How Important are Human Capital, Physical Capital and Total Factor Productivity for Determining Economic Growth in the United States, 1840 - 2000

Scott Baier, Sean Mulholland, Robert Tamura, and Chad Turner[1]

November 2005

Abstract

This paper presents new data on physical capital at the state level for the United States from 1840 - 2000. After combining this new data with recently created state level human capital and income data, we are able to use standard growth accounting techniques to estimate the contribution of aggregate input growth and total factor productivity (TFP) growth on income growth across the states of the United States. In existing growth accounting literature, limitations on data availability typically confine this type of analysis to cross-country comparisons for relatively short periods of time. TFP, or the residual component, is found to be large relative to the portion of output growth that is explained by measured aggregate input growth. Often the explanation for this result appeals to institutional heterogeneity across countries. The data used in this paper includes measures of human capital, physical capital and output growth from 1840 – 2000, a period that is longer than typical for the existing literature. In addition, as our data is across states of the United States instead of across countries, one would expect less institutional heterogeneity in this study than in others in the literature, and therefore expect a larger fraction of income growth to be explained with measured input growth. We find that that 64% of output growth from 1840 – 2000 is accounted for by input growth. We conduct a variance decomposition to examine the role that aggregate input growth and TFP growth have in explaining the cross-sectional variance of income growth across states. As the results are sensitive to the treatment of the observed correlation between aggregate input growth and TFP growth, we construct an upper and lower bound. We find the variance of TFP can explain between 48% and 88% of the variance of income growth, leaving between 12% and 52% to be explained by aggregate input growth. However, these results are sensitive to initial conditions.

[1] Baier and Tamura are affiliated with Clemson University, Mulholland with Moravian College, and Turner with Nicholls State University. Baier and Tamura are also affiliated with the Atlanta Federal Reserve Bank. All views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

HOW IMPORTANT ARE HUMAN CAPITAL, PHYSICAL CAPITAL AND TOTAL FACTOR PRODUCTIVITY FOR DETERMINING ECONOMIC GROWTH IN THE UNITED STATES, 1840 – 2000

Why are some states richer than others? What are the sources of economic growth? In the mid-twentieth century, economist such as Solow (1956, 1957) and Kuznets (1966) began formulating theories for why economies grow. Their ideas created the neoclassical paradigm in which capital and technology are the driving forces of economic growth. The resulting growth rate of income could therefore be accounted for by changes in capital, labor, and technology. Research by Denison (1969) however, found that little of growth could be accounted for by inputs, leaving the overwhelming majority to be explained by technology. Theoretical research by Romer (1986) and Lucas (1988), combined with data on income and investment provided by Summers-Heston (1988, 1991) gave researchers new means with which to empirically test theoretical relationships between labor, capital, technology, and income growth across countries. These data were later expanded to include education by Barro and Lee (1993). The confluence of theory and data allowed researchers like Mankiw, Romer, and Weil (1992) to use growth regressions to determine what factors help explain differences in income and income growth rates across countries. Behnabib and Spiegel (1994) employed growth accounting to investigate the relative contributions of inputs, such as labor and capital, and technology or total factor productivity (TFP). TFP represents the residual portion of income growth not explained by changes in inputs.

Though much of the cross-country analyses have increased our knowledge on 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. Our analysis overcomes both of these criticisms by analyzing cross-state income and TFP 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 reduce both the possible problems associated with institutional heterogeneity and the errors that can be induced by business cycles when shorter time frame.

We use standard growth accounting methodologies to estimate the contribution of aggregate input growth and total factor productivity (TFP) growth on income growth across states. Assuming constant returns to scale and perfectly competitive factor markets, we decompose output growth into two components. The first component is the portion of output growth that is implied by the accumulation of inputs, and is called aggregate input growth. The second component is simply the residual not explained by aggregate input growth, and is called TFP or the "Solow residual." We find average TFP growth across states from 1840 – 2000 is 0.50% per year, which is approximately 36% of the output growth per worker, leaving 64% of output growth accounted for measured input growth. Clearly, there a large portion of output growth is accounted for by input growth, but a significant portion of output growth remains unexplained. We show that variability across states and census regions is non-trivial.

Following the suggestion of Klenow and Rodriguez-Clare (1997) and Easterly and Levine (2000), we consider the possibility that while TFP does not account for an overwhelming fraction of output growth, it may account for a large fraction of the cross-sectional variance in output growth. While the institutions, legal systems, educational systems, and tax rates of the various states are not entirely homogeneous, it is likely the case that these institutions display less heterogeneity than would typically be observed across countries. Having potentially mitigated this source of variation, we would expect to see a larger fraction of the variance in income growth explained by aggregate input growth. As these measures are sensitive to the treatment of any observed correlation between the growth rates of TFP and aggregate input, we construct plausible upper and lower bounds following the methodology of Baier, Dwyer, and Tamura (2004). For the entire 1840 – 2000 period, we find that aggregate input growth can explain at most 52% of the cross-sectional variance of income growth, while TFP growth could explain as much as 88% of the cross-sectional variance in income growth. However, limiting the data to 1880 – 2000, we find that the upper and lower bounds are much wider. Aggregate input can explain up to 86% while TFP growth could explain up to 89% of the cross-sectional variance in income growth. Utilizing state data with relatively homogeneous institutions has not driven the TFP component to zero. We also conduct further analysis of the sensitivity these measures display in regard to initial conditions and the time frame considered.

In order to complete the analyses described above, we require a time series on measures of inputs for each of the states. To this end, an additional purpose of this paper is to provide estimates of physical capital for each of the states of the United States from 1840 - 1900 at a decadal frequency and annually from 1902 – 2000. Primary data sources on physical capital at the state level become available only at the very end of the period we are examining. However, information on the amount of physical capital in each industry for the United States as a whole is available from the Bureau of Economic Analysis after 1902 and can be derived using information from Gallman (1960) for years prior to 1900. For years prior to 1900, we assume the capital-output ratio is identical across states within a given industry, and allow for differing capital-labor ratios across industries. For years after 1900, we assume the ratio of capital per worker in a state to the corresponding value for the nation as a whole is equal to the ratio of output per worker in a state to its corresponding value for the nation as a whole. These assumptions, while not ideal theoretically, enable us to create an estimate of the fraction of each industry's capital located in each state. By simply adding across industries in each state, we arrive at an estimate of physical capital for that state. Additional details of these calculations are provided in the following section.

We combine this newly created physical capital data with human capital and income measures provided by Baier, Mulholland, Tamura and Turner (2005), henceforth BMTT. BMTT create annual state level measures of human capital from 1840 – 2000. They also create state income measures for 1850, 1860, 1870, 1890, and 1910, which we they combine with decadal state income measures provided by Easterlin and annual figures produced by the Bureau of Economics analysis. The result is a dataset that contains income, human capital, and physical capital measures from 1840 – 2000 at the decadal frequency. Neither the physical capital nor the human capital measures have been previously analyzed in a growth accounting framework.

The organization of the remainder of the paper is as follows. The first section outlines the creation of the physical capital data and displays the results of these calculations for census regions. The second section provides a summary of the human capital and output measures created by BMTT and displays these measures for census regions. The third section presents the growth accounting framework and presents the analysis of the growth rates for states and for census regions. The fourth section analyzes the variance of the growth rates across states for the nation as a whole and within broader regions. The final section offers a brief conclusion and outlines future work.

Physical Capital

The first step in the construction of state level physical capital estimates is to calculate physical capital stocks at the industry level for the United States as a whole. For 1840 – 1900, we use data provided by Gallman

(1960) which contains data at the industry level for the nation as a whole. For each industry,[2] Gallman provides an estimate of that industry's share of total value added (a sectoral share), as well as an estimate of the capital-value added ratio for that industry. Letting Y_j denote value added (output) in industry j, it is simple to calculate K_j, the amount of capital employed in industry j:

$$K_j = \left(\frac{Y_j}{\sum_j Y_j} \right) \left(\frac{K_j}{Y_j} \right) \sum_i Y_i$$

The first term in the expression is Gallman's estimate of the sectoral shares of output, the second term is Gallman's estimate of the capital-value added ratio in each industry, and the final term is value added for the nation as a whole from BMTT.[3] The result is total physical capital in each industry from 1840 – 1900.

For 1902 – 2000, we use data provided by the Bureau of Economic Analysis in the Fixed Reproducible Tangible Wealth series. This source provides an estimate of the capital stock at the industry level for 1925 – 2000.[4] This source does not provide data on physical capital stocks for the period 1902 – 1925, but does provide figures on gross investment flows into each industry which we use to derive estimates of the capital stock for this period.

Let K_{jt} denote the capital stock for industry j in period t, let I_{jt} be the gross investment flow into industry j in period t, and define $\Delta K_{jt} = \ln(K_{jt}) - \ln(K_{jt-1})$. In order to estimate the capital stock for the period 1902 – 1925, we fit the following equation with ordinary least squares for each industry:

$$\Delta K_{jt} = \beta_0 + \beta_1 * \Delta K_{jt+1} + \beta_2 \left(\frac{I_{jt}}{K_{jt}} \right)$$

We then use the result of this estimation to go backward in time from the 1925 value of the capital stock to determine the values of the capital stock for years prior to 1925. We find this algorithm generates a reasonable

[2] Gallman's data subdivides the economy into the following industries: Agriculture, Manufacturing, Transportation, Commerce, Government, and Residences. For additional details, see Appendix B.

[3] We utilize the state level measures of output constructed in BMTT. BMTT provide estimates of output at the state level, rather than the industry level. Since $\sum_i Y_i = \sum_j Y_j$, one could substitute the later in the expression for capital in each industry above.

[4] The BEA data uses the following industry classifications: Agriculture Forestry and Fishing, Mining, Construction, Manufacturing, Transportation, Wholesale Trade, Retail Trade, FIRE, and Services. For additional details, see Appendix B.

capital stock path for all industries with the exception of agriculture industry. For this industry, we estimate the following equation and use the predicted values to determine the values of capital stock in for 1901 – 1925.

$$\ln(K_{ag,t}) = \beta_0 + \beta_1 * \ln(I_{ag,t})$$

For all industries except government and residences, the BEA series provides investment data from 1901 – 2000. Thus by utilizing the algorithm described above, we are able to derive values of the capital stock going back to 1902. However, data is more limited for government capital and residential capital. As we have no data for the either the stock of capital or investment flows in the government sector prior to 1925, we simply assume the growth rate of government capital in this period is 6% per year, which corresponds to the growth rate of government capital observed over the 1925 – 2000 period and is very similar to the rate observed from 1840 – 1900. Residential capital is discussed directly below.

Residential Capital

Residential capital makes up nearly half of all physical capital throughout the period and thus is quite important. While the Gallman data enables us to calculate an estimate of residential capital from 1840 – 1900 and the BEA data provides an estimate from 1947 – 2000, we have no data from these sources for the period 1900 – 1947. In order to fill in the gap, we use two additional data series taken from Historical Statistics of the United States (HSUS). HSUS provides annual estimates for 1925 to 1970 of Net Stocks of Residential Structures (Series F213) and provides estimates in 1900, 1912, 1922, and 1929 of Wealth in Residential Nonfarm Structures (Series F425). Unfortunately, the coverage in the sources varies particularly depending on treatment of farm residences, government-owned housing, and other methodological considerations. We assume that the growth rates of residential capital are those that are implied by the data in these sources. Thus we use the BEA data for 1947 - 2000, and use the growth rates implied by the two HSUS series and Gallman's data to go backward in time to create the residential capital series for years prior to 1947. We assume constant growth rates between 1900 - 1912, 1912 - 1922, and 1922 - 1929. The time path of residential capital and of non-residential capital, which includes all other industries, is displayed below in Figure 1. The data is displayed on a logarithmic scale. Annual estimates are displayed after 1902, though in the growth accounting analyses that follows, only decadal values will be utilized. These measures, and all subsequent measures throughout this paper, are reported in 2000 dollars.

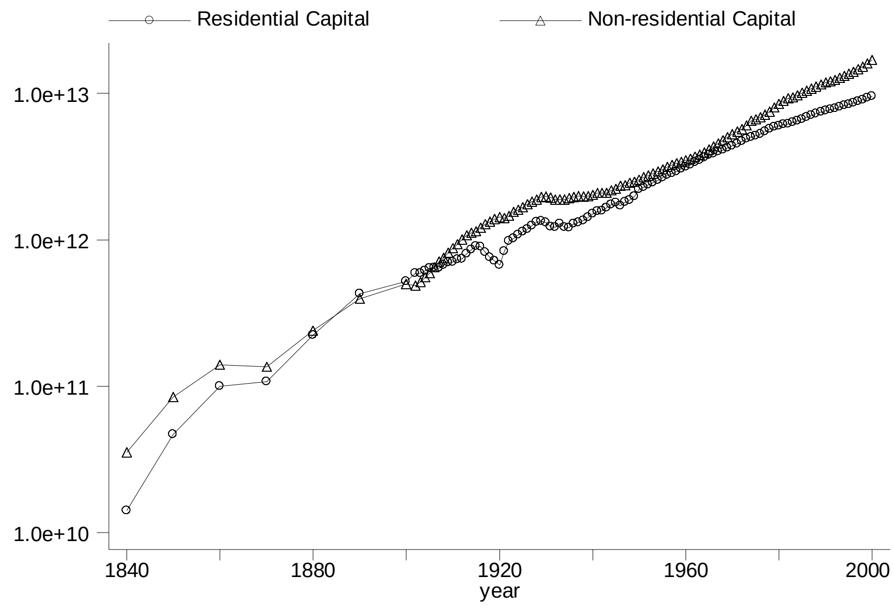


Figure 1 Residential Capital and Non-Residential Capital in United States, 1840 – 2000, logarithmic scale, measured in 2000 dollars.

State Level Estimates

We next need an algorithm to determine what fraction of each industry's capital resides in each state. We begin with the years prior to and including 1900. Recall that we have physical capital data by industry at the national level. The state level data available to us for this period includes employment in each industry, and output in each industry from BMTT, albeit output for a broader industry classification.[5] One alternative would be to allocate capital to states assuming a constant capital-labor ratio in each industry, utilizing state variation in the fraction of total employment in each industry. A second alternative would be to allocate capital across states assuming a constant capital-output ratio for each industry, utilizing state variation in the fraction of total output in each industry. While ideally we would allow these ratios to vary within each industry, these assumptions are motivated directly by data availability issues. With the first assumption, the fraction of industry j's capital that is located in state i will be identical to the faction of industry j's *employment* that is located in state j. Under the second assumption, the fraction of industry j's capital that is located in state I will be identical to the fraction of industry j's *output* that is located in state j.

[5] The state level data on output is available only for the agriculture, manufacturing, and non-agricultural non-manufacturing sectors. See Appendix C in BMTT for further details.

Let E_{ijt} represent employment in state i and industry j, during period t. Let K_{jt} represent physical capital in industry j during period t. Then under the first assumption:

$$K_{ijt} = \left(\frac{E_{ijt}}{\sum_i E_{ijt}} \right) * K_{jt}$$

Next we simply sum across all industries for each state to arrive at the state level measure of capital:

$$K_{it} = \sum_j K_{ijt}$$

Let Y_{ijt} represent output in state i and industry j, during period t. Then under the second assumption:

$$K_{ijt} = \left(\frac{Y_{ijt}}{\sum_i Y_{ijt}} \right) * K_{jt}$$

And again, we sum across all industries for each state to arrive at the state level measure of capital in each industry. We choose the second assumption, as we feel that allocating on the basis of capital-labor ratios assigns too much physical capital to newly formed states.

After 1900, we have a different set of data at our disposal. While employment data at the state level continues to be available, output at the industry level is not available between 1920 and 1978. One alternative is to proceed as described above using the first assumption above and allocating based on employment figures. A second alternative is to assign capital to each state by assuming that physical capital per worker in each state relative to the value of capital per worker for the nation as a whole is identical to output per worker in each state relative to the value of output per worker for the nation as a whole.[6] Letting L_{it} denote the labor force in state i, and time t:

$$\frac{K_{it}}{L_{it}} = \left(\frac{\frac{Y_{it}}{L_{it}}}{\frac{\sum_i Y_{it}}{\sum_i L_{it}}} \right) * \frac{\sum_i K_{it}}{\sum_i L_{it}}$$

We choose the second alternative, and thus assign capital by utilizing information on relative productivity values.[7]

[6] A third possibility, which we intend to pursue in future work, is to use available information from 1978 – 2000 that contains output by sector, as well as the sectoral composition data from 1840 – 1900, to predict the sectoral composition from 1900 – 1978. From there, we would proceed by assuming the capital-output ratio is identical for states within an industry. Covariates for predicting sectoral shares would include the employment figures across industry, human capital measures, and controls for regions.

While the framework above is quite simple, the details of the calculations become slightly more complex when the data is examined. The first issue is that Gallman's data separates output into a different set of industries than the industries provided by the BEA data.[8] In addition, when we utilize data on employment by industry, the data in fact comes to us in a variety of occupation groups that must be mapped into each of these industries.[9] The source of employment data is reports of United States Census of Population reports. Throughout the period examined the census reports occupations of workers by state which are then aggregated into broader classes or industries. However, as mentioned above, these occupations do not necessarily correspond to the sectoral breakdown in the Gallman data and the BEA data.[10]

Physical Capital Estimates

Table 1 below summarizes the natural log of physical capital per worker. Physical capital falls only between 1860 and 1870 and between 1890 and 1900. While new states entering the data tend to increase the coefficient of variation early in the period examined, there is a slow downward trend from 1900 until 1940 and another downward trend from 1950 to 2000. There is a sharp declined in the coefficient of variation from 1940 to 1950 which is largely a result of rapid capital growth in Southern states which had been low capital states. In the first hundred years of the data, the states that have the highest values of capital per worker are consistently located in the Mid Atlantic and New England census regions and later the industrial states of the East North Central region. Those states with the lowest values of capital per worker typically are located in the South. In the later portion of

---

[7] As a robustness check, we also calculate capital allocating on the basis of the employment data. The results are extremely similar qualitatively.

[8] As the industry classifications do not match between the two sources of data, it is not informative to graph the time paths of the fraction of all employment that occurs in each industry or the capital stock in an industry. Each graph would have a discrete jump in 1900.

[9] In the current version of this paper, the employment by industry data is being utilized only in the robustness check described in footnote 7. However, in the next version of this paper, we again plan to utilize this data as described in footnote 6.

[10] A detailed listing of the mapping is included in Appendix B. In addition, occupations in 1860 are reported in the census reports, but are not aggregated by industry. While it would be possible to aggregate from occupation to industry, the number of occupations is quite large and the occupation listings are quite esoteric. Rather than attempt this process, we simply linearly interpolate the fraction of total industry employment located in state i in 1860 using the corresponding 1850 and 1870 values.

the data, states in the northern portion of the South Atlantic and the Pacific census regions are consistently among the leaders.

Table 1 – Summary statistics on physical capital per worker, labor force weighted, measured in 2000 dollars.

Year	States	Mean	Std. Dev	Coeff. Var.	Min	Max
1840	28	10,330	3,129	0.30	5,943	19,748
1850	34	17,359	7,171	0.41	8,445	63,231
1860	38	21,834	6,575	0.30	11,006	43,031
1870	42	19,518	5,777	0.30	8,339	44,366
1880	44	26,782	8,480	0.32	11,708	45,692
1890	44	36,694	12,344	0.34	12,830	55,909
1900	47	35,431	11,239	0.32	13,283	55,480
1910	47	42,319	13,532	0.32	16,805	68,676
1920	48	50,862	13,769	0.27	21,216	72,229
1930	49	67,295	20,176	0.30	23,538	113,642
1940	49	70,770	19,052	0.27	27,551	111,196
1950	51	79,076	11,072	0.14	47,828	108,148
1960	51	95,229	12,569	0.13	64,637	124,326
1970	51	117,295	13,070	0.11	86,524	149,850
1980	51	135,532	13,451	0.10	101,635	200,054
1990	51	154,776	13,056	0.08	114,434	206,253
2000	51	187,993	20,018	0.11	133,191	263,609

While we have estimates of physical capital per worker for each state, it would be cumbersome to display the graphs for each state. Figures 2 - 4 show real physical capital per worker for the nine census regions across time on a logarithmic scale.[11] The slopes of the lines are growth rates. The measure of physical capital for each region is the labor force weighted average of each state's physical capital per worker. During the early portions of the dataset, and for some regions, data on income per worker is not available for all states within a region. Where this is the case, the levels displayed in Figures 2 - 4 are the levels implied by the growth rates for those states that we observe income in the both the current and preceding period.[12]

[11] See Appendix A for a listing of the states in each region. We start with the New England census region and move clockwise around the census regions until we arrive at the East North Central.

[12] We begin with output per worker in 2000 and then calculate the growth rate for 1990 – 2000 for the states with data in both 1990 and 2000. The value in 1990 is the level implied by this growth rate. We continue working backwards in this fashion. While the level of output for a region is not necessarily the actual level of output per

Figure 2 displays that New England is a high physical capital region early, but by the end of the period is below the national average. The Middle Atlantic census region, not surprisingly, is higher than average throughout the entire period. Figure 2 also shows that the South Atlantic census region is well below the national average for much of the period, begins to catch up to the rest of the nation beginning in 1900, and by 1980, this region has surpassed the national average.

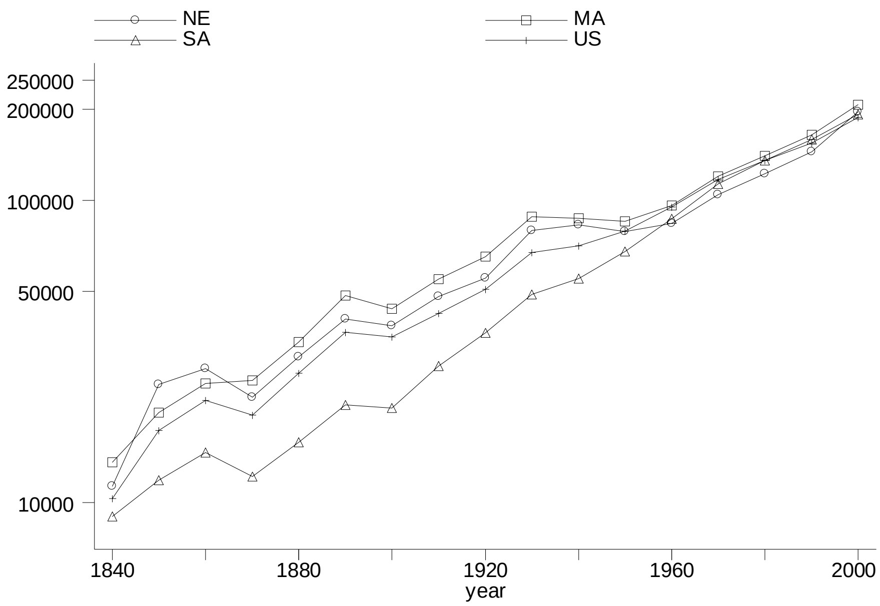


Figure 2 Physical capital per worker in New England, Mid Atlantic, South Atlantic, and United States, 1840 – 2000, logarithmic scale, measured in 2000 dollars.

Figure 3 displays the physical capital per worker measures for the East South Central, West South Central, and Mountain census regions. The Mountain census region begins with a very high capital measure, but remains near the national average throughout the period. Surely the emphasis on mining in here and in the Pacific census region is resulting in high values of capital per worker.[13] The East South Central and West South Central census regions follow very similar paths, though the East South Central has a lower value of capital per worker throughout the entire period. Both regions are well below the national average in 1930, but make up ground substantially thereafter, with the West South Central region having a value above the national average after 1960. Figure 4

worker, the growth rate of output is always the growth rate of output for the states in that region for which we have data. While this adjustment does not change the levels of output in a region appreciably, failing to make this adjustment would tend to bias growth rates of output downward slightly. See Appendix A for the year in which data is first available for each state.

[13] This high measure of capital may also be reflecting mismeasurement of price levels in the Mountain and Pacific regions. This issue is discussed in greater detail below.

illustrates that the Pacific regions behaves much like the Mountain region. Also, the East North Central and West North Central region follow the national average very closely from 1840 – 2000, with the West North Central trailing the East North Central region slightly.

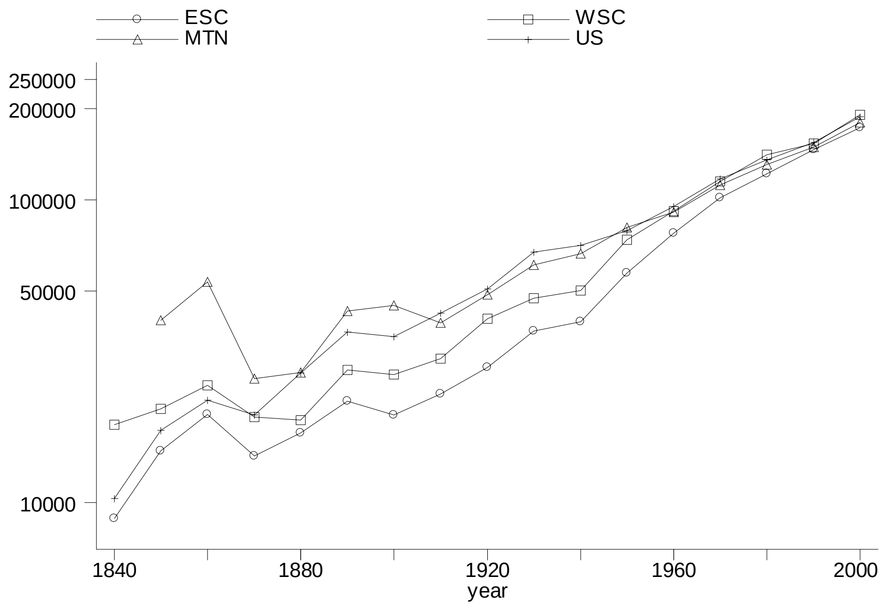


Figure 3 Physical capital per worker in East South Central, West South Central, Mountain, and United States, 1840 – 2000, logarithmic scale, measured in 2000 dollars.

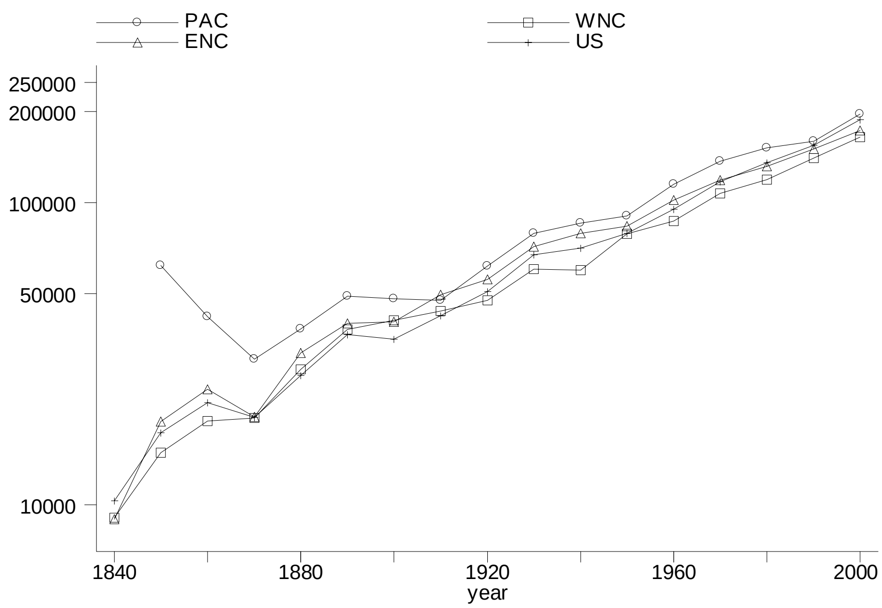


Figure 4 Physical capital per worker in Pacific, West North Central, East North Central, and United States, 1840 – 2000, logarithmic scale, measured in 2000 dollars.

Table 2 below presents the log of physical capital per worker for each region and the United States as a whole. The region with the highest per worker physical capital measure is in bold.

Table 2 Physical capital per worker for census regions and United States, labor force weighted, measured in 2000 dollars.

Year	NE	MA	SA	ESC	WSC	MTN	PAC	WNC	ENC	US
1840	11,329	13,603	8,155	8,854	**18,026**	.	.	8,012	8,606	10,330
1850	24,598	19,847	10,731	14,817	19,735	21,195	**62,554**	13,156	18,912	17,359
1860	27,730	24,823	13,282	19,564	23,650	28,435	**42,310**	18,811	24,178	21,834
1870	22,308	25,422	11,540	14,283	18,540	25,897	**30,607**	19,251	19,570	19,518
1880	30,401	34,073	14,989	17,006	18,156	27,058	**38,472**	27,929	31,822	26,782
1890	40,545	48,441	19,925	21,649	26,592	43,088	**49,416**	37,811	39,872	36,694
1900	38,546	43,796	19,449	19,494	25,637	44,927	**48,422**	40,894	40,322	35,431
1910	48,076	**54,858**	26,789	22,822	28,952	39,246	47,872	43,855	49,660	42,319
1920	55,358	**65,099**	34,437	28,011	40,575	48,723	62,058	47,574	55,836	50,862
1930	79,622	**88,259**	48,953	36,981	47,309	60,918	79,595	60,221	71,579	67,295
1940	83,088	**87,418**	55,133	39,539	50,170	66,596	86,117	59,908	79,205	70,770
1950	78,872	85,201	67,759	57,384	73,969	80,999	**90,380**	78,978	83,759	79,076
1960	84,028	96,328	87,060	77,734	92,025	91,222	**114,991**	87,091	102,092	95,229
1970	104,649	120,204	113,226	101,740	115,225	111,941	**137,274**	107,199	118,691	117,295
1980	122,619	140,632	135,450	122,055	141,206	131,044	**151,962**	119,006	131,952	135,532
1990	144,804	**164,851**	159,347	146,799	153,888	149,699	159,941	140,390	150,731	154,776
2000	196,420	**207,073**	192,551	173,102	191,326	179,954	196,254	164,764	173,192	187,993

As much of the literature discussing state income convergence focuses primarily on the role that Southern and Western states had in producing income convergence, we find it useful to aggregate into even broader regions. We thus aggregate into what we call the North, South, and West. We define the North region as the New England, Mid Atlantic, and East North Central census regions. We define the South region as the South Atlantic, East South Central, and West South Central census regions. We define the West as Pacific, Mountain, and West North Central regions.[14] Figure 5 displays physical capital per worker for each of these broad regions, while Table 3 displays the

[14] While the other pairings seem fairly natural, we decided to include the East North Central region with the Mid Atlantic and New England states. We find throughout the entire period, the income and human capital levels of the East North Central region are very similar to those of the other census regions we include in the North category.

numerical values and summary statistics. Both display that significant regional convergence has occurred, especially from 1910 to 1980.[15]

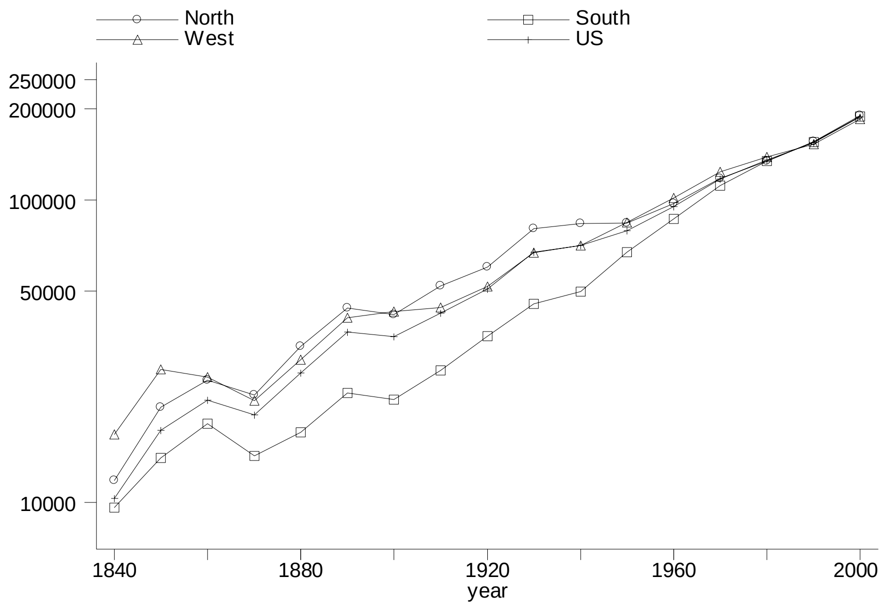


Figure 5 Physical capital per worker for North, South, West and United States, labor force weighted, logarithmic scale, measured in 2000 dollars.

---

[15] The pattern of capital convergence across states will be identical to the pattern of income convergence across states after 1900 as a result of the way capital was allocated across states.

Table 3 – Summary statistics on physical capital per worker, labor force weighted, in 2000 dollars.

Year	North	South	West	US	Std. Dev.	Coeff. Var.
1840	**11,563**	9,007	8,012	10,330	1,589	0.15
1850	**20,563**	13,256	25,584	17,359	4,951	0.29
1860	25,197	17,196	**25,279**	21,834	4,604	0.21
1870	**22,489**	13,625	21,783	19,518	4,567	0.23
1880	**32,514**	16,305	29,691	26,782	8,223	0.31
1890	**43,638**	21,958	40,773	36,694	10,846	0.30
1900	41,520	20,906	42,840	35,431	11,070	0.31
1910	**51,730**	26,097	44,211	42,319	12,970	0.31
1920	**59,840**	34,563	51,983	50,862	12,431	0.24
1930	**80,166**	45,308	67,228	67,295	16,841	0.25
1940	**83,452**	49,743	70,890	70,770	16,013	0.23
1950	83,731	67,256	**84,473**	79,076	8,358	0.11
1960	97,198	86,627	**101,767**	95,229	6,827	0.07
1970	117,388	111,598	**123,888**	117,295	5,875	0.05
1980	134,044	134,754	**138,633**	135,532	2,632	0.02
1990	155,433	**155,540**	153,016	154,776	1,513	0.01
2000	**189,570**	188,894	185,083	187,993	2,544	0.01

To further explore convergence at the state level and census regional levels, Table 4 reports the gap between the states (or regions) with the highest and lowest value of physical capital per worker. The measure of the gap reported is the difference between the logarithms of the state with the largest value and the state with the lowest value. The row marked R presents gaps between census regions while the row marked S represents gaps between states. We also present the gap between states excluding the District of Columbia in the row marked $\tilde{S}$, omitting the entry if is value is identical to the value in the row above. The gap between regions decreases consistently from 1930 to 1990, as does the gap between states.[16]

Table 4 Difference between logarithms of regions (states) with highest and lowest value of capital per worker.

	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
R	0.87	0.80	0.81	0.76	0.73	0.75	0.66	0.41	0.38	0.30	0.24	0.16	0.22
S	1.36	1.47	1.43	1.41	1.23	1.57	1.40	0.82	0.65	0.55	0.68	0.59	0.68
$\tilde{S}$												0.45	0.53

[16] See previous footnote.

Growth Accounting Framework and Data

There has been little growth accounting analysis of state level data over a historical perspective. Proper analysis of growth requires a long time series of data to ensure that business cycles or other high frequency data are not confounding the results. Many of the analyses that have been conducted consider only the United States as a whole, and include Abramovitz (1956), Solow (1957), Kendrick (1961), Denison (1985) and others. Others using data across countries and include a measure of the physical capital stock, including Maddison (1995) and Jones (1997). More recent analyses have included measures of human capital as an input, including Klenow and Rodriguez-Clare (1997). While the basic growth accounting techniques are standard, the analysis of explaining cross-sectional variation in growth rates of output is similar to Baier, Dwyer, and Tamura (2004). We use historical data on state output per worker, state level measures of human capital, and state level measures of physical capital that have not been previously analyzed. As far as we know, this is the first empirical analysis of its kind in regard to the scope of the data utilized.

Output Per Worker

Annual data on income by state is available beginning 1929 from the Bureau of Economic Analysis. Data on output is available from Easterlin in 1840, 1880, and 1920. BMTT estimated measures of state output in 1850, 1860, 1870, 1890, 1900, and 1910. Given information on the size of the labor force from various issues of the decennial census and *Historical Statistics of the United States: Colonial Times to 1970*, a measure of income per worker (or output per worker) can be calculated.[17]

All income data provided by the sources above are nominal figures. BMTT utilizes information on both annual *national* price level variation and less frequent observation of *interregional* price variation. Data on national price level comes from Gordon (2000) for 1875 – 2000, while data prior to 1875 is from *Historical Statistics of the United States*. In addition, BMTT uses three sources of information on relative price levels across regions. By combing data from Mitchener and McLean (1999), Willamson and Linder (1980), and Berry, Fording, and Hanson (2000), BMTT have observations on relative price levels in each region dating back to 1840 for all census regions

[17] Details on data sources for the labor force are included in Appendix B in BMTT.

except the Mountain and Pacific census regions, which are first observed in 1880.[18] BMTT assumes the relative regional price difference observed in these regions in 1880 persists from 1840 to 1870, and then normalize regional price levels to the national price level figures given in Gordon.[19] All income measures are reported in 2000 dollars.

Summary statistics on real income are displayed in Table 5. Mean income falls only between 1890 and 1900. A striking feature is the convergence in income across states. While the introduction of new states in the early portion of the data tends to increase the coefficient of variation across states, this measure declines nearly uniformly from 1890 until 1990, reversing only between 1920 and 1930. Interestingly, the coefficient of variation increases between 1990 and 2000. A surprising value in the table is the maximum income level in 1850. The state with this maximum income figure in 1850 is California, a result of large mining income. We find that other states in the Pacific region have large values of income per worker. In addition to mining income, we feel this region's income may be overstated as a result of failing to account for all of the regional variation in prices.[20]

[18] Additional details on both national and regional price levels are included in Appendix B in BMTT.

[19] BMTT surmises the relative price level in the Pacific and Mountain regions may have been even larger between 1840 and 1870 than its level in 1880. This supposition is based on the trend observed from 1880 to 1920 and the nominal income figures observed for these regions.

[20] See previous footnote.

Table 5 – Summary statistics on real income per worker, labor force weighted, measured in 2000 dollars

Year	States	Mean	Std. Dev	Coeff. Var.	Min	Max
1840	28	4,950	1,706	0.34	2,660	9,218
1850	34	7,034	3,468	0.49	2,766	27,828
1860	38	7,490	2,838	0.38	3,144	16,672
1870	42	7,939	2,945	0.37	2,836	20,672
1880	44	9,448	3,510	0.37	3,297	18,972
1890	44	11,821	4,739	0.40	3,175	19,195
1900	47	11,477	3,983	0.35	3,678	17,088
1910	47	12,531	4,007	0.32	4,976	20,335
1920	48	14,430	3,906	0.27	6,019	20,492
1930	49	16,442	4,929	0.30	5,751	27,766
1940	49	18,328	4,934	0.27	7,135	28,797
1950	51	24,286	3,401	0.14	14,689	33,215
1960	51	29,514	3,895	0.13	20,032	38,531
1970	51	39,139	4,361	0.11	28,871	50,002
1980	51	42,083	4,177	0.10	31,558	62,117
1990	51	48,552	4,096	0.08	35,897	64,700
2000	51	58,791	6,260	0.11	41,653	82,438

Figures 6 – 8 display the regional aggregates for real income (output) per worker on a logarithmic scale. We use the same algorithm to correct for new states entering regions; see footnote 9 for details.

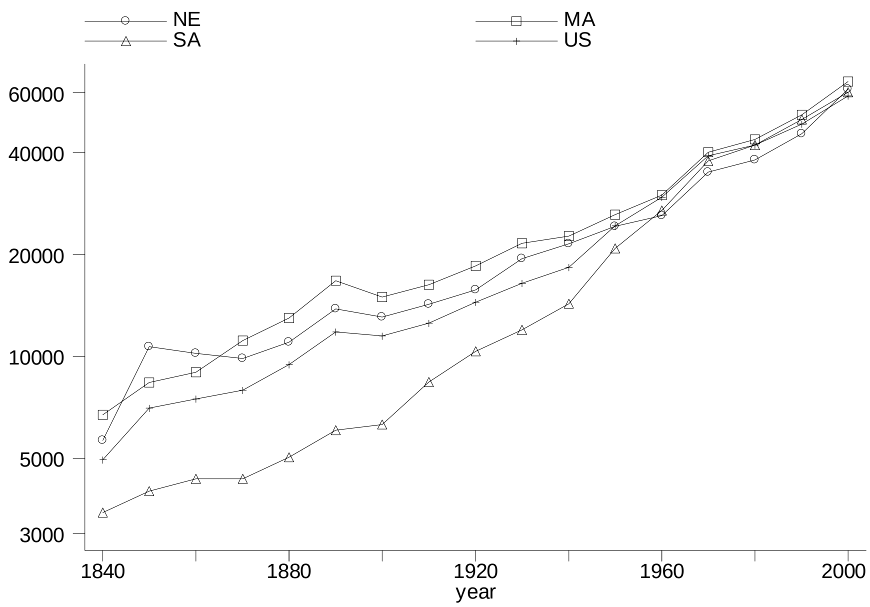


Figure 6 Real output per worker in New England, Mid Atlantic, South Atlantic, and United States, 1840 – 2000, logarithmic scale, measured in 2000 dollars.

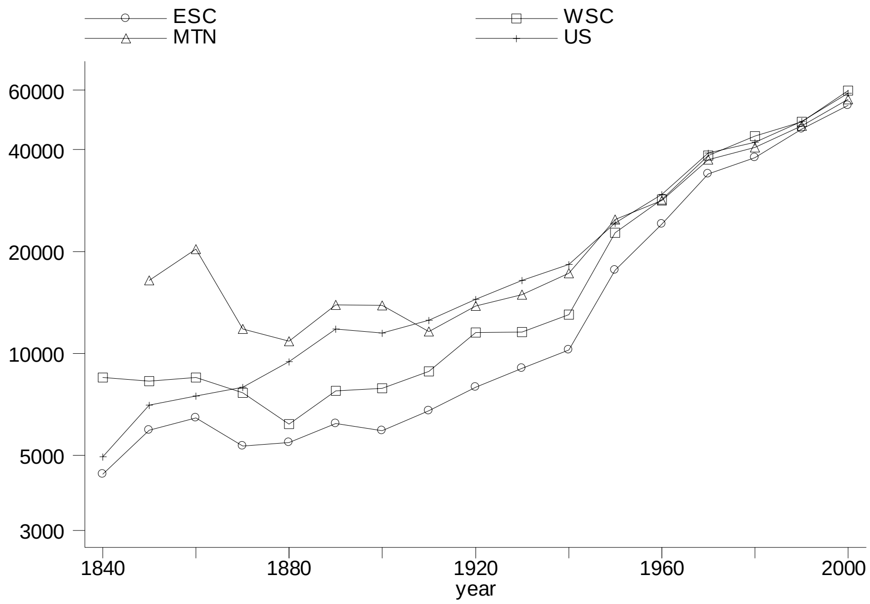


Figure 7 Real output per worker in East South Central, West South Central, Mountain, and United States, 1840 – 2000, logarithmic scale, measured in 2000 dollars.

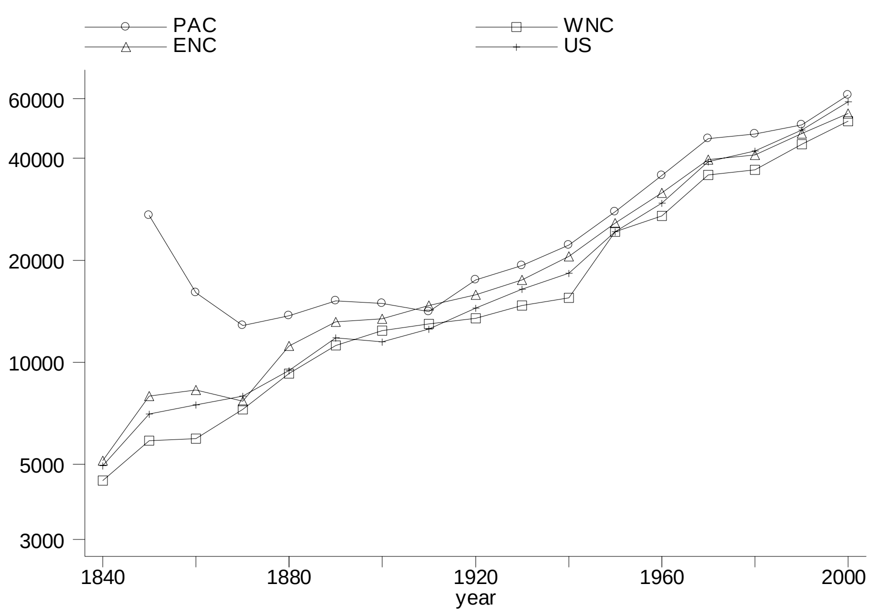


Figure 8 Real output per worker in Pacific, West North Central, East North Central, and United States, 1840 – 2000, logarithmic scale, measured in 2000 dollars.

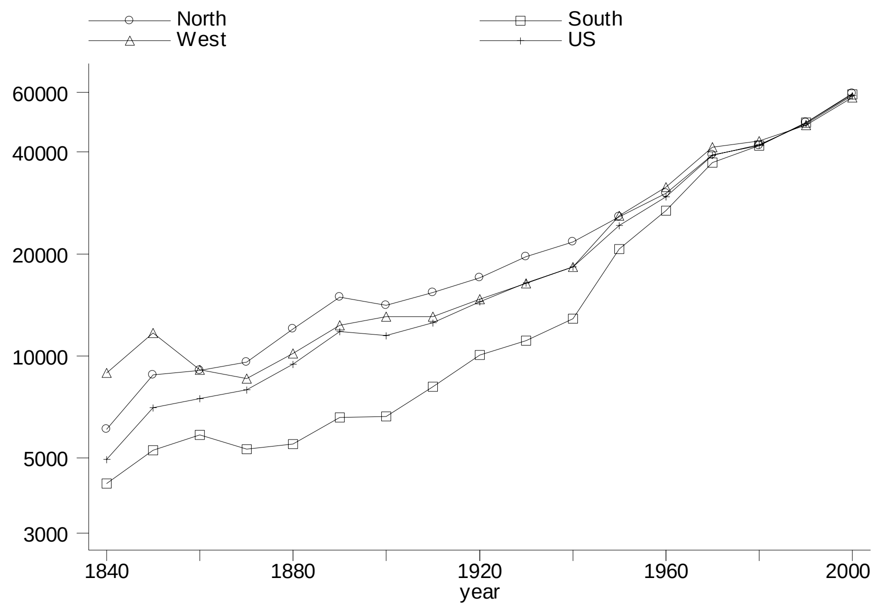
What is clear is that the Pacific, the Mid Atlantic, to a lesser extent, the East North Central census regions are income leaders throughout the period. New England begins as a leader, falls below the national average briefly in 1950, and then follows the national average very closely. The South Atlantic, East South Central and West South Central census regions all follow similar paths and are consistently below the US average, particularly so in the early portion of the data. The West North Central and Mountain regions follow the US average very closely after 1880. The income levels for each census region and decennial year are listed in Table 6.

Table 6 Real income per worker for census regions and United States, labor force weighted, measured in 2000 dollars.

Year	NE	MA	SA	ESC	WSC	MTN	PAC	WNC	ENC	US
1840	5,640	6,709	3,089	4,391	**8,363**	.	.	3,825	4,867	4,950
1850	10,668	8,360	3,581	5,926	8,019	8,261	**27,446**	5,013	7,941	7,034
1860	10,216	8,952	3,882	6,442	8,209	10,236	**16,167**	5,945	8,265	7,490
1870	9,832	11,112	4,106	5,316	7,392	11,889	**12,913**	7,247	7,668	7,939
1880	10,998	12,954	4,751	5,447	5,971	10,913	**13,787**	9,248	11,147	9,448
1890	13,810	**16,743**	5,724	6,191	7,503	14,007	15,294	11,222	13,132	11,821
1900	13,073	14,947	5,929	5,900	7,641	13,838	**14,992**	12,395	13,440	11,477
1910	14,236	**16,244**	7,932	6,758	8,573	11,621	14,175	12,986	14,705	12,531
1920	15,706	**18,469**	9,770	7,947	11,512	13,823	17,607	13,497	15,841	14,430
1930	19,454	**21,564**	11,961	9,035	11,559	14,884	19,447	14,714	17,489	16,442
1940	21,518	**22,639**	14,278	10,240	12,993	17,247	22,302	15,515	20,512	18,328
1950	24,224	26,168	20,811	17,624	22,718	24,877	**27,758**	24,256	25,725	24,286
1960	26,042	29,854	26,982	24,092	28,521	28,272	**35,638**	26,991	31,641	29,514
1970	34,919	40,110	37,781	33,949	38,449	37,353	**45,806**	35,770	39,605	39,139
1980	38,074	43,667	42,058	37,899	43,845	40,690	**47,185**	36,952	40,972	42,083
1990	45,424	**51,713**	49,986	46,050	48,273	46,959	50,172	44,039	47,283	48,552
2000	61,426	**64,758**	60,216	54,134	59,833	56,277	61,374	51,527	54,162	58,791

Inspection of Figures 6 - 8 shows visually that there has been regional income convergence during the period. Figure 9 displays income for the North, South, West, and the United States, while Table 7 details the income levels. What is noticeable in Figure 9 is the convergence brought on by the South beginning in 1900 and continuing throughout the period. The gap between the region with the highest income (most often the North) and the region with the lowest income per worker (almost always the South) narrows between 1890 and 1990, with the exception of the period from 1920 to 1930. Table 8 reports the gap between the states (and regions) with the highest and lowest value of physical capital per worker. The measure of the gap reported is again the difference between the values of the state with the highest value and the state with lowest value, expressed as a fraction of the average value of for the nation as whole. The row marked R presents gaps between census regions, the row marked S represents gaps between states, and the row marked $\tilde{S}$ the gap between states eliminating the District of Columbia.[21] Both show evidence of broad convergence for states and census regions until at least 1970.

Figure 9 Real output per worker in North, South, West, and United States, 1840 – 2000, logarithmic scale, measured in 2000 dollars.



[21] For values after 1900, the measures reported in Table 8 are identical to the measures reported in Table 4. Because capital was assigned on the basis of relative output per worker measures, the capital per worker measures will display the same pattern of convergence as the output per worker measures. In the subsequent version of this paper, we plan to allocate capital across states using a different algorithm utilizing sectoral shares of output. See footnote 5.

Table 7 Real income per worker for North, South, West and United States, labor force weighted, measured in 2000 dollars.

Year	North	South	West	US	Std. Dev.	Coeff. Var.
1840	**5,895**	3,898	3,825	4,950	1,208	0.24
1850	**8,720**	4,945	10,632	7,034	2,456	0.35
1860	**8,975**	5,493	8,763	7,490	1,974	0.26
1870	**9,491**	5,057	8,605	7,939	2,236	0.28
1880	**11,877**	5,231	10,214	9,448	3,342	0.35
1890	**14,787**	6,269	12,368	11,821	4,190	0.35
1900	**14,026**	6,318	13,073	11,477	3,923	0.34
1910	**15,318**	7,727	13,091	12,531	3,841	0.31
1920	**16,977**	9,806	14,748	14,430	3,527	0.24
1930	**19,587**	11,070	16,426	16,442	4,115	0.25
1940	**21,612**	12,882	18,359	18,328	4,147	0.23
1950	25,716	20,656	**25,944**	24,286	2,567	0.11
1960	30,124	26,848	**31,540**	29,514	2,116	0.07
1970	39,170	37,238	**41,339**	39,139	1,960	0.05
1980	41,621	41,841	**43,046**	42,083	817	0.02
1990	48,758	**48,791**	48,000	48,552	475	0.01
2000	**59,284**	59,073	57,881	58,791	796	0.01

Table 8 Difference between logarithms of regions (states) with highest and lowest value of capital per worker.

	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
R	0.95	0.93	0.79	0.76	0.73	0.75	0.66	0.41	0.38	0.30	0.24	0.16	0.22
S	1.66	1.36	1.17	1.23	1.00	1.34	1.18	0.76	0.63	0.54	0.73	0.59	0.69
$\tilde{S}$												0.42	0.49

Human Capital Per Worker

Information on the level of educational attainment of the labor force and human capital per worker was developed in BMTT, and is similar to those measures developed in Barro and Lee (1993). The dataset contains information both on years of schooling and average experience of members of the labor force. Data on enrollment in elementary, secondary, and higher education was collected, along with data on the age distribution of population. Utilizing this information, BMTT were able to calculate elementary, secondary, and higher education enrollment rates by state. BMTT then calculated the fraction of the labor force that falls into the categories of having no schooling, some elementary schooling but no more, some secondary schooling but no more, and some higher educational exposure using a perpetual inventory methodology. They were able to derive the time path of these shares of the population in each educational class on an annual basis. Utilizing decennial census data, they calculated the expectation of the number of years of school attended conditioned on being in each educational class. The result is an estimate of the years of schooling of the average worker in each state for decadal years from 1840 to 1920, and annually from 1929 – 2000.

Next, Mincerian measure of human capital was specified:

$$human\,capital_{it} = \exp\left(\beta E_{it} + \gamma_1 x_{it} - \gamma_2 x_{it}^2\right)$$

where E_{it} is years of schooling as described above and x_{it} is experience (per worker) in state i in period t. Experience was defined as the average age of those persons in the labor force less 6 years, and was estimated by using the age distribution of the population (of those aged 65 and under) and adjusting using the age profile of those currently enrolled in school.

From here it seems plausible to make of two choices concerning the parameters β, γ_1, and γ_2 in the human capital equation. One option is to choose parameter estimates put forth by earnings regression in the literature as is suggested by Jorgenson and Griliches (1967). The other is to use BMTT's own estimates of these parameters. The general specification is of the form:[22]

$$\ln income_{it} = \ln A_{it} + \alpha \ln\left(\frac{w_t}{r_t}\left(\frac{\alpha}{1-\alpha}\right)\right) + \beta Ed_{it} + \gamma_1 x_{it} - \gamma_2 x_{it}^2$$

where w_t and r_t represent the wage rate and rental rate on capital in time period t, respectively. A_{it} measures a technology term and α is capital's share of income. There are a number of specifications outlined in BMTT depending on whether experience in separated into male and female components, the inclusion of the quadratic term on experience, the inclusion of dummies for Alaskan effects, the general structure of fixed or random effects, as well as estimation techniques used due to deal with potential endogeneity of income and investment in human capital

[22] As in most studies physical capital data is not available, in deriving this equation we have assumed the capital use is proportional to human capital use. The regression estimate of β will represent the return to the aggregate input. To determine the return to human capital we must multiply β by labor's share of income (1-α).

accumulation. Clearly, measures of the growth rate of human capital are sensitive to the choice of these parameters, and thus, so too are measures of the growth rate of TFP.

In the analysis that follows, we choose to use generally accepted parameters from the literature, primarily to make our results comparable to existing literature. Returns to education are divided into components for primary schooling, P, intermediate schooling, I, and secondary and higher education, S. The assumption is made that primary schooling must be completed to attend intermediate schooling, and intermediate schooling must be complete to attend secondary and higher education. Primary schooling is assumed to last 4 years, while intermediate schooling is also assumed to last 4 years.[23] Suppressing subscripts for states and years, human capital can be expressed as:

$$human\,capital = h_0 \exp\left(\phi_P\, P + \phi_I\, I + \phi_S\, S + \gamma_1\, x - \gamma_2\, x^2\right)$$

where h_0 is the level of human capital with no schooling or experience, ϕ_P, ϕ_I, and ϕ_S are parameters on years of primary, intermediate, and secondary and higher education, γ_1 and γ_2 are parameters on experience and experience squared. We follow Hall and Jones (1999) and assign $\phi_P = 0.134$, $\phi_I = .101$, and $\phi_S = 0.068$. We use estimates for the return to experience and experience squared from Klenow and Rodgriguez-Clare (1997), assigning $\gamma_1 = 0.049$ and $\gamma_2 = 0.0007$.[24]

Given parameter values and the human capital production function, we can calculate the measure of human capital for each state. Table 9 displays summary statistics on human capital per worker for each decade. There is a nearly uninterrupted reduction in the coefficient of variation over time. The only reversals occur between 1920 - 1930, and 1970 – 1980. As with income, we see divergence across states between 1990 and 2000.

[23] If average years of schooling is less than 4, P = *E*, I = 0, S = 0. If average years of schooling is between 4 and 8, P = 4, I = *E* – 4, S = 0. Finally, if average years of schooling is greater than 9, P = 4, I = 4, S = *E* – 8.

[24] This methodology and parameter values are identical to those in Baier, Dwyer, and Tamura (2004).

Table 9 Summary statistics of human capital per worker, labor force weighted.

Year	State	Mean	Std. Dev	Coeff. Var.	Min	Max
1840	28	2.18	0.414	0.190	1.76	3.28
1850	33	2.50	0.526	0.211	1.91	3.89
1860	39	2.71	0.550	0.203	2.00	4.05
1870	43	2.86	0.541	0.189	2.00	3.87
1880	46	3.20	0.573	0.179	2.19	4.12
1890	49	3.49	0.478	0.137	2.46	4.22
1900	49	3.75	0.464	0.124	2.67	4.43
1910	49	3.96	0.462	0.116	2.87	4.69
1920	49	4.39	0.429	0.098	3.30	5.07
1930	49	4.90	0.483	0.099	3.73	5.61
1940	51	5.47	0.422	0.077	4.32	6.07
1950	51	5.65	0.350	0.062	4.73	6.11
1960	51	5.92	0.271	0.046	4.78	6.32
1970	51	6.39	0.214	0.033	5.38	6.77
1980	51	6.68	0.244	0.037	5.84	7.10
1990	51	6.97	0.188	0.027	6.22	7.32
2000	51	7.42	0.210	0.028	6.81	8.07

Again, to give a sense of the data, Figures 10 – 13 show human capital per worker for each census region and for the entire United States. These measures are calculated in the same fashion as the output and capital measures, correcting for entry of new states. The New England, the Mid Atlantic, and the East North Central regions are education leaders throughout the period, while the East South Central, West South Central, and South Atlantic regions are educational laggards. The Mountain region follows the national average quite closely. The West North Central region is above the national average after 1900, but is very near the average by 1980. The Pacific region is leader throughout the period until it falls behind in 1990. It is clear there is convergence in human capital across regions. Table 10 displays the measures for human capital per worker by census region.

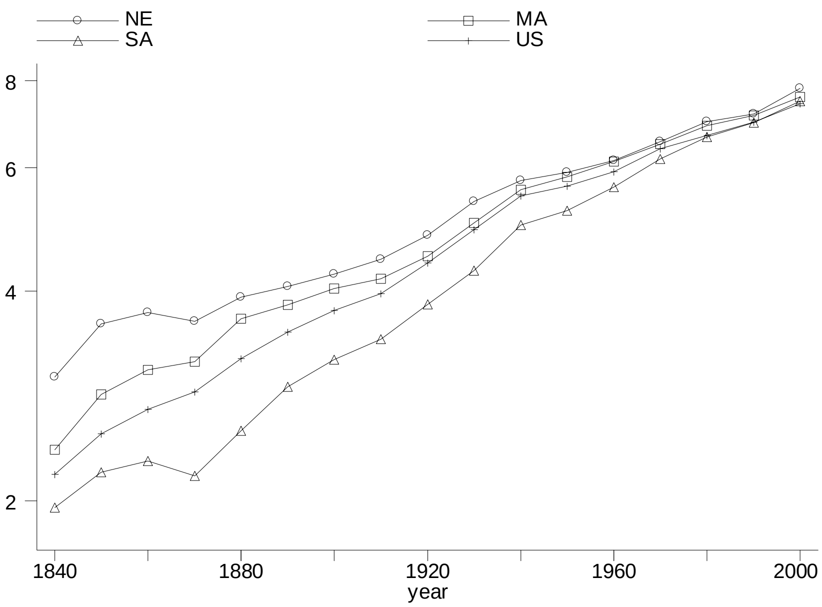


Figure 10 Human capital per worker in New England, Middle Atlantic, South Atlantic, and United States, 1840 – 2000, logarithmic scale.

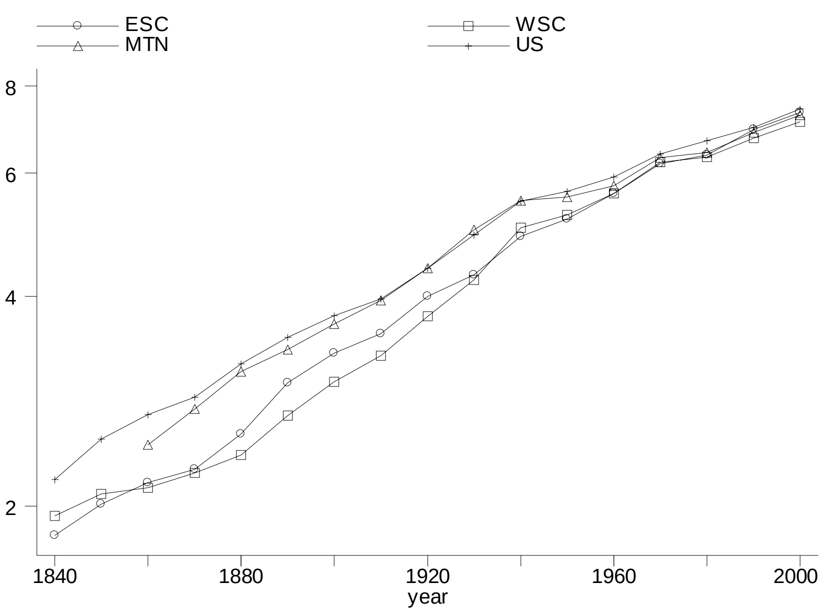


Figure 11 Human capital per worker in East South Central, West South Central, Mountain, and United States, 1840 – 2000, logarithmic scale.

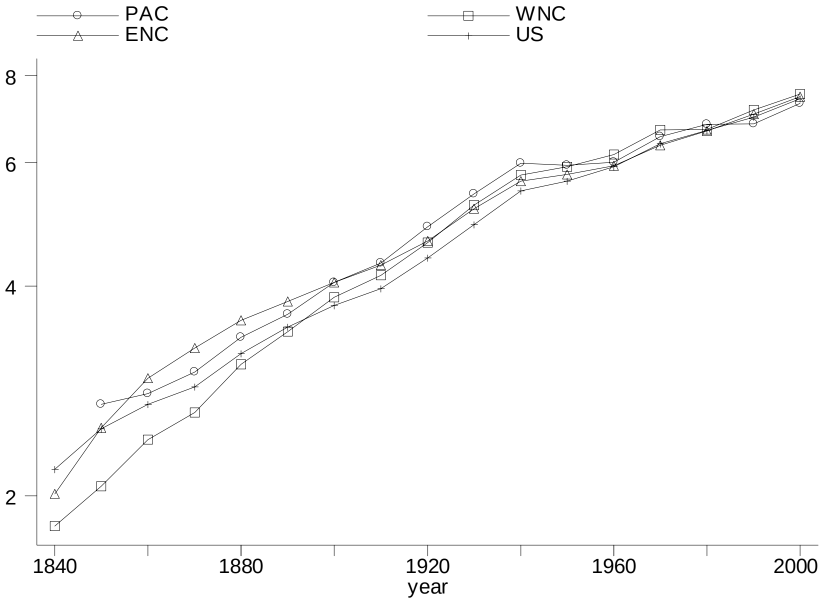


Figure 12 Human capital per worker in Pacific, West North Central, East North Central, and United States, 1840 – 2000, logarithmic scale.

Table 10 Human capital per worker for census regions and United States, labor force weighted

Year	NE	MA	SA	ESC	WSC	MTN	PAC	WNC	ENC	US
1840	**3.01**	2.37	1.95	1.82	1.92	.	.	1.82	2.04	2.18
1850	**3.59**	2.84	2.19	2.01	2.05	.	2.74	2.08	2.50	2.50
1860	**3.72**	3.08	2.28	2.16	2.09	2.14	2.83	2.43	2.95	2.71
1870	**3.62**	3.17	2.17	2.26	2.19	2.75	3.04	2.66	3.26	2.86
1880	**3.92**	3.65	2.52	2.54	2.33	3.13	3.41	3.11	3.57	3.20
1890	**4.06**	3.82	2.91	3.00	2.66	3.36	3.67	3.44	3.80	3.49
1900	**4.23**	4.03	3.19	3.31	2.97	3.65	4.08	3.85	4.04	3.75
1910	**4.44**	4.16	3.41	3.54	3.27	3.94	4.35	4.14	4.29	3.96
1920	4.80	4.49	3.83	3.99	3.74	4.39	**4.91**	4.61	4.64	4.39
1930	5.37	5.00	4.28	4.29	4.22	4.97	**5.47**	5.22	5.16	4.90
1940	5.75	5.58	4.97	4.87	5.01	5.48	**5.96**	5.77	5.65	5.47
1950	5.92	5.82	5.21	5.15	5.23	5.54	**5.95**	5.93	5.78	5.65
1960	6.15	6.13	5.63	5.62	5.62	5.76	6.01	**6.17**	5.95	5.92
1970	6.54	6.48	6.18	6.18	6.22	6.32	6.54	**6.68**	6.36	6.39
1980	**6.99**	6.89	6.64	6.36	6.33	6.42	6.82	6.70	6.66	6.68
1990	**7.16**	7.13	6.96	6.93	6.73	6.86	6.83	7.14	7.05	6.97
2000	**7.80**	7.58	7.47	7.33	7.11	7.27	7.30	7.53	7.46	7.42

Again, to consider convergence, we aggregate into broader regions. Table 11 shows human capital levels for the North, South, and West. From 1860 to 2000, we see only one period, 1920-1930, where the coefficient of variation increases. Looking at the gap between the region with the highest human capital (usually the North) and lowest human capital (always the South) for the same period, we see decreases throughout, with the exception of 1910 – 1930. Figure 13 displays these measures graphically. Clearly the South is the laggard for the entire period, while the North region is a clear leader early. Interestingly, the West region overtakes the North region by 1920, only to fall behind the North region again by 1980.

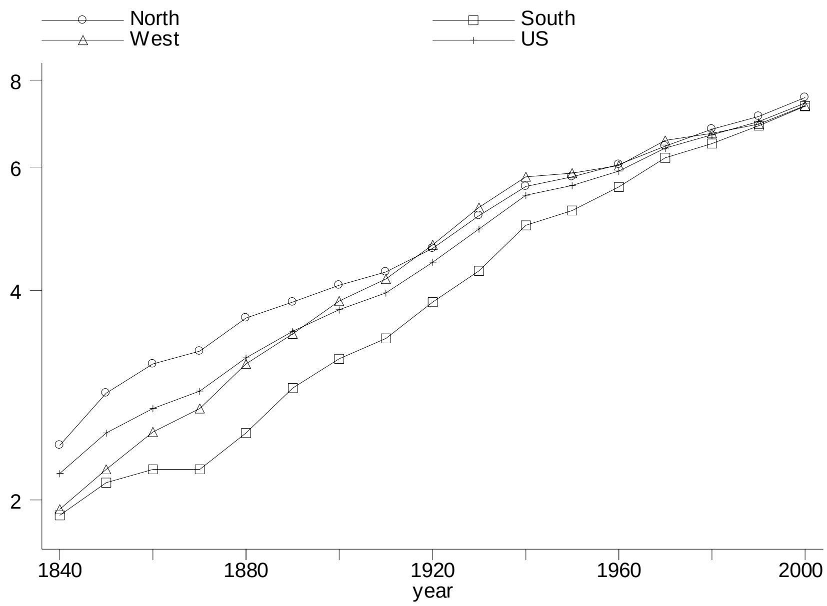


Figure 13 Human capital per worker in North, South, West, and United States, 1840 – 2000, logarithmic scale.

Table 11 Human capital per worker for North, South, West and US, labor force weighted.

Year	North	South	West	US	Std. Dev.	Coeff. Var.
1840	2.45	1.90	1.82	2.18	0.335	0.153
1850	2.89	2.11	2.25	2.50	0.467	0.187
1860	3.17	2.21	2.52	2.71	0.543	0.200
1870	3.29	2.21	2.73	2.86	0.553	0.193
1880	3.66	2.49	3.17	3.20	0.594	0.186
1890	3.85	2.89	3.48	3.49	0.477	0.137
1900	4.07	3.17	3.87	3.75	0.449	0.120
1910	4.25	3.41	4.17	3.96	0.439	0.111
1920	4.60	3.85	4.66	4.39	0.398	0.091
1930	5.12	4.26	5.27	4.90	0.468	0.095
1940	5.63	4.96	5.81	5.47	0.384	0.070
1950	5.82	5.20	5.88	5.65	0.322	0.057
1960	6.05	5.62	6.03	5.92	0.206	0.035
1970	6.44	6.19	6.55	6.39	0.166	0.026
1980	6.80	6.49	6.71	6.68	0.140	0.021
1990	7.10	6.88	6.92	6.97	0.104	0.015
2000	7.55	7.34	7.35	7.42	0.108	0.015

Land Per Worker

In a previous version of this work, we had considered only physical capital and human capital. Especially in the early portion of our dataset, land is an important input into production. At a basic level, the quantity of land available to workers in a state is fixed, the population of each state is increasing, and thus there is a decreasing amount of land per worker. We found that failing to account for the general decrease in land per worker created estimates of TFP growth that were implausibly low for the first hundred years of our data. Utilizing total land area in a state was also unsatisfactory, as there a sizable differences in the arability of land across states. As a result, we proxy land use in each state by using the series of "Land in Farms" which is derived from various issues of the Census of Agriculture and reported in *Historical Statistics of the United States*[25]. For brevity, we do not illustrate the time path of land per worker, but the averages for each census region are reported in Table 12.

[25] Land in farms is not available in 1840. We assume that the amount of land in farms in 1840 is identical to its value in 1850.

Table 12 Land per worker (acres) for census regions and United States, labor force weighted

Year	NE	MA	SA	ESC	WSC	MTN	PAC	WNC	ENC	US
1840	24.7	21.6	40.6	47.2	23.0	.	.	52.7	42.4	35.2
1850	23.4	21.7	40.9	45.9	46.2	16.4	53.1	52.6	42.5	36.6
1860	18.2	16.8	43.4	45.4	59.5	36.3	45.8	53.3	39.3	36.3
1870	15.1	15.4	46.3	45.2	52.5	13.2	51.8	44.7	31.9	32.7
1880	13.7	12.4	38.8	41.0	49.4	14.0	46.9	49.9	29.3	30.8
1890	9.8	8.7	33.2	36.8	52.1	29.5	38.9	50.5	22.6	27.5
1900	8.6	7.2	26.9	28.1	76.0	69.9	45.6	54.4	19.8	29.0
1910	6.8	5.3	20.6	22.6	48.2	53.7	26.5	52.3	16.2	23.1
1920	5.3	4.4	19.2	23.8	46.7	93.5	23.3	56.0	13.8	23.1
1930	4.2	3.2	14.9	19.5	40.7	112.9	16.9	52.5	11.0	20.3
1940	4.0	3.0	14.7	20.9	44.2	136.5	16.4	56.2	11.3	21.5
1950	3.2	2.5	13.0	19.6	39.6	132.2	12.3	51.3	9.0	19.4
1960	2.1	1.9	8.5	15.8	33.1	104.4	9.3	47.7	7.2	16.1
1970	1.1	1.3	5.4	12.2	26.7	77.2	6.7	41.8	5.7	12.7
1980	0.8	1.1	3.5	7.9	17.4	46.3	4.6	32.1	4.5	9.4
1990	0.6	0.9	2.3	6.1	13.9	34.9	3.4	28.1	3.9	7.6
2000	0.5	0.8	2.0	5.4	12.2	25.4	2.7	25.6	3.5	6.7

Growth Accounting

We assume an aggregate production function characterizes the relationship between productive resources and output, and further assume TFP is Hicks-neutral:

$$Y(t) = A(t)F(K(t), N(t), H(t))$$

where $Y(t)$, $K(t)$, $N(t)$ and $H(t)$ are output per worker, physical capital per worker, land per worker, and human capital per worker in period t. $A(t)$ denotes the level of technology in period t. We make the twin assumptions that private marginal products are equal to social marginal products and perfect competition characterizes markets. Then, it can be shown, suppressing subscripts for each state and time period, that:

$$a = y - \alpha k - \beta n - (1 - \alpha - \beta)h$$

where lowercase variables represent growth rates of variables, α denotes physical capital's share of income, and β denotes labor's share of income. We assume that physical capital's share of income is 0.283 and that labor's share of income is 0.05.

Of course, TFP per worker is simply the residual portion of the growth rate of income that is not be explained by the growth in measured inputs. This residual should be interpreted cautiously, for what is being called TFP may actually be identified as technological progress, or instead could result from mismeasurement of income, human capital, physical capital, or simply a failure of the assumptions of perfect competition and constant returns to scale. There may be interactions between technological progress and the stocks of human and physical capital. There are certainly other interpretations as well.

Figures 14 - 16 display TFP for each census regions and the nation as a whole. These figures are constructed in the same fashion as the income and human capital measures, correcting for states entering the region.[26] What seems obvious from these figures is that those census regions in the South display dramatic increases in TFP over the period, surely some of which is related to the Civil War, but much of which occurs after 1900. The Pacific and Mountain regions begin with very high levels of TFP that likely reflect transitory mining income that is not captured by physical capital or human capital. While the New England regions display high levels of TFP early in the period. One could argue that the Civil War or reaction in the post war period has had a lasting effect on the East South Central and West South Central. There seems to be a structural break in TFP series in 1940. This may be exaggerated by business cycles, but this break seems to be quite pervasive.[27]

[26] As we have normalized the value of human capital associated with no experience and no years of schooling, there is implicitly a normalization with calculation of TFP. For each state (region) in the year 2000, TFP is calculated as $TFP_{2000} = Y_{2000} / (K_{2000}^{0.283}\ N_{2000}^{0.05}\ H_{2000}^{0.67})$.

[27] One possibility we considered in explaining the TFP break in 1940 was a change in reporting in the series that we use to construct our measures of human capital, physical capital or income. None of our data sources change significantly in 1940.

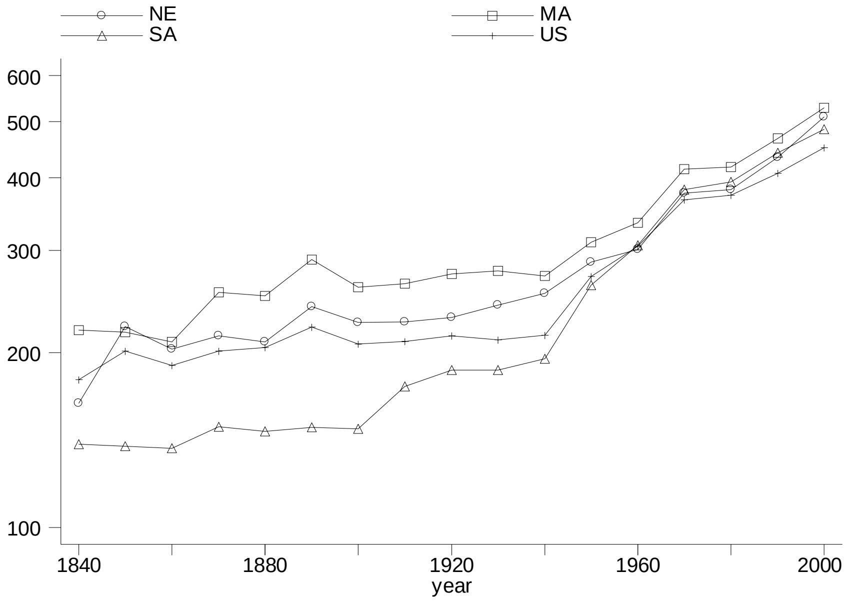


Figure 14 TFP per worker in New England, Middle Atlantic, South Atlantic, and United States, 1840 – 2000, logarithmic scale.

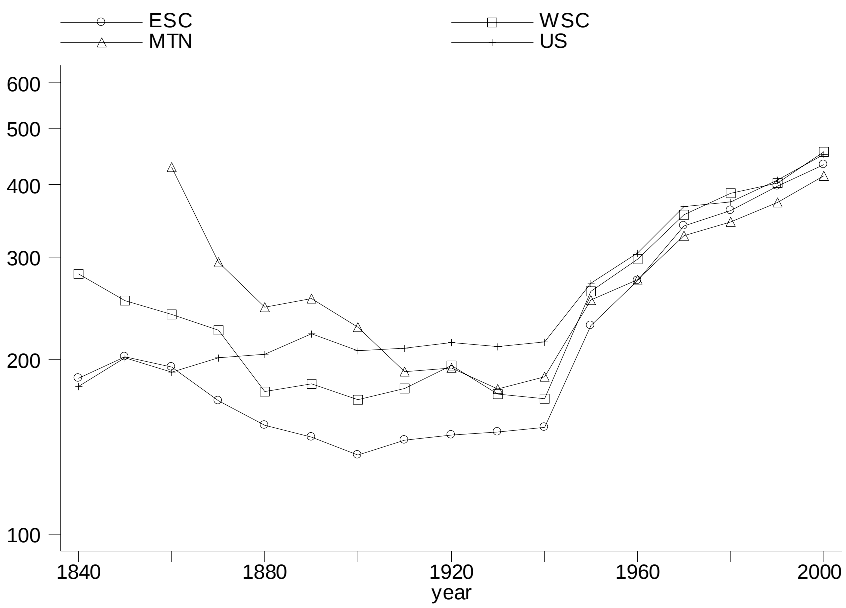


Figure 15 TFP per worker in East South Central, West South Central, Mountain, and United States, 1840 – 2000, logarithmic scale.

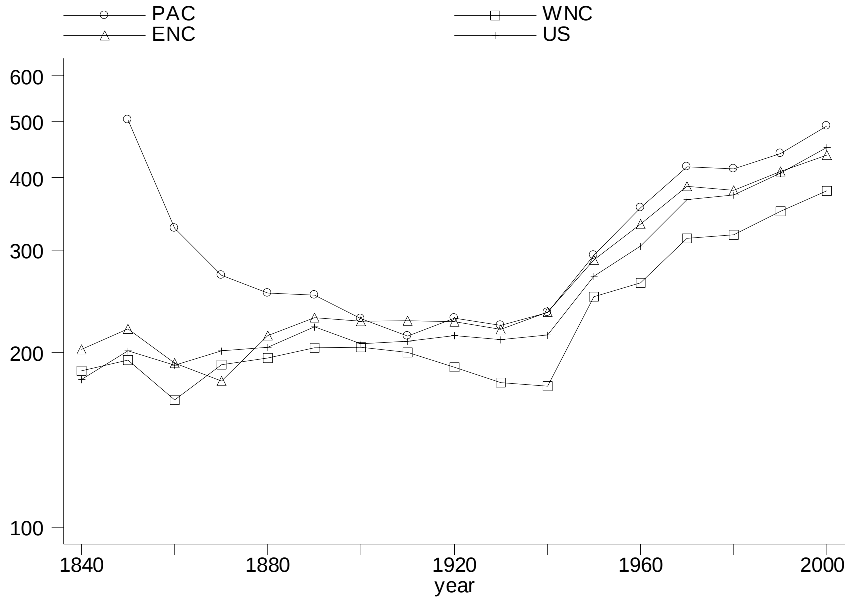


Figure 16 TFP per worker in Pacific, East North Central, West North Central, and United States, 1840 – 2000, logarithmic scale.

Figure 17 displays TFP for the North, South, and West. There is evidence for convergence in these broader regions. This appears to be primarily driven by the South, and seems to have occurred quite rapidly between 1940 and 1970 after a long period of little convergence. Aggregated into broader regions, we see an acceleration of TFP growth in 1940 in both the South and the West, and arguably the North.

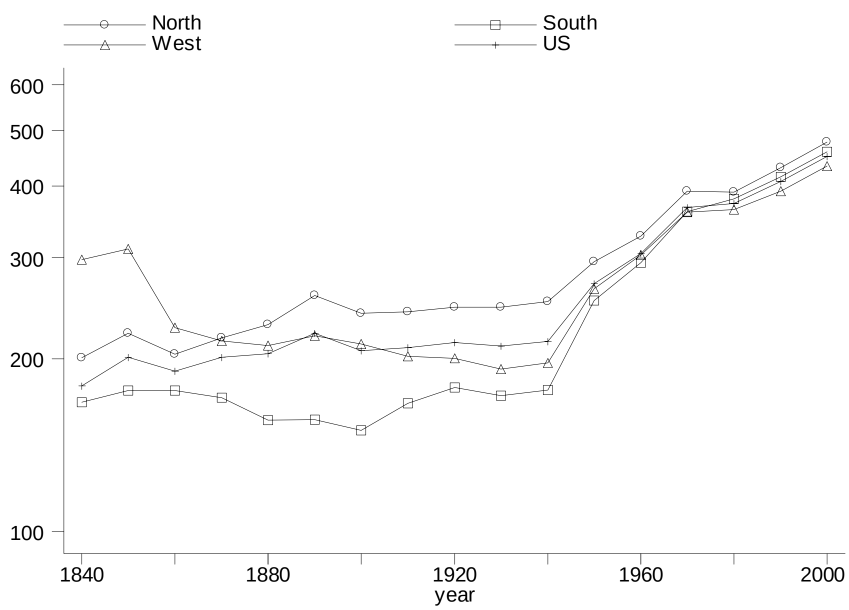


Figure 17 TFP per worker in North, South, West, and United States, 1840 – 2000, logarithmic scale.

Returning to rates of growth, Table 13 shows summary statistics for the average growth rates of real income per worker, human capital per worker, and TFP per worker. The data is reported on an annualized basis, and is calculated using the entire period for each state in which we have human capital, physical capital, land, and income measures. The number of years from which this figure is calculated varies from region to region, depending on data availability.[28] These figures are not weighted by labor force, nor weighted by the number of years included.[29] For the period 1840 - 2000, the annualized rate of growth of output per worker is 1.40% for the United States as a whole. Of this 1.40% annual growth in income per worker, 0.50% (36%) is explained by TFP, while 0.90% (64%) is explained by aggregate input growth. There is variation in growth rates across regions. The East South Central and South Atlantic census regions have the highest growth rates, while the Pacific and Mountain census regions have the lowest.[30] The New England census region has the highest fraction of growth explained by TFP (46%), while the Mountain census region has the lowest fraction of growth explained by TFP (26%). When considering broader regions, the South displays faster output and human capital growth rate than the North, fueling convergence in both. The fraction of growth explained by TFP ranges from 29% to 39% across these broader regions. Overall, the input measures explain a significant fraction of output growth, but a meaningful fraction remains.

---

[28] For instance, for the NE census region, we have data on all inputs and income beginning in 1840. The growth rate reported would be [log($income_{2000}$) – log($income_{1840}$)] / 160. However, we do not observe income and human capital for the Pacific census region until 1850. Hence, the growth rate reported for would be [log($income_{2000}$) – log($income_{1850}$)] / 150.

[29] As a result, the average value reported for the United States is not necessarily the growth rate of output in the United States as a whole. A state with a small labor force receives the same weight as a state with a large labor force in the calculation reported in Table 12. Similarly, a state for which we have data beginning in 1840 will receive the same weight as a state for which we have data beginning in 1920.

[30] As these calculations utilize information only in 2000 and the year in which income, human capital, physical capital are first observed, they are very sensitive to this initial condition. This issue is explored further in a later portion of this paper. Thus, the high estimate of income per worker in 1850 for the Pacific census region has a large impact on the annualized growth rate calculated from 1850 – 2000.

Table 13 Average growth of output and inputs by region, 1840– 2000

Region	Output	Capital	Growth Rate Per Worker Human Capital	Land	TFP	TFP growth relative to Output Growth
All Regions	1.40%	1.58%	0.78%	-1.31%	0.50%	0.357
NE	1.45%	1.74%	0.60%	-2.34%	0.67%	0.462
MATL	1.38%	1.67%	0.74%	-2.25%	0.53%	0.383
SATL	1.72%	1.79%	0.83%	-1.74%	0.74%	0.431
ESC	1.55%	1.84%	0.87%	-1.31%	0.52%	0.333
WSC	1.44%	1.58%	0.83%	-1.09%	0.50%	0.344
MTN	1.20%	1.37%	0.73%	0.21%	0.31%	0.260
PAC	1.08%	1.16%	0.80%	-2.12%	0.33%	0.304
WNC	1.33%	1.41%	0.82%	-0.52%	0.41%	0.308
ENC	1.41%	1.76%	0.82%	-1.59%	0.45%	0.318
North	1.42%	1.73%	0.71%	-2.05%	0.56%	0.394
South	1.61%	1.75%	0.84%	-1.49%	0.63%	0.390
West	1.22%	1.33%	0.78%	-0.63%	0.35%	0.288

Table 14 duplicates the methodology of Table 13, but chooses 1880 as an initial condition rather than 1840. For the United States as a whole from 1880 - 2000, the growth rate of output is slightly higher, 1.58% per year, and now 44% of output growth is explained by TFP growth. Census regions included in the South display the fastest growth, while the Mid Atlantic, East North Central, and Mountain regions display slower growth. The regions with the highest and lowest fraction of growth explained by TFP growth are the New England and Mountain regions respectively, while the fraction explained by TFP varies from 38% to 49% between the North, South, and West. The growth rate of physical capital, human capital, and TFP are highest in the Southern region. The input measures explain slightly more than half of output growth.

Table 14 Average growth of output and inputs by region, 1880 – 2000

Region	Output	Capital	Human Capital	Land	TFP	TFP growth relative to Output Growth
	Growth Rate Per Worker					
All Regions	1.58%	1.63%	0.74%	-1.51%	0.69%	0.441
NE	1.41%	1.50%	0.57%	-2.67%	0.74%	0.527
MATL	1.34%	1.51%	0.61%	-2.43%	0.63%	0.467
SATL	1.99%	1.96%	0.86%	-2.13%	0.97%	0.486
ESC	1.92%	1.93%	0.88%	-1.67%	0.87%	0.453
WSC	1.87%	1.86%	0.89%	-1.27%	0.82%	0.435
MTN	1.37%	1.54%	0.71%	0.21%	0.45%	0.330
PAC	1.45%	1.49%	0.75%	-2.22%	0.64%	0.443
WNC	1.41%	1.40%	0.76%	-0.63%	0.54%	0.384
ENC	1.33%	1.42%	0.62%	-1.78%	0.61%	0.454
North	1.37%	1.47%	0.60%	-2.30%	0.67%	0.489
South	1.95%	1.93%	0.87%	-1.82%	0.91%	0.467
West	1.41%	1.48%	0.74%	-0.69%	0.53%	0.378

Estimates of Variability Across States

Klenow and Rodriguez-Clare (1997) and Easterly and Levine (2000) suggest that while the growth rate of TFP does not account for the vast majority of the growth rate of output, the cross-sectional variance of the growth rate of TFP may account for the vast majority of the cross-sectional variance of the growth rate of output. We continue with the production relationship specified above, expressed in terms of growth rates, with y denoting the growth rate of output per worker, k denoting the growth rate of capital per worker, n denoting the growth rate of land per worker, and h representing the growth rate of human capital per worker:

$$y = a + \alpha k + \beta n + (1 - \alpha - \beta)h$$

We define x as the growth rate of aggregate input per worker:

$$x = \alpha k + \beta n + (1 - \alpha - \beta)h$$

which allows us to simply express the output growth as a function of TFP growth and aggregate input growth:

$$y = a + x$$

By the definition of variance we have:

$$Var(y) = Var(x) + 2Cov(x, a) + Var(a)$$

which implies:

$$1 = \frac{Var(x)}{Var(y)} + \frac{Var(a)}{Var(y)} + 2\,\frac{Var(x)^{1/2}Var(a)^{1/2}}{Var(y)}\,\rho_{x,a}$$

where $\rho_{x,a}$ is the correlation between the growth rate of x and the growth rate of a. If TFP growth and aggregate input growth were uncorrelated, the first term would be the fraction of output growth variance caused by aggregate input growth variance, while the second term would be fraction of output growth variance explained by TFP growth variance. However, this correlation is not zero empirically.

There are several alternatives to deal with the covariance term. First, ignore the covariance term entirely. If so, the relative variances will necessarily not sum to unity unless the covariance is in fact zero. A negative covariance term would result in the relative variances summing to a figure in excess of unity. We do not pursue either of these methods.

We follow the methodology of Baier, Dwyer, and Tamura (2004), and make two alternative estimates of the relative variances, and in doing so, we create an upper and lower bound on each source of variance. We alternately assign all of the correlation between aggregate input and TFP to either aggregate input growth, or to TFP growth. This implies that each estimate will have a complement that adds to unity. The first assumes that all changes in output growth that are predictable by aggregate input growth are due to aggregate input growth, or stated differently, that the correlation between input growth and TFP reflects unmeasured effects of input growth. The second assumes that all changes in output growth that are predictable by TFP growth are due to TFP growth, or, assumes the correlation reflects unmeasured effects of TFP growth.

The first decomposition attributes to aggregate input growth all growth of output predictable by aggregate input growth:

$$\frac{\left(SD(x) + SD(a)\rho_{x,a}\right)^2}{Var(y)} + \frac{\left(1 - \rho_{x,a}^2\right)Var(a)}{Var(y)} = 1$$

This assumption creates and upper bound on the fraction of the variance of output growth that can be explained by variance of input growth, and thus creates a lower bound on the fraction of the variance of output growth that can be explained by TFP growth.

The second decomposition attributes to TFP growth all of the growth predictable by TFP growth:

$$\frac{\left(1-\rho_{x,a}^{2}\right)Var(x)}{Var(y)}+\frac{\left(SD(a)+SD(x)\rho_{x,a}\right)^{2}}{Var(y)}=1$$

This assumption then creates the upper bound on the fraction of the variance of output growth that can be explained by TFP growth, and thus a lower bound on the fraction of the variance that can be explained by aggregate input growth.

Results for the Entire United States

Table 15 displays information on the results of the decomposition using data from 1840 – 2000. Columns 3 through 5 report the standard deviations of the growth rates of output per worker, aggregate input per worker, and TFP per worker across states. Columns 6 and 7 display the variances of the growth rate of aggregate input and the growth rate of TFP, each relative to the variances of output growth. The next two columns display the fraction of the cross-sectional variance of output growth that is explained by the variance of input growth under each of the treatments of the observed correlation. The column labeled lower and upper and the lower and upper bounds, respectively. Each of the two methods has a complement that adds to unity, which is the fraction of output growth variance explained by TFP growth, and is displayed in the following two columns. The final column reports the correlation between aggregate input growth and TFP growth.

Table 15 Variance decomposition, average growth rates for 1840 – 2000, period is state specific.[31]

		Standard Deviation			Relative Variance of		Fraction of Variance Explained by Aggregate Input		Frac. of Variance Explained by TFP		Corr (X,TFP)
States	Time Interval	Output	Aggregate Input	TFP	Aggregate Input	TFP	Lower	Upper	Lower	Upper	
US	Varies	0.32%	0.12%	0.24%	0.150	0.592	0.122	0.519	0.481	0.878	0.433
North	Varies	0.20%	0.10%	0.19%	0.247	0.912	0.114	0.240	0.760	0.886	-0.167
South	Varies	0.21%	0.07%	0.19%	0.119	0.821	0.118	0.186	0.814	0.882	0.095
West	Varies	0.35%	0.14%	0.24%	0.162	0.479	0.094	0.720	0.280	0.906	0.645

[31] The observed correlation between the growth rate of TFP and the growth rate of aggregate input is a negative value for the North region. As a result, the first decomposition referred to the in the text no longer creates an upper bound of the fraction of the variance that can be explained by aggregate input. The entry in the columns marked lower and upper in the table above thus represents the lower and upper bounds for the appropriate column, not the outcome of the first and second decompositions, respectively.

Overall, it is clear the variance of the growth rate of TFP is larger than the variance of the growth rate of the aggregate input. For the US as a whole, the variance of the growth rate of aggregate input is only 15% of the variance of the growth rate of income, while the variance of the growth rate of TFP is 59% the variance of the growth rate of income. Under the alternate assumptions for the decomposition, we can explain between 12% and 52% of the variance of income growth from the variance of the aggregate input, leaving between 48% and 88% due to the variance of TFP. The range is a result of an observed correlation of approximately 0.4 between the growth rate of aggregate input and the growth rate of TFP.

We also report the variance decomposition for the broader regions in Table 15. We are hesitant too much stock in the results here. As we have seen, the results are quite sensitive to the correlation between aggregate input growth and TFP growth. In subdividing by broad regions, this correlation is based on as few as 14 observations. Caveat aside, the results vary significantly from region to region, most notably because the observed correlation differs within region. Within the North and South regions, TFP growth explains the vast majority of the variance of output growth, while in the West region, input growth is capable of explaining more of the variance in output growth.

Growth rates are sensitive to the initial conditions. We repeat the analysis, but utilize only data from 1880 – 2000. The results of this decomposition are reported in the Table 16. For the US as a whole, the results are quite different. We find that again the variance of the growth rate of aggregate inputs can explain between 11% and 86% of the variance of income growth, leaving TFP to also explain between 14% and 89%. The results show that the variance decomposition is clearly dependent on the treatment of the observed correlation of aggregate input growth and TFP growth. Depending on how this correlation is allocated, either aggregate input or TFP could be responsible for explaining nearly all of the variance in output growth.

Table 16 Variance decomposition, average growth rates for 1880 – 2000, period is state specific.[32]

		Standard Deviation			Relative Variance of		Fraction of Variance Explained by Aggregate Input		Frac. of Variance Explained by TFP		Corr (X,TFP)
States	Time Interval	Output	Aggregate Input	TFP	Aggregate Input	TFP	Lower	Upper	Lower	Upper	
US	Varies	0.34%	0.17%	0.19%	0.247	0.325	0.106	0.861	0.139	0.894	0.756
North	Varies	0.08%	0.03%	0.08%	0.129	1.203	0.011	0.106	0.894	0.989	-0.422
South	Varies	0.26%	0.14%	0.15%	0.282	0.345	0.182	0.777	0.223	0.818	0.596
West	Varies	0.23%	0.12%	0.14%	0.253	0.347	0.138	0.811	0.189	0.862	0.675

[32] See previous footnote.

Results for Subperiods With Common Data Availability

Given the sensitivity to initial conditions, we conduct an analysis where we restrict our sample to states for which we have data over identical periods. That is, we first consider those 28 states for which we have human capital, physical capital, land, and income data for the entire 1840 – 2000 period. We then conduct the variance decomposition for only these 28 states. Next we repeat the analysis using those same 28 states and the period 1860 – 2000. We continue in 20 year increments until we arrive at 1960. Next, we consider 44 states in which data is available from 1880 – 2000, and then continue in 20 year increments until again we arrive at 1960. We continue in a similar fashion until we have data on all 51 states. The results are reported in Table 17.

The variance decomposition results are noticeably different across subperiods. The relative variance of aggregate input declines nearly uniformly as the time periods get shorter, which tends to decrease the fraction of cross-sectional variance of output growth that could be explained by input growth. However, the correlation between TFP growth and aggregate input growth first increases, reaching a peak when examining the 1940 – 2000 period, then decreases nearly uniformly as the period being examined is shorter. For the 28 states for which we have data for the entire 1840 – 2000 period, this correlation has a peak value of 0.91. With this large observed correlation, the upper and lower bounds of the variance decomposition contain a wide range of possibilities. As the time period gets shorter, both this correlation and the relative variance of aggregate input tend to reduce the amount of output growth's variance that can be explained by aggregate input growth, leaving the lion's share to be explained by TFP growth. Results are reported for all subperiods and groups of state in Table 17.

Table 17 Variance decomposition for subperiods.

			Standard Deviation			Relative Variance of		Fraction of Variance Explained by Aggregate Input		Fraction of Variance Explained by Aggregate TFP		Corr (X,TFP)
Time Period	States		Output	Aggregate Input	TFP	Aggregate Input	TFP	Lower	Upper	Lower	Upper	
1840-2000	28	1840-2000	0.19%	0.11%	0.16%	0.339	0.652	0.339	0.348	0.652	0.661	0.009
		1860-2000	0.26%	0.15%	0.17%	0.351	0.431	0.323	0.603	0.397	0.677	0.280
		1880-2000	0.37%	0.20%	0.20%	0.288	0.294	0.140	0.857	0.143	0.860	0.716
		1900-2000	0.44%	0.20%	0.26%	0.198	0.352	0.054	0.904	0.096	0.946	0.853
		1920-2000	0.39%	0.19%	0.22%	0.234	0.326	0.085	0.881	0.119	0.915	0.797
		1940-2000	0.54%	0.22%	0.33%	0.172	0.368	0.029	0.937	0.063	0.971	0.911
		1960-2000	0.33%	0.11%	0.25%	0.102	0.590	0.062	0.641	0.359	0.938	0.626
		1980-2000	0.41%	0.12%	0.33%	0.086	0.648	0.059	0.557	0.443	0.941	0.562
	44	1880-2000	0.35%	0.18%	0.22%	0.269	0.407	0.205	0.690	0.310	0.795	0.488
		1900-2000	0.42%	0.18%	0.27%	0.187	0.409	0.087	0.809	0.191	0.913	0.730
		1920-2000	0.34%	0.17%	0.20%	0.243	0.352	0.127	0.816	0.184	0.873	0.691
		1940-2000	0.46%	0.21%	0.28%	0.199	0.363	0.067	0.877	0.123	0.933	0.813
		1960-2000	0.33%	0.13%	0.24%	0.158	0.560	0.122	0.566	0.434	0.878	0.474
		1980-2000	0.43%	0.15%	0.34%	0.124	0.650	0.104	0.454	0.546	0.896	0.399
	48	1920-2000	0.34%	0.16%	0.21%	0.232	0.399	0.147	0.748	0.252	0.853	0.607
		1940-2000	0.45%	0.20%	0.27%	0.201	0.364	0.071	0.872	0.128	0.929	0.805
		1960-2000	0.32%	0.13%	0.24%	0.159	0.573	0.127	0.541	0.459	0.873	0.445
		1980-2000	0.45%	0.15%	0.38%	0.113	0.704	0.101	0.371	0.629	0.899	0.326
	51	1960-2000	0.33%	0.14%	0.25%	0.175	0.596	0.153	0.479	0.521	0.847	0.355
		1980-2000	0.55%	0.18%	0.45%	0.102	0.647	0.078	0.507	0.493	0.922	0.488

Conclusion

This paper creates and utilizes a new state level physical capital measure, and utilizes this measure to conduct a growth accounting for the United States from 1840 – 2000. We find that our measure of aggregate input is able to account for 64% of average output growth per worker, leaving 36% to TFP growth. The measure of aggregate input growth explains a large fraction of output growth, but a significant fraction remains. TFP growth rates are different across census regions.

We find that conclusions concerning the fraction of the cross sectional variance of income growth that is explained by TFP growth and aggregate inputs depend highly on the treatment of the observed correlation between TFP growth and aggregate input growth and sensitive to time periods considered. For longer periods, the upper and lower bounds created suggest that variance in TFP growth is able to account for the bulk of the variance in output growth. This is suggestive that something other than institutional heterogeneity may be responsible for difference in TFP.

Interestingly, TFP growth seems to accelerate across the entire United States in 1940. Before 1940, TFP growth rates are substantially lower than those observed after 1940. As a consequence, TFP growth comprises a much larger fraction output growth in the post-1940 period than in the pre-1940 period. In future work, we endeavor to identify the source of this structural break. In addition, in light of the TFP differences persisting despite relatively homogeneous institution across states, we feel future research should be directed at explaining differences in TFP across states. We have begun to consider institutional heterogeneity between states, most notably the institution of slavery and the subsequent variance in educational attainment by race. Other candidates we intend to pursue are education quality differences, particularly higher education quality, tax differences, and sectoral labor allocation differences. Finally, we plan to replicate the methodology of Klenow and Rodriguez-Clare (1997) to make the state results we have comparable to the results they report across countries.

References Cited

Abramovitz, Moses. 1956. "Resource and Output Trends in the United States since 1870." *American Economic Review Papers and Proceedings* 46 (May), pp. 5-23.

Barro, Robert J., and Jong-Wha Lee. 1993. "International Comparisons of Educational Attainment." *Journal of Monetary Economics* 32 (December), pp. 363-94.

Baier, Scott, Sean Mulholland, Robert Tamura, and Chad Turner. 2005. "Income and Education of the States of the United States: 1840 – 2000", Clemson University working paper.

Baier, Scott, Gerald Dwyer, and Robert Tamura. 2004. "How Important Are Capital and Total Factor Productivity for Economic Growth?", Clemson University working paper.

Benhabib and Spiegel. 1994.

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

Dension, Edward F. 1969.

Denison, Edward F. 1985. *Trends in Economic Growth, 1929-1982*. Washington: Brookings Institution.

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

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

Easterlin, Richard. 1960. "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.

Easterly, William and Ross Levine. 2001. It's Not Factor Accumulation: Stylized Facts and Growth Models." Unpublished paper, World Bank.

Gallman, Robert E. 1960. "Commodity Output in the United States," in Conference on Income and Wealth, *Trends in the American Economy in the Nineteenth Century*, 24, *Studies in Income and Wealth*. Princeton University Press: Princeton.

Gordon, Robert. 2002. *Macroeconomics* (8th edition), Pearson Education.

Jones, Charles I. 1997. "On the Evolution of the World Income Distribution." *Journal of Economic Perspectives* 11 (Summer) pp. 19-36.

Joregenson, Dale W., and Zvi Griliches. 1967. "The Explanation of Productivity Change." *Review of Economic Studies* 34 (July), pp. 249-83.

Klenow, Peter J., and Andres Rodriguez-Clare. 1997. "The Neoclassical Revival in Growth Economics: Has it Gone Too Far?" In NBER *Macroeconomics Annual 1997*, ed. By Ben S. Bernanke and Julio J. Rotemberg, pp. 73-103. Cambridge: MIT Press.

Kuznets. 1966.

Lucas. 1988.

Maddison, Angus. 1995. *Monitoring the World Economy, 1820-1992*. Paris: Development Centre of the Organization for Economic Co-operation and Development.

Mankiw, N. Gregory, David Romer and David N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107 (May), pp. 407-37.

Mitchener, Kris J. and McLean, Ian W. 1999. "U.S. Regional Growth and Convergence 1880 –1980." *The Journal of Economic History*, 59, pp. 1016-42.

Romer. 1986.

Solow. 1956.

Solow. 1957.

Summer-Heston.

Temple. 1999.

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

U.S. Census Bureau. Census of Population, various issues, U.S. Government Printing Office, Washington, D.C.

Williamson, Jeffrey G. and Linder, Peter H. 1980. *American Inequality*, Academic Press.

Appendix A

Regional Divisions

There are nine census regions in the US. Table A.1 provides the regional divisions.

Table A.1 List of states contained in each census region

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				

Year in which income, physical capital, and human capital data is available for each state:

1840: Alabama, Arkansas, Connecticut, Delaware, Florida, Georgia, Illinois, Indiana, Iowa, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Mississippi, Missouri, New Hampshire, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Rhode Island, South Carolina, Tennessee, Vermont, Virginia

1850: California, New Mexico, Oregon, Texas, Utah, Wisconsin

1860: Kansas, Minnesota, Nebraska, Washington, New Mexico*, Utah*

1870: Colorado, Montana, Nevada, West Virginia

1880: Arizona, Idaho

1900: North Dakota, South Dakota, Wyoming

1920: Oklahoma

1930: District of Columbia

1950: Alaska

*Income and physical capital data available in 1850, but human capital data not available until 1860.

Appendix B

Employment mapping, 1840 – 1890

	1840[2]	1850[3]	1860	1870[4]	1880	1890
Agriculture	Agriculture	Agriculture		Agriculture	Agricultural Pursuits	Agricultural Pursuits
Mining & Manufacturing	Manufactures and Trades, Mining	Commerce Trade Manufactures Mechanic Arts and Mining		Manufactures and Mechanical and Mining	Manufacturing and Mechanical Pursuits	Manufacturing and Mechanical Pursuits
Transportation & Public Utilities	Navigation of the Ocean, Navigation of Canals Lakes and Seas	Sea and River Navigation		1	1	1
Commerce	Commerce, Learned Professions and Engineers	Law Medicine & Divinity, Other Pursuits Requiring Education, Labor Not Agricultural, Domestic Servants		Profession and Personal Services, Trade and Transportation	Professional Service, Domestic and Personal Service, Trade & Transportation	Professional Service, Domestic and Personal Service, Trade & Transportation
Government & Education	1	Army, Government Civil Service		1	1	1

Notes:

1 Capital for this industry included with capital in Commerce industry.
2 Includes females and slaves.
3 Includes males age 15 or older.
4 Includes all people age 10 or older.

Employment mapping, 1900 – 1950

	1900[5]	1910[5]	1920[5]	1930[5]	1940[4,6]	1950[6]
Agriculture & Construction	Agricultural Pursuits	Agriculture Forestry and Animal Husbandry	Agriculture Forestry and Animal Husbandry	Agriculture Forestry & Fishing	Agriculture Forestry & Fishery, Construction	Agriculture Forestry & Fisheries, Construction
Mining	[7]	Extraction of Minerals	Extraction of Minerals	Extraction of Minerals	Mining	Mining
Manufacturing	Manufacturing & Mechanical Pursuits	Manufacturing & Mechanical Pursuits	Manufacturing & Mechanical Pursuits	Manufacturing & Mechanical Industries	Manufacturing	Manufacturing
Transportation	Trade & Transporation[1,2]	Transportation[1]	Transportation[1]	Transportation & Coummunication[1]	Transportation Communication & Other Utilities[1]	Transportation Communication & Other Utilities[1]
Communications	Trade & Transportation[1,2]	Transportation[1]	Transportation[1]	Transportation & Coummunication[1]	Transportation Communication & Other Utilities[1]	Transportation Communication & Other Utilities[1]
Utilities	Trade & Transportation[1,2]	Transportation[1]	Transportation[1]	Transportation & Coummunication[1]	Transportation Communication & Other Utilities[1]	Transportation Communication & Other Utilities[1]
Trade	Trade & Transporation[1,2]	Trade[3]	Trade[3]	Trade[3]	Wholesale and Retail Trade	Wholesale and Retail Trade
FIRE	Trade & Transporation[1,2]	Trade[3]	Trade[3]	Trade[3]	FIRE	FIRE
Services	Professional Service, Domestic and Personal Service	Professional Service, Domestic and Personal Service, Clerical Occupations	Professional Service, Domestic and Personal Service, Clerical Occupations	Professional Service, Domestic and Personal Service, Clerical Occupations	Business and Repair Services, Personal Services, Amusement Recreation and Related Services, Professional and Related Services	Business and Repair Services, Personal Services, Entertainment and Recreation Services, Professional and Related Services
Government		Public Service	Public Service	Public Service	Government	Public Administration

Notes:

1 Employment in this category was split into components (Transportation, Communications, and Utilities) by assuming the fraction of the total in each subgroup is identical to the fraction observed in 1970.

2 Prior to the splitting into components per footnote 1, the Trade & Transportation is split into components (Trade, Transportation) by assuming the fraction of the total in each subgroup is identical to the fraction observed in 1910.

3 Employment in this category was split into components (Trade, Fire) by assuming the fraction of the total in each subgroup is identical to the fraction observed in 1940.

4 Employment figures do not include those on public emergency work

5 Includes age 10 or above

6 Includes age 14 or above

7 Capital for this industry included with capital in Manufacturing.

Employment mapping, 1960 – 2000

	1960	1970	1980	1990	2000
Agriculture & Construction	Agriculture Forestry & Fisheries, Construction	Agriculture Forestry & Fisheries, Construction	Agriculture, Forest and Fisheries, Construction	Agriculture, Forest and Fisheries, Construction	Agriculture Forestry Fishing and Hunting, Construction
Mining	Mining	Mining	Mining	Mining	Mining
Manufacturing	Manufacturing	Manufacturing	Manufacturing	Manufacturing	Manufacturing
Transportation	Transportation Communication & Other Utilities[1]	Railroads and Railway Express Service, Trucking Service and Warehousing, Other Transportation	Transportation Communication & Other Utilities[1]	Transportation Communication & Other Utilities[1]	Transportation & Warehousing
Communications	Transportation Communication & Other Utilities[1]	Communications	Transportation Communication & Other Utilities[1]	Transportation Communication & Other Utilities[1]	Information
Utilities	Transportation Communication & Other Utilities[1]	Utilities and Sanitary Services	Transportation Communication & Other Utilities[1]	Transportation Communication & Other Utilities[1]	Utilities
Trade	Wholesale and Retail Trade	Wholesale Trade, Retail Trade, General Merchandise Retailing, Motor Vehicles Retailing and Service Stations, Other Retail Trade	Wholesale Trade, Retail Trade	Wholesale Trade, Retail Trade	Wholesale Trade, Retail Trade
FIRE	FIRE	Banking and Credit Agencies, Insurance Real Estate and Other Finance	FIRE	FIRE	FIRE
Services	Business and Repair Services, Personal Services, Entertainment and Recreation Servies, Professional and Related Services	Bus. Serv., Repair Serv., Pri. Households, Other Personal Serv., Entertain and Rec. Serv., Hospital, Health Serv. Except Hospitals, Elem. and Sec. Schools and Colleges, Other Ed. and Kindred Serv., Welfare Religious and Nonprofit Membership Organizations, Legal Engineering and Misc. Professional Serv.	Services	Services	Professional Scientific Management Administrative and Waste Management Services, Educational Health and Social Services, Arts Entertainment Recreation Accommodation and Food Services, Other Services (Except Public Administration)
Government	Public Administration	Public Administration	Public Administration	Public Administration	Public Administration