U.S. Historical National Accounts 1929–1947: Sources & Methods

Gerben Bakker, Nicholas Crafts and Pieter Woltjer*

This appendix describes the methods and sources which we use to develop new, industry level estimates extending beyond Kendrick's original 1929-37 figures. These new estimates allow us to measure productivity growth over the period suggested by Field, 1929-41, while matching the full sectoral detail realized by Kendrick. We construct new estimates of output, labor- and capital input for the American economy between 1929 and 1947.

As emphasized by Field (2003), the assessment of productivity trends during the 1930s is highly sensitive to the choice of beginning- and end-point. To prevent cyclical effects from influencing the measurement of productivity growth it is best to choose business-cycle peaks as reference years. The choice of Kendrick (1961) to comparing the depressed American economy in 1937 to the peak-year of 1929 conflicts with this principle. Field (2003, 1403) argues instead that 1941, with an unemployment rate of 9.9 percent, compares much more favorably to the fully employed economy of 1929 than the year 1937 (14.3 percent unemployment). Regrettably, little productivity data is available – at least not beyond the total economy trends – for the early 1940s. This led Field (2006) to restrict his analysis of technological change between 1929 and 1941 to a 4-sector breakdown of TFP growth. We require a much finer breakdown to fully decompose the sectoral contribution to TFP and labor productivity growth.[1]

1. *Output*

Instead of estimating value added ('VA')[2] based on industry output less purchases of materials and services, we obtain nominal gross value added by summing over total compensation, gross operating surplus, and taxes on production less subsidies. The components of value added at the industry level are compiled by the U.S. Bureau of Economic Analysis (BEA, 2009) and listed in the National Income and Product Accounts (NIPA). Table 1 provides an overview of the relevant variables, the exact source-tables, the number of industries differentiated, as well as the share of value added covered by each respective variable in the year 1947.

The NIPA tables provide annual data at the industry level, allowing us to estimate net output for a set of 35 (disaggregate) industries, completely covering the domestic economy. As illustrated in table 1, the NIPA provides full industry coverage for the most influential variables which, together, make up over 80 percent of gross value added. For the remaining variables the BEA supplies data at a higher level of aggregation, distinguishing between either 12 separate industries (e.g. proprietors' income) or listing the total-economy value only (e.g. taxes on pro-

* Pieter Woltjer: Wageningen University, Hollandseweg 1, 6706KN Wageningen, The Netherlands, p.j.woltjer@gmail.com.

[1] Methods and sources for Bakker, Crafts and Woltjer (2015).

[2] For a complete list of variables in the database see the appendix on p. 13.

3

TABLE 1—COMPONENTS OF VALUE ADDED BY INDUSTRY, U.S. 1929-1947

Variable	NIPA	Cov.	Shr.
va	...	...	100
.comp	6.2A	35	54
.txpixs	1.7.5 L18	1	7
..nint	6.15A	12	1
..proinc	6.12A	12	9
..frminc	2.1 L10	1	6
..pbt	6.17A	35	13
..ccca	6.22A	35	3
..ncca	6.13A	12	3
..bctp	7.7 L1	1	0
..iva	6.14A	12	-3
..ccadj	7.6	1	0
..gcfc	7.5 L21	1	4
..rip	2.1 L11	1	3

Legend: *va*: value added; *comp*: compensation of employees; *txpixs*: taxes on production less subsidies; *nint*: net interest; *proinc/frminc*: proprietors' income, nonfarm & farm; *pbt*: corporate profits before tax; *ncca/ccca*: (non-)corporate capital consumption allowance; *bctp*: business current transfer payments; *iva*: inventory valuation adjustment; *ccadj*: capital consumption adjustment; *gcfc*: consumption fixed capital government; *rip*: rental income of persons, FIRE
Notes: NIPA lists the source table and line (L); Cov. indicates no. of industries distinguished in source; 35 means full coverage. Shr. indicates the % share of total economy value added covered in 1947.
Source: BEA (2009, 2011).

duction less subsidies). For these variables, we use the detailed industry-level data for the components of value added in 1947 – taken from the BEA (2011) *Historical Industry Accounts Data* – to distribute the aggregate figures over our complete list of industries.

To obtain real value added ('VA_Qi') we deflate the nominal output figures for agricultural, mining, manufacturing, utilities and wholesale trade on the basis of wholesale prices ('VA_Pi') compiled by the U.S. Bureau of Labor Statistics (BLS, 1943, 1949, 1958) supplemented with the production prices listed in the *Historical Statistics of the United States* (Carter et al., 2006, 582-6) and the price index of electrical equipment compiled by the BEA (2010). For the remaining service sectors we apply the relevant price indices for personal consumption expenditure from BEA (1966, tab. 8.6), BEA (2009, tab. 1.5.4) and Kendrick (1961, 543-5, 556, 583-4). We aggregate the price deflators over industries on the basis of an annually chained Fisher index, where nominal gross value added, previously discussed, serves as weights.

2. *Labor input*

For labor input we rely on estimates of total employment ('PEP') by industry, fully compensated for changes in the average annual hours of work ('H_avg') and the growth in the quality of labor ('LQ'). The sources for total employment and the average hours of work are discussed below. The adjustment for labor quality is dealt with in section 4.

In correspondence with Kendrick (1961, 47-9), we define total employment as the sum of the number of employees, converted to a full-time equivalent basis, and self-employed persons. From 1929 onward, the BEA (2009, tab. 6.8A) lists this statistic as the total Persons Engaged in Production at the detailed industry level.

Estimates for the average annual hours of work between 1929 and 1947 for the majority of industries are based on Kendrick (1961, 310, 360-2, 397-8, 543-7, 556, 583-4, 590-8, 611). For construction, other transportation and trade we rely on the Bureau of the Census (1975, 170-3) estimates of changes in the weekly hours of work. In addition, we accounted for differences in the average hours of work in durable and non-durable manufacturing based on data from the Bureau of the Census (1975, 169-70), normalized to fit the total manufacturing estimates by Kendrick (1961, 465-6). Our final measure of labor services ('LAB_Qi') is then derived by multiplying total employment by both the index for the change in the average annual hours of work as well as the index for labor quality (see section 4 below).

3. *Capital input*

For the period 1929 to 1947 we estimate the capital input on the basis of capital services. As opposed to capital stocks, which measure the total value, or wealth of all capital equipment and structures in place, our measure captures the capital service *flows* derived from these capital assets. The difference between both these methods is that capital services weight the growth of capital assets by their respective rental prices, whereas capital stocks weight assets by their asset price. As noted by Jorgenson, Gollop and Fraumeni (1999, 109), "[c]apital input takes the form of services of the capital stock in the same way that labor input involves the services of the work force", making the resulting capital service indices strictly comparable to the measure of labor input discussed above.

In comparison to the stock measure, the capital service flows will allocate greater weight to assets that have shorter asset lifetimes and/or rapidly falling asset prices, as both characteristics will drive up the cost a user would have to pay to hire the asset for a given period. In the 1930s, prime examples of assets that are under-weighted by the traditional capital stock measure are: communication equipment, office and accounting equipment and trucks.

Our capital services differ from the measure of capital adopted by Kendrick (1961), but is consistent with the post-war estimates of capital input by the BLS. This thus allows us to directly compare the 1929-1947 residual in our growth accounting exercise to the official TFP estimates for the decades following the war.

The construction of the indices of capital services ('CAP_Qi') proceeds in two phases. First, we estimate the industry-level stock of capital ('K_Qi') for the private domestic economy between 1929 and 1947 using a Perpetual Inventory Method (PIM) and the investment series taken from the BEA (1993, 2002, 2003, 2010) *Fixed Assets* tables. Second, we estimate the rental price of assets at the industry level based on the imputed industry rate of return, the asset-specific rate of depreciation and capital gains and losses resulting from changing asset prices. Multiplying the stock of an asset by its rental price yields so called 'capital compensation' ('CAP'), which in turn can be used as weights to aggregate the capital stocks to the industry and ultimately the total economy level.

CAPITAL STOCKS

For the construction of the capital stocks ('K_Qi'), we follow the approach set out by the BEA (2003, M-7), where the real investment ($I_{i,k}$) for asset k during year i is assumed to contribute $N_{i,k,t}$ to the real net stock of capital at the end of year t.

$$N_{i,k,t} = I_{i,k}\left(1 - \frac{\delta_k}{2}\right)(1-\delta_k)^{t-i} \tag{1}$$

All investments are expected to have been made during the middle of the calendar year, and are depreciated at an annual geometric rate of depreciation δ_k. By summing the contributions over all investments up to and including year t, the real net stock of capital ($N_{i,k}$) for asset k at the end of year t can be derived.

$$N_{i,k} = \sum_{i=1}^{t} N_{i,k,t} \tag{2}$$

From 1901 onward, the BEA (2010) detailed *Fixed Assets* tables provide annual industry-by-asset investment series for private nonresidential capital. To reliably estimate the starting stock of capital in 1900, we supplement this data with the asset-specific constant-cost investment series for the period 1832-1900, listed in the BEA (1993, 374-81) *Fixed Reproducible Tangible Wealth* report. Unfortunately, the pre-1901 investment series is only available at the total private economy level. We thus distribute the nineteenth-century investment data for each of the 37 assets over our entire industry-list on the basis of the average investment shares for the first decade in the twentieth century – for which we have detailed industry-by-asset data. The geometric rates of depreciation for all our assets, with the exception of automobiles, are taken from Fraumeni (1997). The rate of depreciation for autos was derived implicitly from the standard *Fixed Assets* tables (BEA, 2010, tab. 2.2, 2.8).

On the basis of these investment series, depreciation estimates and equation (2) we compile the real net stock of capital between 1929 and 1947 for all assets and industries distinguished by the BEA (with the exception of the government sector).

CAPITAL SERVICES

Capital services (K) for industry j can be derived by weighting the growth of capital stocks for all m assets by its relative share in total industries capital compensations (v). Dropping the industry subscript j for ease of notation, the growth of capital services can be represented as.

$$\text{(3)} \qquad \ln\left(\frac{K_t}{K_{t-1}}\right) = \sum_{k=1}^{m} \bar{v}_k^K \ln\left(\frac{N_{k,t}}{N_{k,t-1}}\right)$$

Where $\bar{v}_k^K$ is the average share of capital compensation in year t and $t-1$ for asset k.

$$\text{(4)} \qquad \bar{v}_k^K = 0.5\left(v_{k,t}^K + v_{k,t-1}^K\right)$$

As noted previously, capital compensation is the product of the rental price ($p_{k,t}^K$) and the real stock ($N_{k,t}$) of this asset. The share ($v_{k,t}^K$) is then calculated by dividing the assets capital compensation by the total industry's capital compensation. Note that industry j's capital compensation can be obtained from the national accounts as value added minus the compensation of labor (see table 1).

$$\text{(5)} \qquad v_{k,t}^K = \frac{p_{k,t}^K N_{k,t}}{\sum_{k=1}^{m} p_{k,t}^K N_{k,t}}$$

The calculation of the rental price reflects the fact that in equilibrium, an investor is indifferent between two alternatives: either buying a unit of capital at time $t-1$, collecting a rental fee and then selling the depreciated asset in the next period, or earning a nominal rate of return on a different investment opportunity. The capital services thus depend on the asset-specific depreciation rates (δ_k), the (industry-specific) rate of return (r_t) and the capital gains or losses from price changes in an alternative investment ($p_{k,t}^I$).[4]

$$\text{(6)} \quad p_{k,t}^K = p_{k,t-1}^I r_t + p_{k,t}^I \delta_k - 0.5\left[\ln\left(\frac{p_{k,t-1}^I}{p_{k,t-2}^I}\right) + \ln\left(\frac{p_{k,t}^I}{p_{k,t-1}^I}\right)\right] p_{k,t-1}^I$$

For the calculation of the industry rate of return (r_t) we follow the ex-post procedure preferred by the BLS to make our capital service estimates comparable to the post-war figures. The rate of return is the sum of total capital compensation and the total capital gains from changes in investment prices, minus total depreciation, divided by the capital stock in prices of year $t-1$.

$$\text{(7)} \quad r_t = \sum_{k=1}^{m}\Bigg(p_{k,t}^K N_{k,t} - p_{k,t}^I \delta_k N_{k,t} + 0.5\left[\ln\left(\frac{p_{k,t-1}^I}{p_{k,t-2}^I}\right) + \ln\left(\frac{p_{k,t}^I}{p_{k,t-1}^I}\right)\right] \cdot p_{k,t-1}^I N_{k,t}\Bigg) \Bigg/ \sum_{k=1}^{m} p_{k,t}^K N_{k,t}$$

[4] In equations (6) and (7) we rely on a two-period average change in the asset-specific investment prices to smooth out incidental price shocks.

For the estimation of the rental prices we again rely on depreciation rates by Fraumeni (1997), the BEA (2010) price index of investment and the industry-level capital compensation from the NIPA tables (BEA, 2009).

Table 2 shows the difference between Kendrick's original capital input measures, the average annual growth of the capital stock measured using the BEA's investment series and the growth in capital services. Kendrick's estimates are very similar to the growth figures for the capital stocks, but differ substantially from the estimates based on capital services. As previously noted, capital service flows will allocate greater weight to assets that have shorter asset lifetimes (i.e. equipment and machinery), the stock of which expanded more rapidly than for long-lived assets (i.e. structures) during the 1930s. This explains why the growth figures for capital services exceed the capital stock based measures for both the PDE and PNE as well as most of the underlying industries during the years 1929-1941.

TABLE 2—GROWTH OF CAPITAL INPUT, U.S. 1929-1941 (% P.A.).

Private domestic economy (PDE)	
Kendrick	-0.08
Capital stocks	-0.09
Capital services	0.37
Private dom. non-farm economy (PNE)	
Kendrick	-0.13
Capital stocks	-0.05
Capital services	0.48

Source: Kendrick (1961, 333-40); Bakker, Crafts and Woltjer (2015).

VARIABLE RETIREMENT

Gordon (2016, 659-63) argues that the official investment and depreciation rates from the BEA severely underestimate the growth in capital input for the period between 1925 and 1945. In particular, he questions whether the depreciation rates, which are fixed over time, are representative for the Depression era. As investment collapsed, Gordon would expect equipment and structures to be scrapped and depreciated at a slower rate; i.e. he proposes that the expected life-time of all assets should increase substantially during the 1930s. This lower rate of depreciation would lead to a greater increase in the capital stock than the estimates by the BEA would suggest.

As a crude proxy for these varying rates of depreciation, Gordon suggests comparing the ratio of investment to the official capital stock for each year with the average for 1925-1972. A low ratio, as was the case for 1933, would indicate an increase in the asset lifetime, whereas a relatively high ratio would indicate a reduction in the time producers hold on to their aging capital assets.

As a robustness check to our capital input estimates, we apply the same procedure. We estimated the ratio of investment to stock separately for equipment capital and structures. The data was taken from the BEA (2010) *Fixed Assets*, basic tables 2.1, 2.2, 2.7 and 2.8, lines 3 and 37. The official depreciation rates discussed in the previous section were multiplied by the ratio of investment to capital in the respective year, relative to the average of the period 1925-1972. On the basis of these depreciation rates we re-estimated the capital stocks and capital services.

The effect of the adjustment on industry TFP-growth varied substantially, from 0.0 percentage-points for coal mining, electric machinery and furniture, to minus 0.7 percentage-points for oil & gas mining and minus 0.5 percentage-points for both metal mining, and petroleum & coal products. The overall effect on the TFP-growth rate of the PDE was substantial, minus 0.2 percentage-points, which constitutes a ten percent downward adjustment.[5] It should be noted, however, that no evidence is available to validate Gordon's assumptions about delayed retirement of capital. We thus continue to rely on the BEA's fixed depreciation rates.

4. Labor quality

This section considers previous estimates of early-twentieth century labor quality change and discusses the estimation procedures and data behind a new set of industry-specific labor quality estimates for the U.S. between 1899 and 1947.

Kendrick estimates

Kendrick (1961, 31-3) estimated the effect of skill changes on the composition of the labor force between 1869 and 1957. Instead of measuring changes in education attainment, gender and experience directly, however, he measured the changes in the occupational structure. He adjusted labor input by weighting the man-hours of work in separate occupations and industries by their average hourly earnings for a given base year. Kendrick's measure of labor quality thus captures two effects: (1) the relative shifts of workers between occupations, and (2) the relocation of employment between industries. The first effect, the shift of workers from low-paying positions (e.g. laborers) to better-paying jobs (e.g. operatives or clerical staff), reflects a change in the potential output per worker. The higher earnings (measured in terms of base-period compensation) imply a rise in the marginal productivity of that worker and thus a rise in the quality of the labor force – in line with the Jorgenson approach discussed below. Likewise, the shift of workers to better-paying industries also show up as an increase in labor quality.

Kendrick assumes under (1) that labor quality will only change over time if a worker transfers from one occupation to another or if an individual joins (or leaves) the labor force in an occupation that is better (worse) paid than the national average. Kendrick (1961, 33) surmises that "the inherent average physical and mental capacity of the person employed in each occupation is constant over time." The rapid increase in educational attainment during the late nineteenth and early twentieth century casts serious doubt on this assumption, however. The average years of schooling for cohorts born between 1880 and 1950 nearly doubled, increasing from approximately 8 to 14 years (Goldin and Katz, 2008, 20, 113, 170). Part of this increase in skill translated into a shift of employees between occupations and industries, but part also translated into a rise of the labor quality within occupations. For instance, the likelihood for a blue collar worker born around 1885 to have attended high school was substantially greater than it was for its counterpart born only 10 years prior, around 1875. The high-school education gave the blue-collar worker basic knowledge of chemistry, electricity and algebra, allowed him to read manuals and blueprints and made it much easier for him to effectively converse with managers and other professionals, raising his marginal productivity in the process. In addition to undervaluing the impact of the rapid increases in education attainment during the late nineteenth and early twentieth century, Kendrick's method ignores other demographic changes as well which thus biases his labor quality figures downwards compared to our own (see table 3). Changes in the educational attainment, average age, or experience of the workforce and shifts in the gender composition are generally considered to be determining factors in the quality of labor, as we will illustrate below.

New estimates

In order to fully assess the impact of the substantial investments in schooling as well as the structural changes in the gender and age composition of the American workforce

[5] The tabulated results are available upon request from the authors.

TABLE 3—GROWTH OF LABOR QUALITY, U.S. 1899-1941 (% P.A.).

	Kendrick	This study
Private domestic economy (PDE)		
1899-1929	0.36	0.87
1929-1941	0.20	0.59
Private dom. non-farm economy (PNE)		
1899-1929	0.16	0.40
1929-1941	0.14	0.27

Source: Kendrick (1961, 333-40); Bakker, Crafts and Woltjer (2015).

during the early twentieth century, we turn to an approach developed by Jorgenson and Griliches (1967). The key innovation in their work was to adjust the traditional measure of labor input – i.e. total hours of work or employment – for improvements in quality. The main principle behind the labor quality ('LQ') adjustment is the distinction among several different types of labor inputs characterized by one or more quantifiable factors that affect the productivity potential of the worker (e.g. educational attainment, age, gender). By then assigning weights to these categories ('LAB') usually in the form of average wages and earnings one can measure the change in the productivity 'potential' of the workforce. The rationale for this procedure is that differences in average earnings between the labor categories can be thought of as reflecting differences in their marginal productivity. When this new measure of labor input ('LAB_Qi') is used in a growth accounting framework, output growth as a result of better educated and trained workers is ascribed to input growth, rather than productivity or technology growth (Jorgenson, Ho and Stiroh, 2008). Previous studies have shown that this quality adjusted measure can account for a substantial part of the residual or Total Factor Productivity (TFP) growth within traditional growth accounting studies (Denison, 1962; Griliches, 1963; Denison and Poullier, 1967; Gordon, 2010). Therefore, the labor quality adjustment allows for a purer measure of both labor input as well as technical change within a growth accounting framework.

METHODOLOGY

We assume that labor input ($HK_{j,t}$) for industry j at time t can be expressed as a translog function of its individual components (Jorgenson, Gollop and Fraumeni, 1999, 92-3). We form indices of labor input from data on employment by industry, cross-classified by gender, age and education.[6] Dropping the industry subscript j for ease of notation, the growth of labor input for industry j can thus be represented as.

$$(8) \qquad \ln\left(\frac{HK_t}{HK_{t-1}}\right) = \sum_{l=1}^{q} \bar{v}_l^L \ln\left(\frac{L_{l,t}}{L_{l,t-1}}\right)$$

Where L_l is employment at the industry level for a given set of q characteristics of the labor force l (gender, age and education) and $\bar{v}_l^L$ is the average of this employment group's share in the total labor income.

$$(9) \qquad \bar{v}_l^L = 0.5\left(v_{l,t}^L + v_{l,t-1}^L\right)$$

The share of labor income ($v_{l,t}^L$) at time t is derived as the product of the average wage (p_l^L) and employment ($L_{l,t}$) for each combination of labor characteristic l, divided by the total wage sum.

$$(10) \qquad v_{l,t}^L = \frac{p_l^L L_{l,t}}{\sum_{l=1}^{q} p_l^L L_{l,t}}$$

Alternatively, the index of labor input can also be expressed as the product of employment (L) and an index of labor quality (LQ)

[6] Age, in our estimate for labor input, serves as a proxy for (work) experience. We thus assume that an individual has held a job his entire life since he turned 16, left high-school or graduated from college, depending on his educational attainment.

or, in growth terms, as.

$$\ln\left(\frac{HK_t}{HK_{t-1}}\right) = \ln\left(\frac{L_t}{L_{t-1}}\right) + \ln\left(\frac{LQ_t}{LQ_{t-1}}\right) \tag{11}$$

Rearranging terms in equation (11) and substituting the index for labor input by (8) we arrive at a direct measure of sectoral labor quality growth.

$$\ln\left(\frac{LQ_t}{LQ_{t-1}}\right) = \sum_{l=1}^{q} \bar{v}_l^L \ln\left(\frac{L_{l,t}}{L_{l,t-1}}\right) - \ln\left(\frac{L_t}{L_{t-1}}\right) \tag{12}$$

The change in labor quality thus reflects the difference between the growth rates of the compensation-weighted index of labor input and sectoral employment.

The drawback of this approach is that it requires highly disaggregate data on employment and compensation, generally not available in the published census reports or secondary sources for the early twentieth century. Fortunately, the Integrated Public Use Microdata Series (IPUMS) has made samples from the decennial population censuses publicly available, providing detailed records for nearly 10 million individuals between 1900 and 1950 (Ruggles et al., 2015). We utilize the microdata from this source to construct our measure of labor quality.

Unfortunately, however, the 1900-1930 American population censuses did not inquire into either the educational attainment of the general population or the compensation of workers and employees. To overcome these data issues, we follow a three-tiered approach to the data preparation for the labor quality estimation. First, we estimate educational attainment at the micro level for the pre-1940 census samples based on the 1940 returns. Second, we construct an employment matrix for the entire period that groups workers according to their (predicted) educational attainment, gender, age and by industry. Lastly, we derive the compensation matrix based on average wages for each labor category taken from the 1940 census of population.[7] These employment and compensation matrices can then be used to calculate labor quality on the basis of equation (12).

EDUCATIONAL ATTAINMENT

For the first stage, we estimate the educational attainment y for an individual i on the basis of his or her occupation, gender, age and place of residence x_i. On the basis of this approach we take both the long-run changes in the average years of schooling as well as the effects of changes in the occupational structure and the gender/age composition of the workforce into account. We define four education categories (see table 4) and we predict the likelihood that an individual i belongs to each of these specific educational categories (e.g. $\Pr\{y_i = 1\}$). This probability should be bounded by 0 and 1, continuous and nonlinear; conditions which are all met by an (ordered) logit model.

$$\Pr\{y_i \le k | x_i\} = \frac{e^{x_i{}'\beta}}{1 + e^{x_i{}'\beta}}\ , k = 1, 2, 3 \tag{13}$$

The right-hand side of equation (13) is a cumulative distribution function with mean 0 and standard deviation 1. The coefficients are estimated using maximum likelihood, which is the optimal parametric estimator in this context (Long and Freese, 2006).[8]

[7] The 1940 census was the first census of its kind to ask about schooling, labor compensation and working hours to all citizens surveyed. In the wake of the depression the 1940 population census dedicated a substantial part of its inquiry into the issue of employment and productivity.

[8] We estimate the cumulative probability for the first three educational categories, since all individuals that are not part of either the first, second or third category will be part of the fourth category. The fourth category can thus be implicitly derived and should be excluded from the model.

TABLE 4—CATEGORICAL VARIABLES LOGIT MODEL

Education:
see labor quality model

Gender:
see labor quality model

Occupation:
(1) professional, technical
(2) farmers (owners and managers)
(3) managers, officials, and proprietors
(4) clerical staff
(5) sales workers
(6) craftsmen
(7) operatives
(8) service workers (household)
(9) service workers (other)
(10) laborers
(11) unemployed or retired

Region:
(1) South
(2) Midwest
(3) West
(4) Northeast

TABLE 5—CATEGORICAL VARIABLES LABOR QUALITY MODEL

Education:
(1) 1-4 years grade school
(2) 5-8 years grade school
(3) 1-4 years high school
(4) 1 or more years college

Gender:
(1) male
(2) female

Age:
(1) 16-17 years
(2) 18-24 years
(3) 25-34 years
(4) 35-44 years
(5) 45 years and over

Industry:
see main text

DATA

For the estimation of the logit model we rely exclusively on the 1940 1-percent sample included in the IPUMS dataset. This sample is limited to include only those citizens aged 16 years and above, leaving approximately 975,000 observations for the logistic regression. The dataset includes a measure of the highest year of schooling or degree completed. As illustrated in table 4, we reclassify this variable to encompass four distinct educational attainment classes. The reason we reclassify the education variable is twofold. First, by treating it as a categorical variable as opposed to a continuous variable (e.g. years of education), we avoid the assumption that the distances between classes are equal; i.e. that an additional year of grade school is identical to one additional year in college. Second, we limit the number of classes to 4 to ensure that each class is covered by a sufficient number of observations. This is important not just for the estimation of educational attainment, but also for the construction of the compensation matrix.[9] For the independent variables, we follow the literature on US labor quality and mark four variables as important predictors of educational attainment, namely: occupation, birth cohort, gender and region.

Individuals are classified into one of eleven main occupational groups which differ markedly in terms of their average educational attainment. For instance, the probability of a professional (e.g. engineers, eco-

[9]Limiting the number of classes for the education variable allows us, for instance, to test the 'parallel regression assumption'; meaning that for each education class (grade-school, high-school, college, etc.) the coefficients for the independent variables (beta) are identical. As it turns out the assumption is violated. Hence we effectively estimate separate regressions for all education classes, obtaining different betas for each.

nomists) having attended high school was substantially greater than was the case for the average laborer. The importance of gender and year of birth is illustrated by Goldin and Katz (2008, 18-22, 170). They observe a rapid increase in the average years of schooling throughout the late nineteenth and early twentieth century. Each successive cohort spent a substantially greater number of years in school compared to the previous generation. In addition, Goldin and Katz (2008, 19) show that women generally attended school for longer than men did throughout most of the early twentieth century. The gender variable was taken directly from the IPUMS dataset while the year of birth was rounded off to the nearest decade. The log of the relative distance in decades to 1930 was then taken as the birth cohort measure. Lastly, the literature points to widespread differences in state support for education and shows that the rise in both high school graduation rates as well as college enrollment rates for states in the North and West of the country were considerably more impressive than for the rest of the nation (Goldin and Katz, 2008, 271-7). We incorporate a variable in the model that differentiates between the four main regions of the country (see table 4).

For the second stage of the labor quality estimation, the construction of the employment matrix, we rely on the IPUMS 1-percent census samples for the decades between 1900 and 1950. To estimate educational attainment we include the occupation, birth cohort, gender and region variables discussed earlier, supplemented by data on the age of the individual and industry in which the subject is engaged. The employment sample is limited to include only those citizens between the ages of 16 and 84, who are part of the labor force. For the employment-matrix our dataset includes roughly 3,135,000 individual observations.

In the third stage of the data preparation we again rely on the 1940 sample to estimate relative compensation per labor category. Here we limit the sample to include only those citizens between the ages of 16 and 84 having worked at least 48 weeks in the previous year and earning an income greater than 0 (Goldin and Katz, 2008). These individuals are allocated to the cells of the matrix cross-classified by gender, age, education and industry as summarized in table 5. Compensation is reported in the census as the respondent's total pre-tax wage and salary income for the previous year, expressed in current dollars. To obtain total personal income, which also includes non-wage income, we multiplied the 1940 compensation figures by the industry specific ratio between wage and salary income and total personal income taken from the 1950 census returns. Non-wage income generally represented only a small part of total personal income, with the notable exception of the agricultural sector. The samples for the logistic regression, the compensation matrix and the employment matrix are all weighted by the IPUMS 'person weight' variable.

ROBUSTNESS

Ideally, we would like to allow the weights for our labor quality index to vary over time, reflecting potential changes in relative compensation between the labor categories. Unfortunately, the censuses prior to 1940 did not inquire into either wages or earnings, impeding the accurate measurement of labor compensation for these earlier decades. Reassuringly, Goldin and Katz (2008, 53-63) demonstrate that the wage structure observed in 1940 was fairly typical for the prewar period. Although they do observe a gradual compression of the wage distribution for production workers between 1890 and 1940, Goldin and Katz conclude that the gap in the skilled/unskilled wage level for 1920 was virtually identical in comparison to 1940.

On the basis of data by Goldin and Katz (2010) for the state of Iowa we performed a more conclusive sensitivity check of our labor quality figures. The Iowa State Census Project provided detailed compensation data, cross-classified by most of the categories that make-up labor input for the year 1915. In addition, we performed the same sensitivity check based on comprehensive income data for 1950, taken from the IPUMS dataset. Overall, on the basis of this analysis we feel confident the use of only 1940 compensation figures does not bias our labor quality estimates.[10]

[10]Full details are available on request from the authors.

REFERENCES

Bakker, Gerben, Nicholas Crafts, and Pieter Woltjer. 2015. "A Vision of the Growth Process in a Technologically Progressive Economy: the United States, 1899-1941." *CEPR Working Paper Series*, 10995.

BEA. 1966. *The National Income and Product Accounts of the United States, 1929-1965: Statistical Tables.* Washington D.C.:U.S. Government Printing Office.

BEA. 1993. *Fixed Reproducible Tangible Wealth in the United States, 1925-89.* Washington D.C.:U.S. Government Printing Office.

BEA. 2002. *Fixed Assets.* Washington D.C.:U.S. Government Printing Office. accessed April 4, 2014.

BEA. 2003. *Fixed Assets and Consumer Durable Goods in the United States, 1925-99.* Washington D.C.:U.S. Government Printing Office.

BEA. 2009. *National Income and Product Accounts.* Washington D.C.:U.S. Government Printing Office. accessed January 11, 2010.

BEA. 2010. *Fixed Assets.* Washington D.C.:U.S. Government Printing Office. accessed December 8, 2011.

BEA. 2011. "Value Added by Industry, Gross Output by Industry, Intermediate Inputs by Industry, the Components of Value Added by Industry, and Employment by Industry, 1947-1997." *Historical Industry Accounts Data.* accessed April 4, 2014.

BLS. 1943. *Wholesale Prices, July-December and Year 1942.* Washington D.C.:U.S. Government Printing Office.

BLS. 1949. *Wholesale Prices, 1947.* Washington D.C.:U.S. Government Printing Office.

BLS. 1958. *Wholesale Prices and Price Indexes 1957.* Washington D.C.:U.S. Government Printing Office.

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

Carter, Susan, Scott Gartner, Michael Haines, Alan Olmstead, Richard Sutch, and Gavin Wright, ed. 2006. *Historical Statistics of the United States: Earliest Times to the Present.* Cambridge:Cambridge University Press.

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

Denison, Edward, and Jean-Pierre Poullier. 1967. *Why Growth Rates Differ: Postwar Experience in Nine Western Countries.* Washington D.C.:Brookings Institution Press.

Field, Alexander. 2003. "The Most Technologically Progressive Decade of the Century." *American Economic Review*, 93(4): 1399–1413.

Field, Alexander. 2006. "Technological Change and U.S. Productivity Growth in the Interwar Years." *Journal of Economic History*, 66(1): 203–236.

Fraumeni, Barbara. 1997. "The Measurement of Depreciation in the U.S. National Income and Product Accounts." *Survey of Current Business*, 77(7): 1–49.

Goldin, Claudia, and Lawrence Katz. 2008. *The Race between Education and Technology.* Cambridge:Harvard University Press.

Goldin, Claudia, and Lawrence Katz. 2010. *The 1915 Iowa State Census Project.* Ann Arbor:Inter-university Consortium for Political and Social Research. accessed on 14 December, 2010.

Gordon, Robert. 2010. "Revisiting U.S. Productivity Growth over the Past Century with a View of the Future." *NBER Working Paper*, 15834: 1–30.

Gordon, Robert. 2016. *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War.* Princeton:Princeton University Press.

Griliches, Zvi. 1963. "The Sources of Measured Productivity Growth: United States Agriculture, 194060." *Journal of Political Economy*, 71: 331–46.

Jorgenson, Dale, and Zvi Griliches. 1967. "The Explanation of Productivity Change." *Review of Economic Studies*, 34: 249–83.

Jorgenson, Dale, Frank Gollop, and Barbara Fraumeni. 1999. *Productivity and U.S. Economic growth.* . 2nd ed., New York:toExcel.

Jorgenson, Dale, Mun Ho, and Kevin Stiroh. 2008. "A Retrospective Look at the U.S. Productivity Growth Resurgence." *Journal of Economic Perspectives*, 22: 3–24.

Kendrick, John. 1961. *Productivity Trends in the United States.* Princeton:Princeton University Press.

Long, Scott, and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata.* . 2nd ed., College Station:StataCorp LP.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. *Integrated Public Use Microdata Series.* . 6th ed., Minnesota:University of Minnesota.

Appendix

Table A1—List of variables

	Values		
VA	Value added at current prices	($)	p. 1
PEP	Persons engaged in production	(thousands)	p. 2
H_avg	Average weekly hours of work	(number)	p. 2
	Prices		
VA_Pi	Gross output price deflator	(1929=100)	p. 2
	Volumes		
VA_Qi	Value added at constant prices	(1929=100)	p. 2
H_Qi	Total hours of work	(1929=100)	...
LQ	Labor quality	(1929=100)	p. 7
LPE	Output per person engaged	(1929=100)	...
LPH	Output per hour worked	(1929=100)	...
	Growth accounting: 'capital stocks'		
Ks	Share of capital	(share)	...
K_Qi	Capital stocks (total) in constant prices	(1929=100)	p. 3
KP	Output per unit of capital	(1929=100)	...
TFP_k	Total Factor Productivity, capital stocks	(1929=100)	...
	Growth accounting: 'capital services'		
LAB	Labor compensation	($)	p. 7
CAP	Capital compensation	($)	p. 3
LAB_Qi	Labor services	(1929=100)	p. 7
CAP_Qi	Capital services	(1929=100)	p. 3
LABP	Output per unit of labor service	(1929=100)	...
CAPP	Output per unit of capital service	(1929=100)	...
TFP_cap	Total Factor Productivity, capital services	(1929=100)	...