NBER-CES Manufacturing Industry Database: Technical Notes

Randy Becker, Wayne Gray, Jordan Marvakov
April 2016

1 Introduction

The NBER-CES Manufacturing Industry Database contains annual data from the United States manufacturing sector for the period from 1958 to 2011. The data used for the development of the database come from various sources, but chiefly from three government agencies: the U.S. Census Bureau, the Bureau of Economic Analysis (BEA), and the Bureau of Labor Statistics (BLS). The goal is to provide a long time-series of data for a large number of industries, adjusting for changes in industry definitions, and creating price deflators and capital stocks. This paper describes issues related to the most recent update of the database (1997 to 2011). For a more detailed discussion of an older version of the data see Bartelsman and Gray (1996).

The key feature (and complication) of recent updates is the adjustment to the NAICS industry definitions. The database is published in two versions: one based on the 1987 Standard Industrial Classification (SIC), containing 459 industries, and the other based on the 1997 North American Industrial Classification System (NAICS), containing 473 industries. The boundaries of the manufacturing sector shifted between SIC and NAICS, with some SIC industries leaving manufacturing and other NAICS industries entering manufacturing. To maintain consistency in the available data, the benchmark year in the NAICS version was switched to 1997, while the SIC version continues to use 1987 as the benchmark. This version of the database recalculates all the data post-1996; earlier updates of the database (1958-2002, 1958-2005, and 1958-2009) are superseded and should not be used.

The database is available for download at www.nber.org/data/nberces.html in several different formats.

2 Data and Sources

2.1 ASM/Census Data

The majority of the variables provided in the database were extracted from the Annual Survey of Manufactures (ASM). The ASM includes 50,000-65,000 establishments selected from the approximately 350,000 establishments included in the Census of Manufactures (CMF). The ASM is conducted annually, except for years ending in 2 and 7, when it is part of the CMF.

ASM/CMF data is used for eleven of the nineteen variables in the NBER-CES database: total employment (*emp*), total payroll (*pay*), production workers (*prode*), production worker hours (*prodh*), production worker wages (*prodw*), total value of shipments (*vship*), total cost of materials (*matcost*), total value added (*vadd*), new capital expenditure (*invest*), end-of-year inventories (*invent*), and the cost of fuels and electricity (*energy*). More detailed explanations of these variables can be found in the appendixes of various CMF publications, including the General Summary of the 1997 Economic Census

Manufacturing Subject Series.[1] Also note the data found in Table 1-1d of this and similar publications.

The above variables are denominated in millions of nominal dollars, except for labor variables that are denominated in thousands of workers or millions of worker hours. To convert nominal dollars to real ("fixed-base") dollars and calculate productivity factors, four different deflators are used. These are discussed next.

The only one of these variables to undergo substantial changes in definition in recent years was *invest*. The ASM reported new capital expenditures before 1997 (with used capital expenditures in an additional table), but switched to reporting the sum of new and used expenditures in 1997, with no separate breakout of individual components (even in CMF years). In order to maintain consistency with the earlier data, we adjusted the post-1996 data. The data on new and used expenditures from 1987, 1992, and 1996 were combined to derive a "fraction new" (new/[new+used]) for each industry, which was then used to adjust the post-1996 capital expenditure numbers to reflect only new expenditures. When updating our database through 2011, an error that affected the calculation of the "fraction new" in earlier versions was discovered and fixed. Therefore, it is possible that the values for capital stocks, investment deflators, and the two TFP indexes for the period 2003 – 2009 differ slightly from previous versions.

2.2 Industry Capital Stocks and the Investment Deflator (PIINV)

The ASM provides annual data for gross nominal capital expenditures on equipment and structures. However, it lacks information on depreciation rates and investment deflators, which are necessary in order to grow the real capital stocks in the NBER-CES database.

Depreciation rates are derived from Federal Reserve Board data (FRB). The FRB data contain equipment and structures stocks and expenditures in real and nominal terms for 4-digit NAICS industries, providing less industry detail than the 6-digit ASM/CMF data. Depreciation rates are calculated for each industry, backing them out of a standard perpetual inventory equation:

$$K_t = (1 - d_{t-1})K_{t-1} + I_t \tag{1}$$

In equation (1), *d* is the rate of depreciation, *K* is the real capital stock in years *t* and *t-1*, and *I* is the real capital investment in year *t*.

To calculate the investment deflator, price indexes for equipment and structures are computed from the ratios of nominal and real investment for the 4-digit NAICS industries in the FRB data. The final deflator values are created as the weighted mean of the two types of capital expenditures. The shares used for the weights are averaged over years *t* and *t-1*:

$$\Delta P_t = \sum_{n=1}^{2} \left[\frac{1}{2} \left(S_{n,t} + S_{n,t-1} \right) \Delta P_{n,t} \right] \tag{2}$$

In equation (2), ΔP is the annual change in the price level, *S* is the share of total capital investment for each category, and the summation occurs over *n*=2 categories (equipment and structures). Note that the data on investment shares come from the ASM/CMF, so they are available at the 6-digit level. The depreciation rates and the investment deflators are at the 4-digit NAICS industry level, so we assign the

[1] Available at www.census.gov/prod/ec97/97m31s-gs.pdf.

value of each 4-digit "parent" to each of its 6-digit NAICS "children". We then combine the industry-level depreciation rates and investment deflators with the ASM capital expenditures for equipment and structures, using the perpetual inventory method (equation 1) to calculate the industry's real stock of equipment (*equip*), structures (*plant*), and total real capital stock (*cap = equip + plant*).

2.3 The Shipments Deflator (PISHIP)

The price indexes used in the calculation of the shipments deflator come from the BEA GDP-by-Industry data, found at www.bea.gov under the "Industry" tab.

The BEA price index data contain several "roll-up" 6-digit codes that combined two or more 6-digit NAICS industries. For example, NAICS 311411 (Frozen Fruit, Juice, and Vegetable Manufacturing) and NAICS 311412 (Frozen Specialty Food Manufacturing) were "rolled up" into a single code, 311410. This problem is addressed by assuming that industries within each roll-up has the same price structure, and assigning the values of the parent to all associated 6-digit NAICS children codes.

2.4 The Materials Cost Deflator (PIMAT)

The materials cost deflators are calculated using data from the benchmark "use-make" (input-output) tables and the GDP-by-Industry data of the BEA. The raw data are available at the BEA website (www.bea.gov) under the "Industry" tab.

The use-make tables measure the amount spent by each industry on different material inputs. To create the overall material cost deflator for an industry, a price index for each individual input is identified and the price changes for the different inputs are aggregated, weighting by that industry's use of each input to calculate the year-to-year change in the industry's materials cost deflator.

In the process described above, inputs coming from a manufacturing industry employ that industry's price index for shipments (PISHIP above). Inputs from all other (non-manufacturing) sectors employ the sector's price index for gross output, taken from the BEA integrated industry accounts.

The main issue with the above procedure is the lack of a clear 1-to-1 correspondence between the industry codes used in different datasets. The input-output tables contained "use identifiers" (6-digit industry codes that ended in 0 or a letter) that covered multiple NAICS industries. We assign all 473 6-digit NAICS codes to the 279 IO-use industry identifiers, assuming that if industries fall under the same IO-use identifier, they had the same structure of inputs. BEA concordance tables provided some of the mapping information, but a small number of codes were matched manually based on industry description.

2.5 Energy Usage Data and the Energy Deflator (PIEN)

The energy deflator is based on each industry's expenditures on seven types of energy: electricity, residual fuel oil, light oils, liquefied petroleum, coal, coke, and natural gas. Previous updates used information for six types of energy (the "distillates" category was recently split into light oils and liquefied petroleum).

The data for six of the fuels in the current update (excluding electricity) come from the 2002 Manufacturing Energy Consumption Survey (MECS), available at the U.S. Energy Information Administration website (www.eia.gov/consumption/manufacturing/index.cfm). MECS provides energy consumption in trillions of British thermal units (BTUs) and the average price per million BTUs for each energy type. The MECS data (price × quantity) are used to calculate the share of total energy spending allocated to each energy type. Meanwhile, the ASM data provides information on electricity and non-

electricity energy spending. The MECS data are used to disaggregate the non-electricity spending into the six other fuels.

There are a number of missing values in the MECS data. Extremely small values (less than 0.5), as well as values with high relative standard error (more than 50%) are not published. Additionally, some values are withheld to avoid disclosing information for individual establishments in cases where an industry has relatively few plants.

All extremely small values were replaced by 0.1, consistent with the approach taken in previous updates. Since MECS reports totals by industry and type of energy, it is possible to impute most remaining missing values by running a cross-check against the unassigned residuals from those totals. Where this was not possible, the missing data was extrapolated using the MECS data from the previous data set update. National averages were used in place of missing industry-level energy price data.

Price indexes for each energy type were obtained from the BLS database (www.bls.gov). An earlier 1958-2002 version of our data used energy prices for commercial use. The present version uses industrial price indexes, which seems to fit the nature of the energy deflator better. The codes of the extracted BLS variables are as follows:

WPU051	Coal
WPU052	Coke oven products
WPU0532	Liquefied petroleum gas
WPU054321	Industrial electric power
WPU055321	Industrial natural gas
WPU0573	Light fuel oils
WPU0574	Residual fuels

Since the 1996 update, the *WPU052 Coke oven products* series has been discontinued. Thus, the coke price data used here are extrapolated from coal prices. Electricity prices are calculated from electricity consumption and expenditures data in the ASM tables, where available.

The aggregate energy price index is calculated as the weighted mean of the price changes for the seven different types of energy, following the structure of equation (2). It should be noted that the 1958-2002 version of the dataset appears to have contained some inconsistencies in the deflator calculation, so that version of the deflator series should not be used.

3 Industry Classification Changes

The system of industrial classification used by the Census Bureau to report manufacturing data has changed several times over the lifespan of the NBER-CES database. The original SIC 1958 system underwent two substantial modifications in 1972 and 1987 before it was finally replaced with the NAICS industry codes in 1997. Earlier versions of the NBER-CES database were provided for both SIC 1972 and SIC 1987 industry codes. The 2011 version is provided for NAICS 1997 and SIC 1987 industry codes.

The change in industry codification and the necessary conversion of existing data presented challenges to the extension of the NBER-CES data. First, unlike the earlier SIC-based ASM data, the new NAICS-based ASM data do not provide full industry detail in non-CMF years. Many 6-digit NAICS industries are only reported in “rolled-up” form, with two or more 6-digit industries within the same 5-digit

NAICS combined into a single value (designated by the 5-digit number plus one letter). These roll-ups were broken back down into their component 6-digit NAICS industries, using CMF year values to calculate shares and interpolating the shares between CMF years. When the last year of the dataset is not a CMF year, the shares from the latest available CMF year are used. For example, the current version ending in 2011 uses the 2007 CMF shares to calculate the values for 2008-2011, which should be considered "provisional". When the 2012 CMF data become available, the "final" values will be calculated for 2008-2011, using shares interpolated between the 2007 CMF and 2012 CMF.

The biggest hurdle arising from the SIC to NAICS transition was the entry, exit, and realignment of manufacturing industries. While there was direct correspondence between many 4-digit SIC industries and the new 6-digit NAICS industries, there were also many cases in which SIC industries were split or multiple pieces from SIC industries were joined to form a new NAICS industry. For example, SIC 2077 (Animal and Marine Fats and Oils) was split between NAICS 311613 (Rendering and Meat Byproduct Processing), 311711 (Seafood Canning), and 311712 (Fresh and Frozen Seafood Processing), while NAICS 311712 also included a part of SIC 2092 (Prepared Fresh or Frozen Fish and Seafoods).

The 2002 CMF concordance table linking NAICS 1997 and SIC 1987 codes is instrumental in the conversion process. This table provides industry mappings and shares for three variables in the NBER-CES dataset: employment, payroll, and shipments. The employment shares are used in the calculation of labor-related variables (*emp*, *prodh, prode).* The payroll shares are used in the calculation of compensation related variables (*pay*, *prodw*). The shipments shares are used for all remaining variables.

As pointed out by Bartelsman and Gray (1996), the use of this type of concordance is not without its pitfalls. Most notably, the composition of industries tends to change over time, and the use of a fixed concordance matrix may distort the data. While there is no easy fix for this situation, it should be recognized when analyzing long time-series of data for industries that have been redefined over time, or that come into and out of existence.

While most manufacturing industries could be reconstructed under either SIC or NAICS definitions, there are some cases in which entire industries (or large fractions of them) were reclassified out of (or into) manufacturing. Four industries were classified as manufacturing under NAICS that had been considered non-manufacturing under SIC, including Retail Bakeries, Tire Retreading, Dental Laboratories, and Software Reproducing. The initial capital stocks of these new industries were obtained from 1997 CMF data, though it should be noted that unlike the original creation of the starting capital stocks used for 1958 for all other industries, we have not examined these starting values for their sensitivity to using different accounting methods. The entry of these new industries forced us to shift the benchmark year to 1997 for the NAICS version of the database, so that deflator series could be properly defined (starting as 1.000 in 1997). These new manufacturing industries in the NAICS database have missing values for all years before 1997. On the other hand, six SIC manufacturing industries were moved completely to non-manufacturing under NAICS, and therefore have missing values for all years after 1996. Most notably, in the NAICS system publishing was reclassified in its entirety to the Information sector (NAICS 51).

Some industries were split between manufacturing and non-manufacturing in the move from SIC to NAICS. Unlike earlier versions (where we retained only those industries which kept at least 95% of their activity in manufacturing), we created a new variable (PCTNMFG) that measures the fraction of the SIC manufacturing industry's shipments that moved to a non-manufacturing NAICS industry (or vice versa for the NAICS version of the file). We then adjusted the partial industry numbers after the switch, dividing them by (1-PCTNMFG), to avoid misleading shifts in the industry's data around 1997.

For example, NAICS 336612 (Boat Building) does not include the 13% of SIC 3732 (Boat Building and Repairing) associated with "repairing" which was reclassified out of manufacturing. In this case we multiply the post-1997 values of NAICS 336612 by 1.15 (=1/[1-0.13]) in the SIC version of the database to remove the discontinuity in 1997. Similar adjustments to pre-1997 values were done in the NAICS version of the database for those industries that were partially non-manufacturing under SIC.

Table 1 lists the industries that were part of this realignment of the manufacturing sector between SIC and NAICS.

Table 1. SIC-NAICS Manufacturing Realignment in 1997.
(partial realignment cases – percent non-manufacturing in parentheses)

Entering Industries

NAICS Code	*Name*
311330	Confectionery Manufacturing from Purchased Chocolate (.025)
311340	Non-Chocolate Confectionery Manufacturing (.018)
311612	Meat Processed from Carcasses (.072)
311811	Retail Bakeries
313311	Broadwoven Fabric Finishing Mills (.188)
314121	Curtain and Drapery Mills (.151)
315999	Other Apparel Accessories and Other Apparel Manufacturing (.057)
326212	Tire Retreading
334611	Software Reproducing
334612	Prerecorded Compact Disc (except Software), Tape, and Record Reproducing (.420)
335312	Motor and Generator Manufacturing (.037)
337110	Wood Kitchen Cabinet and Countertop Manufacturing (.073)
337121	Upholstered Household Furniture Manufacturing (.024)
337122	Non-Upholstered Wood Household Furniture Manufacturing (.029)
339116	Dental Laboratories

Exiting Industries

SIC Code	*Name*
2411	Logging
2711	Newspapers: Publishing, or Publishing and Printing
2721	Periodicals: Publishing, or Publishing and Printing
2731	Books: Publishing, or Publishing and Printing
2741	Miscellaneous Publishing
2771	Greeting Cards
3732	Boat Building and Repairing (0.127)

In addition, some NAICS industry definitions in the manufacturing sector changed during the 2002 and 2007 NAICS updates. Several of these changes caused the movement of sizable fractions between industries, making the data inconsistent over time without additional adjustments. NAICS 327992 (Ground/Treated Mineral and Earth) and NAICS 339912 (Silverware/Hollowware) have to be scaled up, since a substantial share of each was reassigned to different industries. NAICS 339111 (Laboratory Apparatus and Furniture) was discontinued in 2007, with most of its contents moving to NAICS 339113 (Surgical Appliances and Supplies). There were also changes between 2002 and 2007 which caused

parts of NAICS 339112 and 339113 to move to NAICS 339111. These latter changes were reversed, and post-2007 NAICS 339111 values are imputed based on the last available ratio of the sizes of 339111 and 339113. Other sets of industries are similarly adjusted for "moving parts": NAICS 331422, 335929, and 331491; NAICS 333311 and 333313; NAICS 334220 and 334419; NAICS 339912, 332812 and 332999.[2]

4 Productivity Measures

The dataset includes two different multi-factor productivity measures: 4-factor productivity, including capital (K), production worker hours (N), non-production worker hours (L), and materials (M), and 5-factor productivity, which splits the materials variable into energy (E) and non-energy materials (M-E). The 5-factor productivity factors are calculated as follows:

$$\widehat{TFP} = \hat{Q} - \sum_i \alpha_i \hat{X}_i \, , i \in K, N, L, M, E \tag{3}$$

Here Q is real output, α_i is the average share (current and lagged year) of factor *i* relative to industry shipments, X_i is the real input of factor *i*, and ^ denotes the first difference of the logarithm (growth rate). The share of capital is computed as one minus the sum of the other factor shares.

5 Conclusions

This document discusses issues related to the 1958-2011 version of the NBER-CES productivity database. Further updates will follow as more years of data become available.

If you have questions, comments or problems to report about the database, please contact Dr. Wayne Gray at wgray@clarku.edu or Dr. Randy Becker at randy.a.becker@census.gov. Feel free to use the data in your research, with a citation of the appropriate NBER technical paper - we hope it will be useful.

6 References

Bartelsman, E. and W. Gray. 1996. "The NBER Manufacturing Productivity Database". Technical Working Paper 205, National Bureau of Economic Research.

[2] For more details of the specific magnitude of these adjustments, please consult the concordance tables at www.census.gov/eos/www/naics/concordances/concordances.html.