate between the secondary and college groups. If $\widehat{E}_{for,t} > yrs_{it}^{\text{college}}$, we allocate between the secondary and college categories, assigning zero for the none and primary.

California Adjustment The algorithm above assumes that the foreign born population is homogeneous. Foreign born workers are assumed to come from only adjacent educational categories. If the number of years assigned to foreigners lies between $yrs_{it}^{\text{primary}}$ and $yrs_{it}^{\text{secondary}}$, the algorithm would assign foreigners a zero share to the college and none categories. If the *actual* distribution of foreigners contains a substantial fraction of workers categorized as none and college, the algorithm would mistakenly assign these workers into the primary and secondary categories. While this is a possibility in all states, we feel this is particularly troublesome in California after 1970. In California it is quite plausible that the foreign born may be comprised of two distinct groups - a highly educated group, and a group of new migrants with low educational attainment levels. Using this algorithm for California, we would overestimate primary and secondary, but more importantly, underestimate college. This problem is further exacerbated by a growing share of the population that is foreign born. This would result in a substantial underestimation of the share exposed to college after 1970. To address this problem, we assign half of the foreign born to the college category after 1980.[30] We then allocate the remaining years to be assigned between secondary and primary. The remaining foreign born are assigned to the none category.

Idiosyncracies

DC / MD / VA.—

[30] We linearly interpolate the value of h_{it}^{college} between 1970 and 1980.

We observe extremely high private enrollment rates for District of Columbia throughout the sample, presumably due to a large number of non-residents attending the District of Columbia schools. We surmise that these enrollment figures are overstated as many residents of Maryland and Virginia are attending District of Columbia schools.

From 1910 − 1999, we assign a private elementary enrollment rate equal to zero for DC. We apportion those private elementary students enrolled in DC into the private elementary enrollment figures for Maryland and Virginia, using the population aged 5-13.

$$\begin{aligned} pri.enroll^{\text{primary}}_{Md,t} &= pri.enroll^{\text{primary}}_{Md,t} \\ &\quad + \left(\frac{\ell[5-13]_{Md,t}}{\ell[5-13]_{Va,t} + \ell[5-13]_{Md,t}} \right) pri.enroll^{\text{primary}}_{DC,t} \qquad (90) \\ pri.enroll^{\text{primary}}_{Va,t} &= pri.enroll^{\text{primary}}_{Va,t} \\ &\quad + \left(\frac{\ell[5-13]_{Va,t}}{\ell[5-13]_{Va,t} + \ell[5-13]_{Md,t}} \right) pri.enroll^{\text{primary}}_{DC,t} \qquad (91) \end{aligned}$$

We allow the private secondary enrollment rate in DC to be no higher than the private secondary enrollment rate in the state of Massachusetts. We first calculate the enrollment rate in excess of the enrollment rate in DC, and then calculate the implied excess enrollment (students). We then apportion the excess enrollment into MD and VA, weighted by the population aged 14-17 in each state.

$$pri.enroll^{\text{secondary}}_{DC,t} = pri.r^{\text{secondary}}_{Ma,t}\, \ell[14-17]_{DC,t} \qquad (92)$$

$$\begin{aligned} pri.enroll^{\text{secondary}}_{Md,t} &= pri.enroll^{\text{secondary}}_{Md,t} \\ &\quad + \left(\frac{\ell[14-17]_{Md,t} \cdot \left(pri.r^{\text{secondary}}_{DC,t} - pri.r^{\text{secondary}}_{Ma,t} \right)}{\ell[14-17]_{Va,t} + \ell[14-17]_{Md,t}} \right) \ell[14-17]_{DC,t} \qquad (93) \\ pri.enroll^{\text{secondary}}_{Va,t} &= pri.enroll^{\text{secondary}}_{Va,t} \\ &\quad + \left(\frac{\ell[14-17]_{Va,t} \cdot \left(pri.r^{\text{secondary}}_{DC,t} - pri.r^{\text{secondary}}_{Ma,t} \right)}{\ell[14-17]_{Va,t} + \ell[14-17]_{Md,t}} \right) \ell[14-17]_{DC,t} \qquad (94) \end{aligned}$$

AK / HA.—
yrs^{college}_t, $yrs^{\text{secondary}}_t$, and yrs^{primary}_t for Alaska in 1939 and for Hawaii in 1940 were set as 14.5, 10.5, and 5.5 respectively.

ND / SD/ Dakota.—
From 1880 through 1890, population and enrollment figures are reported for Dakota, which is the aggregate of North Dakota and South Dakota. In 1890, we first observe separate figures for North Dakota and South Dakota. Where data is available, we allocate a constant fraction of Dakota population and enrollment figures to each of North and South Dakota, based on the population of each state in 1890.

Indian Territory / Oklahoma.—

We first include Oklahoma in our data set only after the *Statistical Abstract* reported data for Oklahoma, rather than Indian Territory.

Table B2: List of first year we observe enrollment data, and first year we observe higher education enrollment data.

State	1^{st} year of obs.	1^{st} year of higher ed.	State	1^{st} year of obs.	1^{st} year of higher ed.
Alabama	1840	1840	Montana	1870	1870
Alaska	1939	1924	Nebraska	1860	1870
Arizona	1872	1899	Nevada	1870	1886
Arkansas	1840	1850	New Hampshire	1840	1840
California	1850	1860	New Jersey	1840	1840
Colorado	1870	1870	New York	1840	1840
Delaware	1840	1840	North Carolina	1840	1840
D.C.	1850	1850	North Dakota	1890	1890
Florida	1840	1870	Ohio	1840	1840
Georgia	1840	1840	Oklahoma	1890	1899
Hawaii	1940	1922	Oregon	1850	1860
Idaho	1870	1899	Pennsylvania	1840	1840
Illinois	1840	1840	Rhode Island	1840	1840
Indiana	1840	1840	South Carolina	1840	1840
Iowa	1840	1850	South Dakota	1890	1890
Kansas	1860	1860	Tennessee	1840	1840
Kentucky	1840	1840	Texas	1850	1850
Louisiana	1840	1840	Utah	1860	1870
Maine	1840	1840	Vermont	1840	1840
Maryland	1840	1840	Virginia	1840	1840
Massachusetts	1840	1840	Washington	1860	1870
Michigan	1840	1840	West Virginia	1870	1870
Minnesota	1860	1860	Wisconsin	1850	1850
Mississippi	1840	1840	Wyoming	1870	1890
Missouri	1840	1840			

APPENDIX C

To analyze the return to schooling, we need information on the income per worker. Since 1929, the Bureau of Economic Analysis has reported state level annual income data. Total and per capita state income for 1840, 1880, 1900 and 1919-1921 are documented by Richard Easterlin in his works,"Interregional Differences in Per Capita Income, Population, and Total Income 1840-1950" in *Trends in the American Economy in the Nineteenth Century* and *Analyses of Economic Change in Population Redistribution and Economic Growth, United States, 1870-1950.* These data exclude transfer payments, likely small during this time period, and the figures for 1840 do not include all components of personal income. For the Census years not reported by Easterlin, 1850, 1860, 1870, 1890, and 1910, we generate the missing state per capita income using data available from the Easterlin sources above, the 1850 through 1910 Censuses, and the *Historical Statistics of the United States: Colonial Times to 1970* (HSUS). In order to calculate state per worker income, we calculate value added by each industry at the state level. Although data is not available for every industry, production value is reported for agriculture in the Census from 1870 through 1910 and production value and materials are reported in the Census from 1850 through 1910 for manufacturing.

Agricultural Production Value

From 1870 to 1910, each Census reports the value of agricultural products at the state level, Y_{it}^{ag}. We would prefer explicit data on agricultural value added rather than agricultural products. However, in the only year of overlapping values, 1880, the Census numbers match the agricultural income reported by Easterlin in *Trends in the American Economy in the Nineteenth Century.* To determine the state values of agricultural production for 1850, and 1860, we estimate the relationship of the production value of agricultural products sold within a state on the total value of farmland and buildings and agricultural labor force.

Agricultural labor force is reported in the Census in 1840, 1850, and 1870 through 2000. While the census does report a measure of the agricultural labor force in 1850, it usefulness is diminished because it does not include slave labor.[31] To estimate the total agricultural labor force for 1850 and 1860, we use the agricultural labor force reported in 1840, which includes slaves, and in 1870, which includes freed slaves, to construct the portion of the state labor force engaged in agricultural production, $\text{fraction}_{it}^{ag}$. In non-slave holding regions, where the omission of slave labor is not problematic, we calculate $\text{fraction}_{it}^{ag}$ in 1850 using the Census data.[32] We then linearly interpolate $\text{fraction}_{it}^{ag}$ between 1840 and 1870 (between 1850 and 1870 for slave-holding regions and New England). We complete our measure of agricultural labor force in these intervening years by multiplying $\text{fraction}_{it}^{ag}$ by the total labor force in each state.[33]

For the 1850 and 1860 values of agricultural products, we estimate the relationship in 1870 and 1880. For 1920, we estimate the relationship in 1910 and 1930.[34] The Census reports the production value of agricultural products and data on total farmland value comes from HSUS. With our measures of agricultural capital, farmvalue_{it}, and labor, aglabor_{it}, we estimate the value of products produced in 1850, 1860, and 1920 by regressing the following:

$$\ln(Y_{it}^{ag}) = \beta_1 \ln(farmvalue_{it}) + \beta_2 \ln(aglabor_{it}) + \beta_3 Z \tag{95}$$

where Z is the vector of region dummies and year_t is a time trend. We then take the exponential of the predicted value, $\widehat{Y_{it}^{ag}}$, to estimate state level agricultural production value for 1850, 1860, and 1920. Results of these regressions are reported in Table C1 below.

Table C1: Regressions of Natural Log Agricultural Production

[31] The 1860 census reports data hundreds of detailed occupations, but we do not attempt to map these occupations into the broader agricultural labor force.

[32] These regions are the Middle Atlantic, Mountain, Pacific, East North Central, and West North Central regions. We do not include the New England region because data in 1850 appear unreliable.

[33] No data on agricultural labor force is reported for Kansas, Nebraska, Texas, and Washington in 1840, therefore, we are unable to calculate the fraction of the labor force in agriculture using the methodology described above. For 1860, we proxy the agricultural labor force for these states by the number of persons listing their occupation as farmers.

[34] Additionally, data on agricultural products is not available in Arizona and New Mexico in 1890. We again regress using Eq. 94 and use data from 1880 and 1900 to estimate values for these two states.

variable	coefficient	std.error	coefficient	std. error
ln(farmvalue)	0.290	0.061	0.892	0.080
ln(aglabor)	0.574	0.069	0.131	0.082
NE	5.304	0.751	-0.817	0.991
MA	5.662	0.837	-0.946	1.079
SA	5.169	0.742	-0.895	1.023
ESC	5.407	0.771	-0.748	1.044
WSC	5.525	0.745	-0.876	1.067
MTN	5.129	0.623	-1.019	1.015
WNC	5.471	0.780	-1.212	1.127
ENC	5.581	0.830	-1.242	1.120
PAC	5.514	0.731	-1.194	1.103
N			96	
$\overline{R}^2$	0.9997		0.9997	
data used	1870, 1880		1910, 1930	
predict	1850,1860		1920	

Manufacturing Value Added

The value added by manufacturers at the state level, Y_{it}^{manu}, is calculated by subtracting the value of materials used from the value of products sold reported in the Census from 1850 through 1920. Because the 1840 Census does not report the value added by manufacturing, we use the relationship between value added and the manufacturing labor force from 1850 through 1860 to determine value added in 1840. We regress the natural log of value added in the manufacturing sector, mvalue_{it}, on the natural log of the manufacturing labor force, mlabor_{it}, interacted with regions as well as individual census region effects, Z:[35]

$$\ln\left(mvalue_{it}\right) = \beta_1 Z + \beta_2 \left(Z \ln\left(mlabor_{it}\right)\right) + \beta_3 year_t \tag{96}$$

Taking the exponential of the predicted $ln\left(\widehat{mvalue_{it}}\right)$ generates the 1840 estimate of value added by manufacturing.

Mining Value Added

The output of precious metals is an important component of state income in the Pacific and Mountain region, particularly so in the early portion of out data set. As will be discussed in the following section, our income calculations allow for a component of income not captured by agriculture and mining. However, our methodology implicitly assumes that this component is relatively stable over time. Given the nature of gold and silver discoveries and subsequent rushes, we find this assumption unsatisfactory for these regions. As a result, we have collected data on precious metals mining output for the Mountain and Pacific regions.

[35] Data on manufacturing labor are not available in 1890 and 1910. We calculate the fraction of the labor force engaged in manufacturing, $\text{fraction}_{it}^{min}$ in 1880, 1900, and 1920. We linearly interpolate the value of $\text{fraction}_{it}^{min}$ in 1890 and 1910, and multiply the result by the total labor force.

Value added in the precious metals mining sector of the economy is calculated by subtracting the value of materials from the value of mining products, product_value$_{it}$, where available. A measure of mining products is available at the state level from the 1890 Census Report on Mineral Industries in the United States for 1870, 1880, and 1890.[36] A measure of materials used and labor is also available. This allows a measure of mining value added in 1890 , $Y^{mn}_{i,1890}$, to be calculated.

$$Y^{mn}_{i,1890} = product\_value_{it} - materials_{it} \tag{97}$$

We next calculate per worker value added in 1890:

$$y^{mn}_{i,1890} = \frac{Y^{mn}_{i,1890}}{L^{mn}_{i,1890}} \tag{98}$$

and fraction of output this is value added, $fracY_{i,1890}$:

$$fracY_{i,1890} = \frac{Y_{i,1890}}{product\_value_{i,1890}} \tag{99}$$

The 1870 Census report, The Statistics of Mining, gives data on employment, materials, and output of precious metals in 1870, but appears to be only a partial sample of all mining establishments. We do not use the measures of total products, value added and employment, but maintain measures of *per worker* products, value added, and employment.[37] Thus, we calculate $y^{mn}_{i,1870}$ and $fracY_{i,1870}$ and then use these values with the 1890 values to interpolate to obtain $y^{mn}_{i,1880}$ and $fracY_{i,1880}$. Prior to 1870, data is not as detailed. We assume that products per worker for each state in 1850 and 1860 is equal to it's value in 1870.[38] Thus:

$$y^{mn}_{i,1850} = y^{mn}_{i,1860} = y^{mn}_{i,1870} \tag{100}$$

We do the same for the fraction of products that is value added.

$$fracY^{mn}_{i,1850} = fracY^{mn}_{i,1860} = fracY^{mn}_{i,1870} \tag{101}$$

We next turn out attention to employment in precious metals mining. Direct measures of precious metals mining employment are available in 1840, and 1890 (and in 1870 we have a sample), as are measures of non-precious metal mining employment. This overlapping data will be exploited below. Data on precious metals employment data do not exist directly in 1850, 1860, and 1880, yet measures of total employment in mining (precious and non-precious) are available in these years.

Let employment in precious metals mining be L^{prec}_{it}, and employment in non-precious metals mining, $L^{nonprec}_{it}$. In 1840, 1870, and 1890 we calculate:

$$fracL^{prec}_{it} = \frac{L^{prec}_{it}}{(L^{prec}_{it} + L^{nonprec}_{it})} \tag{102}$$

For states in which we have no data prior to 1870, we assume that $fracL^{prec}_{it}$ in 1850 and 1860 are identical to the 1870 values in each state. We also interpolate between 1870 and 1890 to acquire 1880 values. Thus:

[36] Data is not readily available from this source for 1890. Instead, we use the values in 1889

[37] In addition, we maintain the fraction of all mining labor that is engaged in precious metals mining. See below.

[38] There is only one state, California, for which we have data in 1850. We make a separate adjustment for this state below.

$$fracL_{i,1850}^{prec} = fracL_{i,1860}^{prec} = fracL_{i,1870}^{prec} \quad (103)$$

Next, we calculate labor in the precious metal sector, L_{it}^{prec}, in 1850, 1860, and 1880 as,

$$L_{it}^{prec} = fracL_{it}^{prec}\left(fracL_{it}^{prec\&nonprec}\right) \quad (104)$$

And to correct for the fact that L_{it}^{prec} in 1870 is a sample, we geometrically interpolate between the value of L_{it}^{prec} in 1860 and 1880.

Finally, we can calculate our measure of Y_{it}^{mn} for 1850, 1860, 1870, and 1880:

$$Y_{it}^{mn} = y_{it}^{mn} L_{it}^{mn} fracL_{it}^{prec} \quad (105)$$

As a check on the reasonableness of our calculations, we compare the sum of mining output across the states to the national output figures given for 1850 and 1860 in the 1890 Census report. We find we overestimate mining output in 1860. We assume that California has the same share of national mining output in 1860 as it does in 1850. We then renormalize all other states so that the sum is equal to the national total.

Total State Income

Adding the value of products produced by manufacturers and mines and the estimated income from agricultural production at the state level generates the total state income attributable to manufacturing, mining, and agriculture:

$$Y_{it}^{ag+manu+mn} = Y_{it}^{ag} + Y_{it}^{manu} + Y_{it}^{mn} \quad (106)$$

for $1840 \leq t \leq 1920$.[39]

Unfortunately for us, this measure of income is not the total state income, but only the of portion of state income resulting from manufacturing, mining, and agriculture. In order to account for the remaining industries in a states' economy, we turn to the total income calculations reported by Easterlin. In *Trends in the American Economy in the Nineteenth Century*, Easterlin calculates the total state income level for 1840 and in *Analyses of Economic Change in Population Redistribution and Economic Growth, United States, 1870-1950,* he reports total state income for 1880, 1900, and 1919-1921(1920). For 1840, 1880, 1900, and 1920, we calculate the difference between our estimated, $Y_{it}^{ag+manu+mn}$, and Easterlin's total state income, Y_{it}^{E}:

$$Y_{it}^{not} = Y_{it}^{E} - Y_{it}^{ag+manu+mn} \quad (107)$$

for t=1840, 1880, 1900, and 1920. We then calculate the ratio of income generated outside agriculture, manufacturing, and mining over income produced by agriculture, manufacturing, and mining:[40]

$$Y_{it}^{notshare} = \frac{Y_{it}^{not}}{Y_{it}^{ag+manu+mn}} \quad (108)$$

For the states with 1840 Easterlin incomes, listed in Table C1, we estimate the ratio of income

[39] We only make our mining adjustments in 1850, 1860, 1870, and 1890 for the Mountain and Pacific regions. We do not adjust mining for states outside of these regions. That is, $Y_{it}^{mn} = 0$ for all other regions.

[40] We occasionally observe a measure of Y_{it}^{not} that is less than zero in 1840. For these states, the sum of agricultural, mining, and manufacturing income exceeds the figure given as total income by Easterlin. We replace the measure of Y_{it}^{not} with zero. Cases are rare and magnitudes are small.

generated outside agriculture, manufacturing, and mining over income produced by agriculture, manufacturing, and mining for 1850, 1860, 1870, 1890, and 1910 using the following methods:

$$\widehat{Y}_{i,1850}^{notshare} = \left(Y_{i,1840}^{notshare}\right)^{.75}\left(Y_{i,1880}^{notshare}\right)^{.25} \quad (109)$$
$$\widehat{Y}_{i,1860}^{notshare} = \left(Y_{i,1840}^{notshare}\right)^{.5}\left(Y_{i,1880}^{notshare}\right)^{.5} \quad (110)$$
$$\widehat{Y}_{i,1870}^{notshare} = \left(Y_{i,1840}^{notshare}\right)^{.25}\left(Y_{i,1880}^{notshare}\right)^{.75} \quad (111)$$
$$\widehat{Y}_{i,1890}^{notshare} = \left(Y_{i,1880}^{notshare}\right)^{.5}\left(Y_{i,1900}^{notshare}\right)^{.5} \quad (112)$$
$$\widehat{Y}_{i,1910}^{notshare} = \left(Y_{i,1900}^{notshare}\right)^{.5}\left(Y_{i,1920}^{notshare}\right)^{.5} \quad (113)$$

For the states without 1840 incomes, listed in Table C2, we use the 1880 ratio of income generated outside agriculture, manufacturing, and mining over income produced by agriculture, manufacturing, and mining, $Y_{i,1880}^{notshare}$, in order to determine $Y_{i,t}^{notshare}$, for t =1850, 1860, 1870. For 1890, and 1910 we use the similar method as above:

$$\widehat{Y}_{i,1850}^{notshare} = \left(Y_{i,1880}^{notshare}\right) \quad (114)$$
$$\widehat{Y}_{i,1860}^{notshare} = \left(Y_{i,1880}^{notshare}\right) \quad (115)$$
$$\widehat{Y}_{i,1870}^{notshare} = \left(Y_{i,1880}^{notshare}\right) \quad (116)$$
$$\widehat{Y}_{i,1890}^{notshare} = \left(Y_{i,1880}^{notshare}\right)^{.5}\left(Y_{i,1900}^{notshare}\right)^{.5} \quad (117)$$
$$\widehat{Y}_{i,1910}^{notshare} = \left(Y_{i,1900}^{notshare}\right)^{.5}\left(Y_{i,1920}^{notshare}\right)^{.5} \quad (118)$$

Using these ratios we calculate our final total state income, $\widehat{Y_{it}^{all}}$, for all non-Easterlin years:

$$\widehat{Y}_{it}^{all} = Y_{it}^{ag+manu+mn}\left[1+\widehat{Y}_{i,t}^{notshare}\right] \quad (119)$$

In order of find our calculated per worker income, we simple take total state income in year and divide it by the states' labor force reported by the census, except 1850 and 1860 where the our labor force figures are adjusted for slaves:

$$y_{it} = \frac{\widehat{Y}_{it}^{all}}{L_{it}} \quad (120)$$

We then put our per worker income measures into real terms by adjusting for both national and regional differences in prices. See Appendix B for more details on price levels.

Table C1: 1840 State Incomes Reported By Easterlin

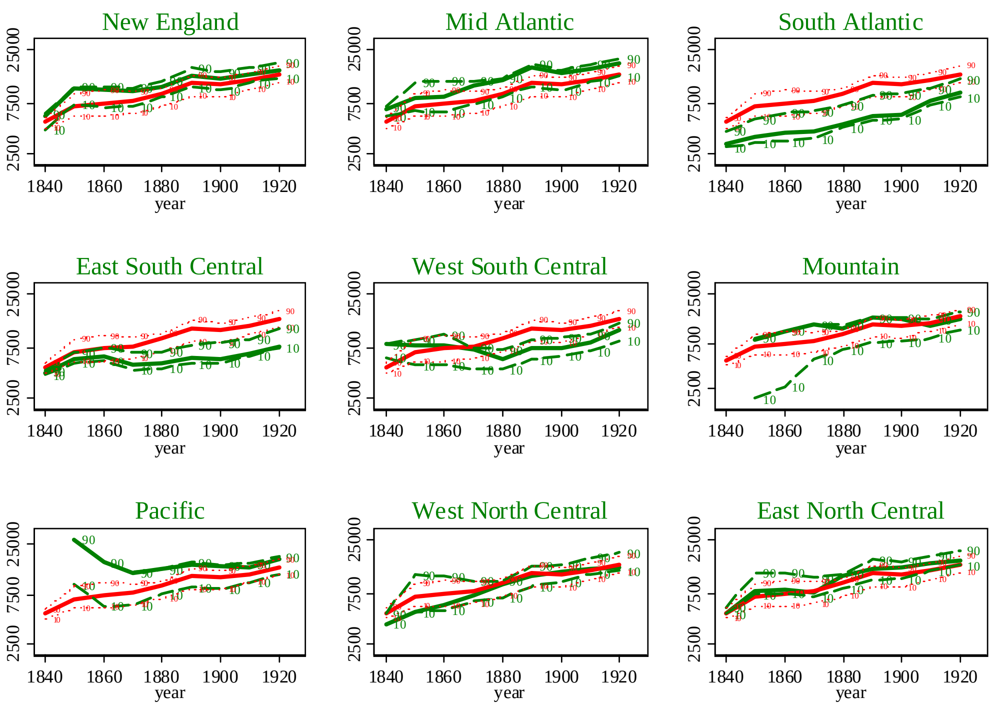
Alabama	Iowa	Mississippi	Pennsylvania
Arkansas	Kentucky	Missouri	Rhode Island
Connecticut	Louisiana	New Hampshire	South Carolina
Delaware	Maine	New Jersey	Tennessee
Florida	Maryland	New York	Vermont
Georgia	Massachusetts	North Carolina	Virginia
Illinois	Michigan	Ohio	Wisconsin
Indiana			

Table C2: 1840 State Incomes Not Reported By Easterlin
(with first year of agriculture and manufacturing data availability)

State	First Year Calculated	State	First Year Calculated
Arizona	1870	New Mexico	1850
California	1850	Oregon	1850
Colorado	1870	South Dakota	1910
Idaho	1870	Texas	1850
Kansas	1860	Utah	1850
Minnesota	1860	Washington	1860
Montana	1870	West Virginia	1870
Nebraska	1860	Wyoming	1870
Nevada	1870		

INCOME BOUNDS

In this section we present income per worker bounds for the period 1840-1920. Since we imputed non agricultural, non manufacturing non mining output per worker, we provide bounds on our estimates in this section. Our procedure takes the 10th (90th) percentile values of non agricultural, non manufacturing output per worker for the country as a whole and replaces our estimate of each state's non agricultural, non manufacturing output per worker with the min (max) of our estimate and the 10th (90th) percentile values. The results do not change substantively if instead we use the census regions or the North, South and West region values instead. The figure below presents the results of this exercise. In the panel graph each panel presents our estimates of the region's output per worker, and the two bounds as well as the US values for each. The census region figures are always in green and the US values are always in red.



The following Table presents our regional estimates as well as the percent deviation between the estimates and the two bounds for each census year.

Table C3: National and Regional Income and Bounds (10th, 90th) percentiles

region	1840	1850	1860	1870	1880	1890
US	4950	7034	7490	7939	9448	11821
	(.934,1.055)	(.895,1.164)	(.884, 1.175)	(.869, 1.235)	(.931, 1.061)	(.873, 1.105)
NE	5640	10668	10216	9832	10998	13810
	(.770, 1.082)	(.706, 1.776)	(.788, 1.151)	(.827, 1.083)	(.899, 1.058)	(.816, 1.101)
MATL	6709	8360	8952	11112	12954	16743
	(.987, 1.014)	(.961, 1.024)	(.970, 1.030)	(.917, 1.089)	(.964, 1.023)	(.916, 1.075)
SATL	3089	3581	3882	4106	4751	5724
	(.936, 1.178)	(.890, 1.187)	(.859, 1.096)	(.798, 1.292)	(.854, 1.170)	(.851, 1.193)
ESC	4391	5926	6442	5316	5447	6191
	(.967, 1.053)	(.941, 1.045)	(.908, 1.053)	(.905, 1.037)	(.938, 1.094)	(.927, 1.080)
WSC	8363	8019	8209	7392	5971	7503
	(.787, 1.011)	(.742, 1.134)	(.797, 1.254)	(.719, 1.104)	(.935, 1.017)	(.954, 1.040)
MTN	-	8261	10236	11889	10913	14007
		(.988, 1.077)	(.957, 1.015)	(.490, 2.779)	(.579, 1.560)	(.766, 1.282)
PAC	-	27446	16167	12913	13787	15294
		(1.000, 1.040)	(.695, 1.050)	(.824, 1.020)	(.891, 1.011)	(.783, 1.090)
WNC	3825	5013	5945	7247	9248	11222
	(1.000, 1.000)	(.999, 1.005)	(.870, 1.477)	(.864, 1.367)	(.960, 1.029)	(.817, 1.160)
ENC	4867	7941	8265	7668	11147	13132
	(.999, 1.004)	(.947, 1.000)	(.915, 1.456)	(.900, 1.516)	(.945, 1.059)	(.882, 1.096)

Table C3 (continued) : National and Regional Income and Bounds (10th, 90th) percentiles

region	1900	1910	1920	average
US	11477	12531	14430	
	(.891, 1.080)	(.877, 1.092)	(.875, 1.095)	(.892, 1.118)
NE	13073	14236	15706	
	(.849, 1.074)	(.900, 1.058)	(.891, 1.047)	(.827, 1.159)
MATL	14947	16244	18469	
	(.933, 1.055)	(.881, 1.100)	(.892, 1.100)	(.936, 1.057)
SATL	5929	7932	9770	
	(.770, 1.279)	(.810, 1.141)	(.860, 1.172)	(.847, 1.190)
ESC	5900	6758	7947	
	(.861, 1.089)	(.850, 1.114)	(.780, 1.118)	(.898, 1.076)
WSC	7641	8573	11512	
	(.845, 1.083)	(.946, 1.039)	(.883, 1.095)	(.845, 1.086)
MTN	13838	11621	13823	
	(.868, 1.086)	(.829, 1.150)	(.863, 1.079)	(.792, 1.378)
PAC	14992	14175	17607	
	(.866, 1.045)	(.878, 1.042)	(.878, 1.051)	(.852, 1.044)
WNC	12395	12986	13497	
	(.943, 1.071)	(.804, 1.131)	(.758, 1.153)	(.891, 1.155)
ENC	13440	14705	15841	
	(.885, 1.063)	(.921, 1.067)	(.926, 1.062)	(.924, 1.147)

These bounds are constructed using the following method. First we calculate the total income produced by workers not in agriculture, manufacturing, mining, $\widehat{Y}_{it}^{not}$, by state, year:

$$\widehat{Y}_{it}^{not} = \widehat{Y}_{it}^{all} - Y_{it}^{ag+man+mn} \tag{121}$$

We then calculate the per worker income for workers not in agriculture, manufacturing, mining, $\widehat{y}_{it}^{not}$, by dividing the total income produced by workers not employed in agriculture, manufacturing, mining by the number of workers in these other industries:

$$\widehat{y}_{it}^{not} = \frac{\widehat{Y}_{it}^{not}}{L_{it}^{not}} \tag{122}$$

In order to compare not per worker incomes across states, we deflate our nominal measures using the same deflator constructed in Appendix B (Price Levels), to create the real not per worker income, $\widetilde{y}_{it}^{not}$. For each state, we generate two real income per worker bound series: one by replacing a state's own real not per worker income with the national 10th percentile real not per worker income, multiplying by the number of workers in the not sector, adding this result to the total income from agriculture, mining, and manufacturing, and dividing by the total number of workers in the state:

$$\widetilde{y}_{it}^{10th} = \frac{\widetilde{y}_{it}^{10th,not} L_{it}^{not} + \widetilde{y}_{it}^{ag+man+mn} L_{it}^{ag+man+mn}}{L_{it}^{not} + L_{it}^{ag+man+mn}} \tag{123}$$

and the other by replacing a state's own real not per worker income with the national 90th percentile real per worker income for employees in the non-sector and repeating the similar process as above:

$$\widetilde{y}_{it}^{90th} = \frac{\widetilde{y}_{it}^{90th,not} L_{it}^{not} + \widetilde{y}_{it}^{ag+man+mn} L_{it}^{ag+man+mn}}{L_{it}^{not} + L_{it}^{ag+man+mn}} \tag{124}$$

For the 90th percentile, if a states actual not per worker income is higher; we simply used the states own. For the 10th percentile: if a state's real not per worker income was lower, we simply used the states own, so that for both series a state's overall real per worker income always lay on or between the constructed 90th and 10th percentile real income per worker bounds.

APPENDIX D

Table D1 below presents the correlations of our years of schooling in the labor force with the two separate state human capital measures of Mulligan and Sala-i-Martin (1997,2000).

D1: Correlation of Years of Schooling in the Labor Force with Mulligan and Sala-i-Martin (1997, 2000)

1940	yrs of schooling	hc1997	hc2000
yrs of schooling	1		
hc1997	.9326	1	
hc2000	.8996	.9747	1
1950	yrs of schooling	hc1997	hc2000
yrs of schooling	1		
hc1997	.8824	1	
hc2000	.8081	.9321	1
1960	yrs of schooling	hc1997	hc2000
yrs of schooling	1		
hc1997	.7766	1	
hc2000	.7955	.9500	1
1970	yrs of schooling	hc1997	hc2000
yrs of schooling	1		
hc1997	.6455	1	
hc2000	.6727	.8403	1
1980	yrs of schooling	hc1997	hc2000
yrs of schooling	1		
hc1997	.8466	1	
hc2000	.7669	.8792	1
1990	yrs of schooling	hc1997	hc2000
yrs of schooling	1		
hc1997	.8449	1	
hc2000	.7797	.9141	1

Table D2 below details how well we fit the census information using labor force weighted regressions.[41] Overall, the our calculations fit the data extremely well, but this could be due to the trend in education. Thus we present the decade by decade results. If our estimates were exactly in line with the census, we would get a slope coefficient of 1 on years and a 0 intercept.

Table D2: Regressions of Average Years of Schooling from the Census on Estimates (standard errors)

variable	ALL	1940	1950	1960	1970	1980	1990	2000
E	1.053	1.028	1.147	1.185	1.161	0.948	0.8101	0.855
	(0.009)	(0.002)	(0.002)	(0.007)	(0.007)	(0.003)	(0.003)	(0.003)
constant	-0.565	-0.355	-1.449	-1.833	-1.835	0.738	2.416	1.961
	(0.103)	(0.145)	(0.193)	(0.693)	(0.830)	(0.462)	(0.505)	(0.461)
N	355	49	51	51	51	51	51	51
$\overline{R}^2$	.9723	.9127	.9183	.7963	.7889	.8413	.7982	.8425
$prob > F$	.0000	.0013	.0001	.0973	.0034	.0001	.0002	.0000

The final row of the table contains the result of the joint test of this hypothesis. Overall we reject the null hypothesis that our estimated slope coefficient is 1 and our intercept is 0, however for 1960 we cannot reject the null. In all regressions, our fit is quite good, with $\overline{R}^2$ over .75.

[41] This seems reasonable as it seems much more important to fit New York or California than to give those states equal weight with states like North and South Dakota.

An alternative way to compare our estimates of years of schooling in the labor force with the values of years of schooling by state from the Census is to compare the means and standard deviations weighted and unweighted. Table D3 provides evidence that our estimates are similar, if not identical with the census values.

Table D3: Average Years of Schooling: Census and New Estimates

year	Census mean	Census std. dev.	Estimate mean	Estimate std. dev.	% dev. mean	Census weighted mean	Estimate weighted mean	% dev. weighted mean
1940	8.24	0.89	8.34	0.86	1.2	8.17	8.29	1.3
1950	8.98	0.87	9.07	0.74	1.0	8.95	9.07	1.3
1960	9.85	0.75	9.85	0.58	0.0	9.82	9.83	0.1
1970	10.68	0.64	10.85	0.49	1.6	10.65	10.75	0.9
1980	11.87	0.59	11.65	0.51	-1.9	11.82	11.68	-1.2
1990	12.45	0.42	12.37	0.43	-0.6	12.43	12.36	-0.6
2000	13.14	0.38	13.04	0.41	-0.8	13.08	13.01	-0.5

Table D3 shows that our average years of schooling measure nearly match Census estimates. The largest weighted difference occurs in 1940 and 1950, while the largest unweighted difference occurs in 1980. The smallest difference occurs in 1960 for both measures. From 1940 onward the mean of our estimates differs from the Census by less than 1.9 percent. One thing evident from Table D3 is the greater amount of dispersion about the mean in our estimates from 1980 to 2000, but smaller dispersion than the Census estimates before 1980.

LABOR'S SHARE OF INCOME

Table 4: Labor Share and Capital Share of Income[42]

[42] Lines (1)-(12) Table reprinted from Table 15, *National Income: A Summary of Findings*, Kuznets, NBER (1946), p. 50.
Lines (13)-(25) from Table 4, Denison, *The Sources of Economic Growth in the United States and the Alternatives Before US*, Committee for Economic Development (1962) p. 30.

line	period	*Emp. Comp.* (1)	*Entrep. Net Inc.* (2)	(1) + (2)	*Div.* (3)	*Int.* (4)	*Rent* (5)	(3) + (4) + (5)
1	1870-1880	50.0	26.4	76.5	15.8		7.8	23.6
2	1880-1890	52.5	23.0	75.4	16.5		8.2	24.6
3	1890-1900	50.4	27.3	77.7	14.7		7.7	22.4
4	1900-1910	47.1	28.8	75.8	15.9		8.3	24.2
5	1899-1908	59.5	23.8	83.3	5.3	5.1	6.4	16.7
6	1904-1913	59.6	23.3	82.9	5.7	5.1	6.3	17.1
7	1909-1918	59.7	23.3	83.0	6.5	4.9	5.7	17.0
8	1914-1923	63.0	20.8	83.8	5.6	5.3	5.3	16.2
9	1919-1928	65.1	18.3	83.4	5.4	6.0	5.2	16.6
10	1919-1928	61.7	19.5	81.2	5.6	6.1	7.1	18.8
11	1924-1933	63.1	16.6	79.7	6.5	7.8	5.9	20.3
12	1929-1938	64.9	15.9	80.8	6.6	8.4	4.3	19.2
13	1909-1913			69.5				30.5
14	1914-1918			67.0				33.0
15	1919-1923			69.5				30.5
16	1924-1928			69.7				30.3
17	1929-1933			69.2				30.8
18	1934-1938			70.4				29.6
19	1939-1943			72.1				27.9
20	1944-1948			74.9				25.1
21	1949-1953			74.5				25.5
22	1954-1958			77.3				22.7
23	1909-1958			71.4				28.6
24	1909-1929			68.9				31.1
25	1929-1958			73.0				27.0