Online Appendix: Data Construction of New BEA-BLS Estimates of Industry-level Sources of U.S. Economic Growth between 1987 and 2016

Corby Garner
Bureau of Labor Statistics
Justin Harper
Bureau of Economic Analysis
Thomas F. Howells III
Bureau of Economic Analysis
Matt Russell
Bureau of Labor Statistics
Jon Samuels
Bureau of Economic Analysis[1]

This appendix provides additional details on the data construction.

Gross Output and Intermediate Inputs

BEA's GDP by Industry statistics provide a time series of nominal and real gross output, intermediate inputs (including a decomposition of energy, materials, and purchased services inputs), and value added by industry, prepared based on the 2007 North American Industry Classification System (NAICS). These data are fully integrated with expenditure-based GDP estimates from the National Income and Product Accounts (NIPAs). In addition, the data are prepared within a balanced supply-use framework that allows for simultaneous and consistent analysis of industry output, inputs, value added, and final demand. These fully integrated GDP by industry accounts were first released in January 2014,

1 The main article is available at http://www.csls.ca/ipm/36/Garner.pdf.

and covered the period 1997-2012. They have subsequently been updated annually and extended to cover the period 1947-2016 (Kim, Strassner and Wasshausen, 2017 and Barefoot, Gilmore and Nelson, 2017).

While certain series in the GDP by Industry dataset extended further back in time, the full suite of integrated make-use tables and GDP by industry statistics extended back only to 1998 prior to the release in early 2014. As a result, the initial version of the industry-level production account was also limited to the time period beginning in 1998. However, in February 2016, BEA released integrated make-use tables and GDP by Industry statistics extending back to 1947, adding a half century of historical data to this time series. The availability of these new historical data opened the possibility of extending the industry-level production account back in time as well (Lyndaker *et al.* 2016).

Gross Output and Intermediate Inputs: Backcasting

Preparation of the historical make-use tables and GDP by Industry statistics relied heavily on a series of benchmark input-output tables prepared by BEA from 1947 to 1992.[2] These tables provide valuable information about the structure of the U.S. economy at various points in history; however, the tables as initially published were designed as standalone snapshots of the economy and could not be treated as a time series. Among other things, earlier tables were not updated to incorporate definitional and conceptual changes introduced in later tables, and the tables were not prepared using a consistent industry classification structure. Transformation of these disconnected benchmark tables into a consistent annual time series took place in four major steps.

First, the tables were updated to incorporate major definitional changes introduced into the NIPAs since the initial publication of each table. Major changes incorporated as part of this step include statistical revisions to autos and trucks, housing and housing services, and non-profits; changes to the treatment of output for insurance and banking; introduction of the "government as producer" treatment; and capitalization of software.

Second, the tables were updated to reflect the 2002 NAICS structure, consistent with data available as of May 2010. Benchmark tables

2 Benchmark tables in this period were prepared for the following years: 1947, 1958, 1963, 1967, 1972, 1977, 1982, 1987, and 1992.

in the historical period were initially published using whichever version of the Standard Industrial Classification (SIC) system was current at the time of preparation. The SIC-to-NAICS concordance that was used for the initial conversion of these tables to NAICS was based on fixed weights from 1997. This 1997 fixed-weight concordance was updated by aggregating SIC-based benchmark data up to roughly a 3-digit NAICS level of detail. The resulting concordance was used to convert both make and use tables to a 2002 NAICS basis. The converted tables were then re-balanced using a RAS balancing technique.

Third, annual tables were prepared for the inter-benchmark periods. To begin, make tables and unbalanced use tables were prepared by interpolating between benchmark tables. A variety of data were used as indicators for the interpolation process, including Census survey data, annual make-use tables and gross output by industry statistics published by BEA, and personal consumption and private investment data from the National Income and Product Accounts(NIPA). Each of these tables was then controlled to be consistent with historical GDP data using RAS balancing.

Fourth, the time series was updated to reflect the 2007 NAICS structure, and definitional and statistical improvements from BEA's 2013 comprehensive update were incorporated into the dataset. Major changes incorporated at this stage include the capitalization of R&D expenditures, own account entertainment originals, and residential housing transfer costs as well as the adjustment of defined pension plans from a cash-accounting basis to an accrual-accounting basis.

For the industry-level production account, intermediate inputs in the historical time series also needed to be decomposed into energy, materials, and purchased services (EMS) components. The time series of make-use tables for 1963-1996 includes a decomposition of intermediate inputs into 75 commodities; however, this level of detail was not sufficient to allow a one-to-one assignment of each commodity to an EMS category. We addressed this using unpublished data from the 1997 use table. The working level of detail for tables beginning in 1997 includes about 5,000 goods and services and about 800 industries. At this level of detail, it is possible to make a direct EMS assignment to each cell in the use matrix. By aggregating these data up to the same level of detail as the historical use tables (about 75 goods and services and 65 industries), we were able to calculate EMS ratios for each cell in the matrix. These fixed ratios were applied to the historical use tables to generate estimates of EMS inputs for each

industry for 1987-1996.

With the nominal data available in the make-use framework, estimates of real GDP by industry were prepared using a double deflation methodology in which gross output and intermediate inputs are deflated separately and real value added is computed as the residual, as is standard in the published GDP by industry statistics. Real gross output by industry was derived by deflating the commodities produced by each industry as reflected in the make table.[3] Similarly, real intermediate inputs were derived by deflating the commodities consumed by each industry as reflected in the use table. Prices used for deflation were developed by Dale W. Jorgenson, Mun Ho, and Jon D. Samuels and are described in more detail in the article that accompanied the initial publication of the historical GDP by industry statistics (Lyndaker, *et al.* 2016).

As in the post-1996 period, the domestic and imported portions of intermediate inputs are deflated separately to account for potential differences in price between commodities purchased from domestic versus foreign sources. Intermediate inputs are disaggregated into domestic and imported components based on the proportionality assumption. More specifically, for each detailed commodity used by an industry, the portion attributable to imports is calculated as the ratio of total imports of the commodity over total domestic supply of the commodity.

Gross Output and Intermediate Inputs: Revisions

This 2017 annual update to BEA's GDP by industry statistics incorporated the Census Bureau's latest Service Annual Survey (SAS) tabulations, which revised statistics for 2014 and 2015. In addition, newly available data for 2016 from SAS replaced estimates based on the Census Bureau's Quarterly Services Survey (QSS). Similarly, revised and newly available data from the Department of Treasury's Statistics of Income (SOI) Division led to revisions to underlying components of the current-dollar estimates of value added by industry, including corporate profits, nonfarm proprietors' income, and net interest for 2014 and 2015. Finally, the annual update incorporated newly available Census Bureau data from the 2015 Annual Survey of Manufactures, the 2015 Annual Retail Trade Survey, and the 2015 Annual Wholesale Trade Survey.

3 This differs from the BLS Productivity Program's use of a sectoral output measure which removes the double counting of the intrasectoral sales between establishments within the same sector.

Capital Services Inputs

Capital Input: Backcasting

By definition, a capital asset is one that lasts more than a year. Therefore, the service flows received from a producing industry or firm are employed over a longer period than the original investment conveys. As a practical matter, real stocks are constructed as vintage aggregates of real historical investments in accordance with an "efficiency" or service flow concept using the perpetual inventory method outlined in the initial release of the BEA-BLS production account. This implies that the hard work of backcasting historical stocks had, in effect, already been accomplished by the BLS with the original release of the BEA-BLS account.

The current-dollar value added components by industry needed to calculate rental prices used to construct capital services were the missing link that BEA's release of historical make-use tables was able to bridge. These new historical measures allowed the BLS to incorporate capital services estimates back to 1987. With the release of the 2016 Multifactor Productivity Trends[4] news release, BLS incorporated these integrated and improved data into their official measures. Thus, this release makes use of all the available historical data from the most recent BLS multifactor productivity update to produce its capital service measure back to 1987 to be consistent with the national accounts.[5]

Capital Input: Revisions

With this update of the account, the capital measures for the finance and insurance industries were revised to more fully integrate the BEA-BLS production account into the national accounts. Previously, in the finance and insurance sectors, controls for capital compensation were not constrained to BEA income estimates because of concerns for some of the features of the more detailed estimates.

More recently, BEA has made a number of improvements to the finance and insurance sectors that made constraining the BLS construction of capital measures to the income pieces available from BEA beneficial from both a consistency standpoint as well as a methodological standpoint. Specifically, BEA improved the insurance and banking es-

4 See https://www.bls.gov/news.release/archives/prod3_03302017.pdf.

5 See https://www.bls.gov/mfp/mprdload.htm.

6 See March 2013 Survey of Current Business "Preview of the 2013 Comprehensive Revision of the National Income and Product Accounts."

timates with the 2013 Comprehensive Revision (Survey of Current Business, 2013).[6] These improvements have made constraining to BEA, as is done with all other industries in the BEA-BLS production accounts, a logical next step.[7] This change follows publication of the BLS Multifactor Productivity Trends 2017[8] release in which this consistent treatment was also adopted by BLS in the finance and insurance sectors. On a real basis, the revisions to the capital input measures were small.

The final revision to the methods for capital service estimates involves an update to how inventories are distributed among most non-manufacturing industries. For all non-manufacturing industries except mining, utilities, and construction, BEA inventories are grouped into an "other" category. BLS uses IRS book value data to distribute "other" inventories to the remaining non-manufacturing industries after moving the data from a company to an establishment basis via establishment-company ratios. Additionally, BLS employed a three-year smoothing of the IRS book values for the data processing, internet publishing industry. Previously BLS had only smoothed the data for the years 1999-2008. BLS now smooths the data from 1999 through the last year available. Because the IRS data are used to determine the share of "other" inventories allocated to each non-manufacturing industry, this change affected the inventory values of all of the other non-manufacturing industries as well.

Labour Services Inputs

Labour Hours

As in the previous set of accounts, BLS prepares a time series of labour hours reflecting annual hours worked aggregated from estimates of more detailed industries. BLS estimates labour hours using payroll employment and hours from the Current Employment Statistics (CES) survey and are supplemented with data for the self-employed and average weekly hours for nonproduction and supervisory workers from the Current Population Survey (CPS).[9] The BLS National Compensation Survey is also used to convert the hours of payrolled workers from a paid to a worked basis when constructing the initial set of employment and hours estimates

7 See https://www.bea.gov/scb/pdf/2013/06%20June/0613_preview_comprehensive_iea_revision.pdf also https://www.bea.gov/scb/pdf/2013/02%20February/0213_nipa-rev.pdf.

8 See https://www.bls.gov/news.release/archives/prod3_03212018.pdf

9 See http://www.nber.org/chapters/c13005.

for the BEA-BLS production account. [10]Sources for industries that are not covered by CES or where data are missing include the Department of Agriculture, BLS Quarterly Census of Employment and Wages (QCEW), and Mine Safety and Health Administration.[11] These estimates are subsequently controlled to BEA National Accounts estimates of hours worked at a more aggregate level before being distributed to demographic groups as outlined in the labour composition section.

Labour Hours: Backcasting

In 2003, the BLS CES program released historical employment and average weekly hours data for detailed industries on a North American Industrial Classification system back to 1990. In addition to the historical data, the BLS CES program made available the bridge ratios used in converting the measures. These data serve as the basis for the estimates of the BEA-BLS Production account prior to 1997 and are consistent with the data after 1997. For the period prior to 1990, the BLS CES provides reconstructed historical measures for two-digit NAICS sectors back to 1939. Reconstructed historical measures for years before 1990 are also available for selected three, four, five, and six-digit NAICS industries.

Where available, complete NAICS-based payrolled employment and hours from the BLS CES series were used in the BEA-BLS production account. For the industries that were not available from BLS CES, employment estimates were developed using historical Standard Industrial Classification (SIC)-based measures and conversion factors published by the BLS CES program to reconstruct NAICS-based series back to 1987. This conversion was carried out at the most detailed industry level for which conversion factors were available by the BLS, that is, using a four-digit SIC to six-digit NAICS-CES bridge. The resulting converted employment estimates were historically linked to the published CES employment data series in 1990. Hours for 1987-89 were derived using an approach similar to that used to develop the employment estimates. Total hours residuals were constructed from the available NAICS-based aggregates and the available NAICS-based detail from BLS CES program. The residual total hours were then distributed to the detail using proportions derived from the initially converted set of NAICS-based employ-

10 See https://www.bea.gov/scb/pdf/2014/08%20August/0814_industry-level_production_account.pdf.

11 See https://www.bea.gov/scb/pdf/2014/08%20August/0814_industry-level_production_account.pdf.

ment estimates, to create total payrolled hours estimates for the remaining component industries at all levels of industry detail.

The CPS assigns respondents to industries using the Census Bureau's Industry Classification System (ICS). There are two different classification systems across the time period of 1987-1996. For the years 1987-1991, the ICS is based on the 1980 Census codes, which uses the 1972 SIC classification, as modified in 1977. The ICS for the 1992-2002 interval is based on the 1990 Census which uses the 1987 SIC classification. The CPS data from 2000-2002 were dual-coded on both an SIC and a NAICS basis.

To estimate consistent historical NAICS-based self-employed and supervisory average weekly hour estimates back to 1987, a multi-step process was followed to convert the historical CPS data from an SIC-based code to a NAICS-based industry code used in the BEA-BLS production account. First, a three-year average SIC-to-NAICS conversion bridge was derived from the dual coded CPS data from 2000-2002. Adjustments to the initial conversion ratios were later made based on a comparison of the NAICS industry employment levels for 2000-2002 generated by applying the bridge with the employment estimates from the CPS data provided on a NAICS basis. NAICS final employment and hours estimates for 1987-2002 were derived by applying the adjusted conversion ratios to the historical SIC-based employment and hours series.

With the annual release of the 2016 data, the BLS instituted an improvement to the hours worked to hours paid ratios necessary to convert the BLS CES data for payrolled workers to the theoretically preferred hours worked basis.[12] From 2005-onward, fourth-quarter NCS data at the three-digit NAICS level are used as a proxy for each annual ratio value. During these years, more than 98 per cent of the sample rotation is isolated to the fourth quarter of each year. These new observations, which represent around 20 per cent of the overall respondents in this quarter, provide a refreshed source of response relative to the three prior quarters of the year, in which carried-over responses are generally held the same as the initial survey response.

From 1996 through 2004, however, an average of the four quarterly NCS ratios is used as the NCS sample rotation was intermittent throughout the year and was not regularly scheduled as the 2005-onward period had been. In order to estimate three-digit ratios

12 See https://www.bls.gov/lpc/hwhpnew.htm.

for 1990-1996, the 1996 NCS ratio values are carried backwards using the BLS Hours At Work Survey (HWS) as an extrapolator series. For 1987-1989, ratios for 14 super sectors from the HWS are utilized to move more detailed industry hour worked to hours paid data backwards.

Labour Hours: Revisions

This update of the BEA-BLS production accounts uses the same source for labour hours as the original release, but includes a number of improvements. The BLS CES made an improvement in educational services, health care and social assistance that has been incorporated into this update the BEA-BLS production accounts.[13] In addition, there has been a revision to incorporate the all-employee hours measure for couriers and messengers within other transportation and support activities which revises the series across the time period.

Labour Composition

For this set of accounts, workers are disaggregated by sex, eight age groups, six education groups, and employment class (payrolled vs. self-employed) for a total of 192 demographic categories. In addition, workers are categorized into one of 63 industries resulting in a total of 12,096 cells in the labour composition matrices for each period.

The estimation process begins by filling out information on employment, hours, and compensation for each cell in these matrices. For 1990 and 2000, the matrices are initialized using the U.S. Census 1990 and 2000 1-per cent Public Use Microdata Sample (PUMS) files. Initial estimates are generated for 1991-1999 by linear interpolation at the cell level. These initial estimates are iteratively adjusted using the RAS balancing technique to match a series of marginal controls developed from the March supplement to the CPS. For years before 1990 the t+1 balanced matrices are used as the initial cell estimates, and for years after 2000 the t-1 balanced matrices are used. As with the periods 1990-2000, these initial matrices are iteratively adjusted to match controls from the CPS.[14]

After balancing, the matrices are scaled in sequence (1) to employment controls from BEA's National Income and Product Accounts (NIPAs) for 63 industries by employment class, (2) to BLS hours for 63 industries by em-

13 https://www.bls.gov/ces/cesbmart13.pdf.

14 Labour composition estimates for the published BLS MFP data are constructed using the Basic Monthly CPS data. BLS and BEA are collaborating to reconcile the labour composition measures produced by BLS for the official MFP estimates and those produced by BEA for the account presented in this article.

ployment class, (3) to NIPA hours for payrolled workers by 17 aggregate industries, and (4) to NIPA compensation for payrolled workers by 63 industries. In the final step, the hourly compensation of self-employed workers is replaced by the rate for payrolled workers in the same cell. This step is taken because reported compensation of self-employed workers cannot be disentangled from compensation accruing to their capital assets. Additional methodological information is described in Fleck, *et al* (2014) with updates in Rosenthal, et al (2014).

Labour Composition: Backcasting

Previous publications of these accounts made use of an SIC-to-NAICS bridge from the BLS CES program to convert SIC-based labour measures beginning in 2003 to NAICS industries. In preparing the new historical period covered by these accounts, a modified SIC-to-NAICS bridge was constructed to incorporate time-varying weights for manufacturing industries. These dynamic, employment-based weights were supplied by the Federal Reserve Board based on research from Bayard and Klimek (2004) which made use of establishment-level microdata from the Census of Manufacturing and the Annual Survey of Manufactures spanning the period from 1963 to 1997. The time-varying weights replaced static weights where available, but were scaled to leave unchanged any weights linking portions of SIC manufacturing industries to NAICS non-manufacturing industries. For the period between 1997 and 2000, all updated manufacturing weights were interpolated to the static weights from the previous bridge.

The modified SIC-to-NAICS bridge was applied to the U.S. Census 1990 PUMS files to develop the initial 1990 labour composition matrix as well as to the 1987-2002 CPS marginal controls. The bridge was also applied to the SIC-based NIPA employment, hours, and compensation scaling controls for 1987-1997; however, these converted results were not used directly. In order to mitigate the possibility of time series breaks, the converted series were used as indicators to backcast a time series beginning with the 1998 levels in the published NAICS-based NIPA tables. Finally, these new NAICS-based employment, hours, and compensation levels were scaled to the SIC-based totals for all industries to ensure that this conversion process left totals unchanged.

In addition to the modified bridge, the 1987-1991 March Supplement of the CPS required special handling for the reported level of educational attainment. The current questionnaire allows respondents to select their

highest degree attained, which aligns well with the education categories chosen for these accounts. However, prior to 1992, respondents were instead asked for the number of years of schooling as well as whether the last year of schooling was completed. This inconsistency was addressed by converting the number of years of schooling to an estimated highest degree attained via a frequency matrix described in Jaeger (1997). That work matched CPS respondents who had reported educational attainment under both versions of the questionnaire, and cross tabulated pairs of responses to create conversion weights.

Labour Composition: Revisions

Revisions to the period 1998-2000 are a result of the interpolated Census PUMs matrices. The process of controlling to the CPS redistributed the marginal matrices based on the shares that resulted from the iterative proportional scaling procedure. Beyond that, revisions reflect updates to incorporate the latest data with hours and compensation estimates and are typically confined to the 2014 forward period.

References

Barefoot, Kevin, Teresa L. Gilmore and Chelsea K. Nelson (2017) "The 2017 Annual Update of the Industry Economic Accounts, Initial Statistics for the Second Quarter of 2017, Revised Statistics for 2014-2016 and the First Quarter of 2017," *Survey of Current Business*, December.

Bayard, Kimberly and Shawn Klimek (2004) "Creating a Historical Bridge for Manufacturing Between the Standard Industrial Classification System and the North American Industry Classification System," 2003 Proceedings of the American Statistical Association, Business and Economic Statistics Section [CD-ROM], pp. 478-84.

Jaeger, David A. (1997) "Reconciling the Old and New Census Bureau Education Questions: Recommendations for Researchers," *Journal of Business and Economic Statistics*, Vol. 15, No. 3, July.

Kim, Donald, Erich H. Strassner and David B. Wasshausen (2014) "Industry Economic Accounts, Results of the Comprehensive Revision, Revised Statistics for 1997-2012," *Survey of Current Business*, February.

Lyndaker, Amanda S., Thomas F. Howells III, Erich H. Strassner and David B. Wasshausen (2016) "BEA Briefing, Integrated Historical Input-Output and GDP by Industry Accounts, 1947-1996," *Survey of Current Business*, February.

Rosenthal, Steven, Matthew Russell, Jon D. Samuels, Erich H. Strassner and Lisa Usher (2014) "Integrated Industry-Level Production Account for the United States: Intellectual Property Product and the 2007 NAICS," paper presented at the Third World KLEMS Conference, Tokyo, Japan, 19-20 May.

Survey of Current Business (2013) "Preview of the 2013 Comprehensive Revision of the National Income and Product Accounts," pp. 6-14.