

NAICS AND THE 2002 HISTORICAL REVISION OF INDUSTRIAL PRODUCTION, CAPACITY, AND CAPACITY UTILIZATION

Norman J. Morin

Board of Governors of the Federal Reserve System, Washington, DC 20551

norman.j.morin@frb.gov

Key words: Industrial production; capacity; capacity utilization; NAICS; industry classification

Abstract. The primary feature of the 2002 historical revision of industrial production, capacity, and capacity utilization was the reclassification of these measures to a North American Industry Classification System (NAICS) basis back to 1972; formerly the production, capacity, and utilization indexes had been classified according to the Standard Industrial Classification (SIC) system. The datasets that required restatement on a NAICS basis in order to reconstruct the industrial production and capacity system are detailed, and the methods used to reclassify these datasets are discussed. The historical consistency of industrial production and capacity was further improved by recompiling the new NAICS-based indexes using methods formerly applied only to recent data. Methods for grouping individual industry indexes useful for analysis of industrial production, by market group and by stage of process, were refined. The overall effects of these changes are explored.

1. Introduction

The Federal Reserve Board began publishing monthly estimates of industrial production (IP) in December 1922 (the time series begins in January 1919); both the scope of the coverage and its detail have increased in the intervening years. The Federal Reserve Board created for internal use estimates of industry capacity and capacity utilization in the mid-1950s and began publishing these measures in the 1960s. In April 1990, the industry structure of the capacity and utilization rate measures was reorganized in order to construct a more integrated system of output, capacity, and utilization measures for total industry and a variety of market and industry sub-aggregates (see Raddock, 1990). Currently, indexes of real output and corresponding measures of capacity and of capacity utilization cover manufacturing, mining, and electric and natural gas utilities (see Stevens, 2003 for a more detailed discussion).

Historically, the industries in the IP system were classified according to various versions of the Standard Industrial Classification (SIC) system. In

December 2002, the Federal Reserve Board published a historical revision of industrial production, capacity, and capacity utilization (henceforth, the IP system) for which the primary feature was the reclassification of these measures on a North American Industry Classification System (NAICS); see Corrado (2003).¹ The restatement of the IP system on a NAICS basis back to 1972 extends further into history and is more extensive than the NAICS conversions for other statistical measures (see Stevens, 2003).²

The real output index for an individual industry is based on a monthly production indicator and a correction factor that aligns the monthly production indicator to comprehensive annual measures of output, or benchmark; typically the benchmark factor is a measure of the value-added from the Census Bureau's Census of Manufacturers (COM) and Annual Survey of Manufacturers (ASM) that is deflated by a price deflator from the Bureau of Economic Analysis or by a price index constructed by the Federal Reserve from detailed data. Where available, an industry's monthly production indicator is a measure of physical output (e.g., tons of steel, barrels of oil) from government or industry sources. If a measure of physical output is unavailable, production is inferred from monthly estimates of inputs to production, either hours worked in the industry (from the Bureau of Labor Statistics Establishment Survey) or the electricity used by the industry (from a Federal Reserve survey).

The Federal Reserve Board's capacity indexes attempt to capture the concept of sustainable practical output. Initial implied capacity indexes are

¹NAICS was created by the Economic Classification Policy Committee of the United States, Statistics Canada, and Instituto Nacional de Estadística Geografía e Informática of Mexico. The intent was to develop in common an industrial classification system that groups together economic units that have similar processes for transforming inputs into outputs. See Office of Management and Budget (2002).

²For example, in May 2001 the Bureau of the Census reclassified on a NAICS basis data from its Monthly Shipments, Inventories, and Orders Survey going back to 1992.

constructed by dividing a production index by a benchmark utilization rate.³ The contours of the implied capacity indexes are further refined to give consideration to additional indicators of capacity; baseline estimates of capacity are derived from the fitted values of regression models that relate the implied capacity indexes to the other measures. For industries whose capacity indexes are based on measures of physical output, the additional indicators include the physical capacity measures and deterministic trends; for industries whose capacity indexes are based on utilization rates from the Census Bureau's Survey of Plant Capacity (SPC), the additional measures of capacity include Federal Reserve estimates of industry capital input.⁴ A monthly time series is formed by interpolating the annual baseline capacity estimates produced by the regression models; for a fuller discussion, see Corrado, Gilbert, and Morin (1999). The industry-level monthly capacity utilization measures are calculated from the ratio of the production and capacity indexes.

2. Translation of Databases to a NAICS basis

As illustrated in the previous section, constructing estimates of industrial production, capacity, and capacity utilization requires several separate datasets, most prominently:

- (1) Bureau of the Census' COM and ASM
- (2) Physical output and capacity data from various government and trade sources
- (3) Hours worked by production workers from the Bureau of Labor Statistics' Establishment Survey.
- (4) Electric power use by industry from the Federal Reserve electric power survey
- (5) Capacity utilization rates from the Bureau of the Census' SPC
- (6) Bureau of Economic Analysis' Capital Flows Tables

³ The benchmark utilization rates are derived from data in physical units from government or trade sources (e.g. tons of steel that the industry can produce) or, absent available physical output data, from the Census Bureau's Survey of Plant Capacity (SPC). The SPC is funded jointly by the Federal Reserve and the Department of Defense.

⁴ The capital input measures are constructed by aggregating asset-by-industry capital stock data that, in turn, are developed from industry-level investment data from the COM and ASM; from asset-level investment data and price deflators from the BEA's National Income and Product Accounts; and from the BEA's capital flows tables, which provide a detailed breakdown of the asset composition of industry investment.

Reconstructing the IP system from 1972 to 2002 on a NAICS basis required each of these databases to be reclassified according to NAICS.

(1) *Bureau of the Census' COM and ASM.* The COM/ASM data are of paramount importance in constructing the IP system.⁵ In addition to the benchmarking role played by COM/ASM data that was mentioned in section 1, the analytically useful production, capacity, and utilization rate industry aggregates and market groups are produced by chain-aggregating individual industry-level indexes weighted by COM/ASM value-added. Moreover, COM/ASM SIC-to-NAICS mappings are the source for the reclassification of the several datasets that are used to build the IP system and to produce the monthly estimates.

At the most disaggregated level, 4-digit SIC to 6-digit NAICS, the majority of SIC industries in manufacturing map to one NAICS industry (335 out of 460). Because the remaining 4-digit SIC industries in manufacturing map to multiple (between 2 and 13) 6-digit NAICS industries, the translation from an SIC-based to a NAICS-based IP system is not straightforward. The first, and most important, step is the SIC-to-NAICS translation of the COM/ASM data. The industry data in previous COMs were classified by SIC, except the 1997 COM. The Census classified industry data in the 1997 COM on both an SIC and a NAICS basis, which allows one to determine the share of each SIC that maps to a corresponding NAICS industry.⁶ These mappings, however, only specifically pertain to 1997, and the industry structure from earlier years need not accurately be reflected in the 1997 distribution.⁷ For all COMs back to 1963, however, Bayard and Klimek (2003) assigned each plant to a NAICS

⁵ The Census of Manufacturers occurs every five years. Between COMs, the Annual Survey of Manufacturers is released, which is based on a statistical sample of plants in the COM; adjustments are made for plant births and deaths.

⁶ This exercise can be performed separately for the dozens of variables in the COM. The most pertinent measures for the IP system are value added, cost of materials, capital spending on equipment and on structures, purchased electric power, production worker hours, shipments, and end-of-year inventories.

⁷ See Stevens (2003) for details of a stark example using autos and light trucks, which are part of the same 4-digit SIC code (3711), but inhabit separate NAICS codes (336111 and 336112, respectively).

category (the number of manufacturing establishments per COM more than 300,000).⁸

Beginning with the assignments of each plant in the 1997 COM to both an SIC and a NAICS industry, Bayard and Klimek (2003) moved sequentially through the previous COMs and classified a plant in a given SIC industry as an exact match for a specific NAICS industry if (1) the SIC industry mapped to a unique NAICS code in 1997; (2) the plant was in the 1997 COM and was, therefore, classified in a particular NAICS industry; or (3) product information related to the plant could align it with a great deal of certainty to a specific NAICS industry. If none of these criteria could be satisfied, a plant was assigned to a specific NAICS industry with a model-based statistical procedure that compared the plant to other plants in the same SIC that had been assigned to specific NAICS industries. Based on the characteristics reported in the COM of the plant relative to those other plants (such as shipments per worker, hourly wages), the probability of the plant being in the various NAICS industries to which that SIC maps were calculated, and the plant was randomly assigned to a specific NAICS industry.

Once all plants in each COM back to 1963 were allocated to NAICS industries, the resulting mappings from SIC to NAICS at the plant level were aggregated over all plants to produce total 4-digit SIC to 6-digit NAICS assignments that varied by Census year. The non-constant nature of U.S. industrial structure is evident in the time-variation of the derived mappings. For example, in SIC 3711 (motor vehicles and car bodies) in the 1977 COM, 78 percent of the industry value-added is assigned to NAICS 336111 (autos); in 1987 the share is 64 percent; and in 1997 the assignment is 40 percent.

To reclassify the variables in the IP system, the variable shares constructed by Bayard and Klimek (2003) for the COMs from 1963 to 1997 were linearly interpolated to an annual frequency. Using these intercensal mappings of the COMs, variables in the ASMs were recalculated on a NAICS basis. This yielded a complete dataset of NAICS-based series from 1963 to 2000 (the COM and ASMs were surveyed on a NAICS basis from 1997 on).

(2) *Physical measures.* Physical output data from various government and trade sources are used as the monthly indicator for about 46 percent of industrial

production, and physical capacity measures account for about 16 percent of total industrial capacity. These data are generally at a very detailed level, therefore for essentially each series there existed a one-to-one mapping between the SIC and NAICS classification.

(3) *PWH.* Detailed industry-level production worker hour data from Bureau of Labor Statistics Establishment Survey were one of the series from the COM/ASM for which a time series of SIC-to-NAICS mapping were constructed (4-digit SIC to 6-digit NAICS), based on the work of Bayard and Klimek (2003). Constructing a dataset of NAICS-based BLS PWH data was a fairly straightforward application of the times series of SIC-to-NAICS shares. Occasionally complicating matters were those industries for which, for disclosure reasons, the BLS combines multiple 4-digit SIC industries.

(4) *KWH.* Similar to the PWH data, translating the SIC-based KWH data was a straightforward application of the 4-digit SIC to 6-digit NAICS shares derived from the COM/ASM series on purchased electric energy. About 25 percent of the Federal Reserve's monthly data on electric power use by industry is actually reported on a NAICS basis; the remainder continues to be reported on according to SIC. To construct industrial production each month, SIC-based and NAICS-based reports on electric power must be combined to produce useable measures of industry-level electric power.

(5) *SPC.* About 80 percent of industrial capacity is constructed from utilization rates from the SPC. It was necessary to recalculate detailed industry-level utilization rates from the SPC on a NAICS basis for pre-1997 data; as with COM/ASM, the SPC was surveyed on a NAICS basis after 1996. The SPC utilization rate for NAICS industry N at time t , $U_{N,t}$, is computed according to equation (2.1). The total value added for the NAICS industry is divided by the NAICS-based industry's implied capacity. The latter is computed by summing over the implied capacities of each SIC-based industry that maps into NAICS industry N , weighted by the shares of each SIC industry that maps to NAICS industry N . Let $share_{i \rightarrow N}^{VA,t}$ represent the share at time t of value added in SIC industry i , $VA_{i,t}$, that is also in NAICS industry N , and let $U_{i,t}$ represent the SPC utilization rate for SIC industry i at time t . The SPC utilization rate for industry N at time t is:

⁸ Bayard and Klimek (2003) dual-coded each plant for Census years 1992, 1987, 1982, 1977, 1972, 1967, and 1963.

$$U_{N,t} = \frac{\sum_i VA_{i,t} \times share_{i \rightarrow N}^{VA,t}}{\sum_i \frac{VA_{i,t}}{U_{i,t}} \times share_{i \rightarrow N}^{VA,t}} \quad (2.1)$$

(6) *CFT*. The BEA's capital flows tables measure the asset allocation of industry investment in economic census years (but lag 5 years behind the most recent COM). It was a relatively straightforward task to use the investment shares for equipment and for structures to translate the SIC-based allocation of asset level investment to a NAICS basis. The resulting tables, which also go back to 1963, were used as initial estimates in the routine that allocates annual industry level capital spending data on equipment and on structures across asset categories. This allows for, ultimately, the construction of the industry-level measures of capital stocks and capital inputs that are a primary determinant of industry capacity. See Gilbert and Mohr (1996).

3. Application of Current Methods to Earlier Data

To ensure consistency of the production, capacity, and utilization rate measures, refinements to methods used in the construction of the IP system that had been introduced during the last several years and implemented only on recent data were applied to the NAICS-based indexes over history. Among the several refinements that were extended back to at least 1972, the five most prominent include (1) chain-weighting, (2) improved seasonal adjustment techniques, (3) removing systematic weather effects from the electric power data, (4) ASM drift, and (5) an updated procedure that allows monthly capacity indexes to change smoothly.

Individual industry-level production indexes are aggregated from 1972 to the present using a chain-type procedure with monthly weights. Formerly, chain weighting was used to aggregate data back to 1977 (originally implemented in the 1997 revision; see Corrado, Gilbert, and Raddock, 1997); the procedure employed annual weights between 1977 and 1992 and monthly weights from 1992 to the present (implemented in the 2001 revision; see Corrado, Gilbert, and Morin 2002). The data from 1972 to 1977 were aggregated with a linked-Laspeyres procedure.

Improvements in seasonal adjustment techniques were applied to the production indexes going back to 1972. These methodological improvements, which had been applied to data back to 1987, include regression-based adjustments for the effects the

timing of holidays and for various calendar effects (e.g., timing factors for 4-week versus 5-week reporting periods for the BLS PWH data, trading day adjustments), and an improved trend extraction technique used by the Federal Reserve in lieu of the standard procedure in the X-12 seasonal adjustment package.

Production indexes that are based on industry electric power use, about 19 percent of total industrial production, were refined by removing the systematic effects of weather on the seasonally adjusted electric power series. These adjustments, which had been introduced in the 1998 revision and applied to data back to 1992, make it possible to, at least in part, disentangle an increase in electric power use from which a gain in production should be inferred versus an increase in electricity use because of unseasonably warm summer or unseasonably cold winter weather (see Gilbert and Raddock, 1999).

The annual estimates of real output with which the monthly estimates of IP are benchmarked primarily derive from the ASM and, in census years, the COM. The COM represents the universe of domestic manufacturing plants and the ASM is a statistical sample based on a panel that is drawn from the most recent COM. Historically, between census years the ASM figures progressively understate economic activity because of sample erosion (called "ASM drift"); conversely, the change between the final year before a census and the census year is magnified as the cumulative understatement is made up in a single year (see Raddock, 1993 and Raddock, 1995). The Federal Reserve had previously made estimates at the detailed industry level of the size of the ASM drift for the period leading up to the 1992 COM by looking at the difference between year-end inventories in the ASM for the year before a census and the beginning-of-year inventories in the COM (the two magnitudes should be essentially identical). In the historical revision the ASM drift adjustments were made back to 1982.

The final notable methodological change applied to earlier data was an interpolation procedure with which the monthly increases in industry capacity evolved more smoothly. These methods were introduced in 1999 (see Corrado, 2000) and had been applied to capacity indexes back to 1992. The prior methods kept monthly rates of change constant during a calendar year, and resulted in possibly abrupt jumps in rates of change that straddled the beginning of a year.

Table 4.1. Major market groups

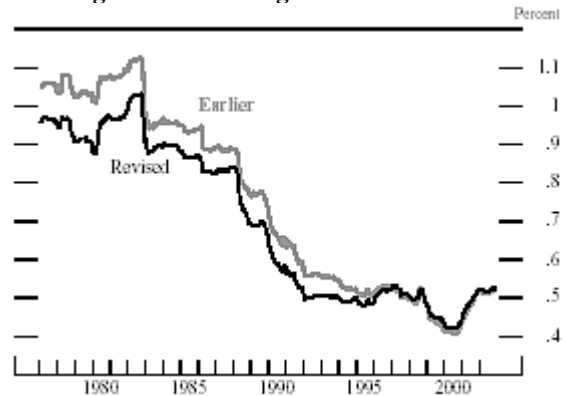
Final products and nonindustrial supplies
Final products
Consumer goods
Equipment
Nonindustrial supplies
Construction supplies
Other business supplies
Materials
Non-energy
Energy

4. Market Groups and Stage of Process Classification

The production indexes in the IP system are combined into analytically useful aggregates that allow a researcher to assess trends in activity according to both (1) the demand categories for which specific industry production is an input; and (2) the stages in the overall production process from crude materials to finished goods of the output of various industries. The former describes the market group structure of industrial output, and the latter describes the stage-of-process categorization of industrial output.

The market group structure of industrial production classifies output in two broad categories, the materials group and the final products and nonindustrial supplied group. The materials group represents output from an industry that will be an input to the production process of another manufacturer. The other major market group, final products and nonindustrial supplies, represents production that is an input to demand outside the industrial sector. Output of a particular industry can be divided among multiple market groups (e.g., personal computers are part of both business equipment and consumer durable goods). Prior to the revision, the output of a fairly small number of industries (26 of more than 275) was split across market groups based on detailed end-use statistics. In the historical revision, the output of all industries has the potential to be split among market groups. To what market groups an industry's output is divided and the relative shares that are allocated to the particular market groups were derived using the BEA's 1992 input-output (IO) tables. The IO tables parse the gross value of output (in producers prices) between its use as an intermediate input and its allocation to final demand categories (such as consumption, investment, government, and exports)

Figure 5.1. Percent Change in Monthly IP Eight-Year Moving Standard Deviation



Industry output is also grouped together by stage-of-process, so that the output of one stage is an input into subsequent stages. Formerly, industries were allocated to primary processing and advanced processing groups. With the historical revision, the number of stage-of-process groups was expanded to four: crude, primary, semi-finished, and finished processing. The IO tables were utilized to array, as close as possible, the individual industries into a block triangular matrix such that output of industry in block i,j can be used as an input to production in blocks further down the matrix.

5. Results of the revision

The cumulative effect on total industrial production of the change in classification systems and the methodological improvements is fairly small; the picture of industrial activity between 1972 and the present is not materially different. Between 1972 and 1987, the average annual increase in total IP (measured fourth-quarter to fourth-quarter) is only 0.2 percent less than was previously estimated. After 1987, the average annual increase was unchanged. The aggregate capacity index for total industry was revised by a similar proportion, so capacity utilization was, on average, little different as a result of the revision.

The business cycle peaks and troughs in monthly IP were also little changed by the revision. Of the five peaks and five troughs since 1972, only two were altered; the peak formerly placed at May 1979 is now estimated to have occurred in March 1979; the peak formerly in April 1989 is placed at September 1990. Industrial activity before both peaks in the revised and in the earlier data was fairly flat, so the picture of the business cycle episodes with the revised peak dates is unchanged.

As indicated above, aggregate industrial activity at the low and medium (business cycle) frequencies is essentially the same as in the earlier data. At the monthly frequency, the extension over history of current methods, such as the chain weighting and the seasonal adjustment procedures, reduced slightly the variance of the monthly rates of increase. As shown in figure 5.1, a moving eight-year standard deviation of the revised monthly percent changes is about 0.1 percentage point less than in the earlier data through early 1980s—the period over which the revised and earlier data have the greatest differences in methodology. The rolling standard deviations are almost identical over the last several years—where there are essentially no methodological changes pre- and post-revision and the high-frequency revisions mainly reflect updated source data. Overall, the correlation between the monthly percent changes of the revised and earlier data is 0.91; between quarterly averages of the monthly data, the correlation is 0.97.

6. Conclusion

The recent reclassification of the IP system from an SIC to a NAICS basis represented a major advance among government statistical agencies. The span over which the industry-level reclassification and market and stage-of-process assignments were performed and over which consistent methods of data construction and aggregation were applied is of great importance and enormous usefulness for economic analysis. The overall effect on aggregate production, capacity, and capacity utilization was very small, and resulted in a materially unchanged picture of the industrial sector in terms of its long-run trends, cyclical activity, and monthly variation.

References

Bayard, Kimberly and Shawn Klimek (2003), “Reclassifying the Census of Manufactures from the Standard North American Classification System to the North American Classification System, 1963 to 1992,” *2003 Proceedings of the American Statistical Association*, Business and Economics Statistics Section, Alexandria, VA.

Corrado, Carol A., Charles E. Gilbert, and Richard D. Raddock (1997), “Industrial Production and Capacity Utilization: Historical Revision and Recent Developments,” *Federal Reserve Bulletin*, v83, pp. 67-92.

Corrado, Carol A., Charles E. Gilbert, and Norman J. Morin (1999), “Industrial Production and Capacity Utilization: The 1998 Historical and Annual Revision,” *Federal Reserve Bulletin*, v85, pp. 151-176.

Corrado, Carol A. (2000), “Industrial Production and Capacity Utilization: The 1999 Annual Revision,” *Federal Reserve Bulletin*, v86, pp. 132-148.

Corrado, Carol A., Charles E. Gilbert, and Norman J. Morin (2002), “Industrial Production and Capacity Utilization: The 2001 Annual Revision,” *Federal Reserve Bulletin*, v88, pp. 132-148.

Corrado, Carol A. (2003), “Industrial Production and Capacity Utilization: The 2002 Historical and Annual Revision,” *Federal Reserve Bulletin*, v89, pp. 151-176.

Gilbert, Charles C., and Michael F. Mohr (1996), “Capital Stock Estimates for Manufacturing Industries: Methods and Data,” Board of Governors of the Federal Reserve System working paper.

Gilbert, Charles C., and Richard D. Raddock (1999), “Industrial Production and Capacity Utilization: The 1998 Annual Revision,” *Federal Reserve Bulletin*, v85, pp.20-33.

Office of Management and Budget (2002), *North American Industry Classification System*.

Raddock, Richard D. (1990), “Recent Developments in Industrial Capacity and Utilization,” *Federal Reserve Bulletin*, v76, pp. 411-435.

Raddock, Richard D. (1993), “Industrial Production, Capacity, and Capacity Utilization Since 1987,” *Federal Reserve Bulletin*, v79, pp. 593-605.

Raddock, Richard D. (1995), “Industrial Production and Capacity Utilization: A Revision,” *Federal Reserve Bulletin*, v81, pp. 16-26.

Stevens, John J. (2003), “Overhauling Industrial Production: The 2002 Annual Revision,” *2003 Proceedings of the American Statistical Association*, Business and Economics Statistics Section, Alexandria, VA.