CREATING A HISTORICAL BRIDGE FOR MANUFACTURING BETWEEN THE STANDARD INDUSTRIAL CLASSIFICATION SYSTEM AND THE NORTH AMERICAN INDUSTRY CLASSIFICATION SYSTEM

Kimberly N. Bayard and Shawn D. Klimek
Board of Governors of the Federal Reserve and the U.S. Census Bureau

I. Introduction

Since the 1930s, the U.S. government has classified industries according to the Standard Industrial Classification (SIC) system. The SIC system has been periodically updated and revised to mirror changes in the underlying structure of the economy; the most recent change was in 1987. In the early 1990s, a movement began to construct a more sweeping revision to industry classification. In 1997, representatives from the statistical communities of the U.S., Canada, and Mexico finalized the new North American Industry Classification System (NAICS). NAICS was designed to reflect the significant changes in the economy in the 1990s and to standardize industry definitions across North America.[1]

The Census Bureau was one of the first U.S. government entities to collect and publish NAICS-based data for the manufacturing sector with the release of the 1997 Census of Manufactures (COM). In addition, the Census Bureau published SIC-based statistics for the 1997 COM, and also issued a correspondence between the two classification systems. However, the Census Bureau did not release NAICS-based manufacturing statistics for any year prior to 1997.

This paper documents a joint Federal Reserve Board-Census Bureau project to convert the plant-level information in the Censuses of Manufactures from the SIC system to NAICS, and describes the subsequent historical SIC-NAICS concordance that arises. Economists at the Federal Reserve Board used this concordance to obtain, back to 1972, NAICS-based output benchmarks. These benchmarks were used in the construction of the recently published NAICS-based historical revision of the index of Industrial Production (IP) and the related measures of capacity and capacity utilization.

The issuance of a NAICS-based industry time series for IP fills a gap for users of industry statistics. The statistical community's transition to NAICS created a break in industry-level data that has made it difficult at best to perform basic manipulations such as trend analysis, forecasting, and seasonal adjustment. The Federal Reserve Board's issuance of historical NAICS-based time series for IP provides data users with many of the tools necessary to perform standard analyses.

II. Motivation and Data Source

One of the most useful data products provided by the Census Bureau following the issuance of the 1997 COM industry statistics (and, indeed all Economic Censuses) is the official bridge table between SIC and NAICS. This bridge provides information for each four-digit SIC industry that allows one to calculate the share of establishments, shipments, employment, and payroll associated with its corresponding six-digit NAICS industries. Table 1 shows an example of the type of information conveyed in the bridge table for SIC 3578, (Calculating & accounting machines, except electronic computers), and how this industry corresponds to its two associated NAICS industries (Office machinery manufacturing and other computer peripheral equipment manufacturing).[2]

Table 1: Sample bridge between SIC and NAICS for SIC 3578 (1997)

NAICS		Value of Shipmts ($1,000)	Shr of Ship.	Paid Empl	Shr of Empl.
	Calc. & acc. machines, ex. comp.	2,014,806	--	7,683	--
333313	Ofc mach. mfg	144,380	0.07	966	0.13
334119	Other comp. peripheral equip. mfg	1,870,426	0.93	6,717	0.87

From the table above, it is clear that, in 1997, most of SIC 3578 corresponded to "other computer peripheral equipment manufacturing" (NAICS 334119), which includes products such as computer monitors, keyboards, printers, and automatic teller machines. A much smaller share of SIC 3578 corresponded to "office machinery manufacturing" (NAICS 333313), which includes

[1] A history and description of the various industry classification systems can be found in the manuals (Executive Office of the President, 1998, 1987, 1972).

[2] Some industry concordances are omitted for failure to meet disclosure requirements. An electronic copy of the complete bridge table can be found at http://www.census.gov/epcd/ec97brdg.

the production of mailhandling machinery and equipment, calculators, and typewriters.

The table also illustrates why it is somewhat limiting that the only publicly available SIC-NAICS bridge that exists is for 1997. One might expect that the value of shipments of typewriter and calculator manufacturing relative to computer monitors and printers was much greater in, say, 1972 than it was in 1997. However, because there is no available information for prior years, data users can only speculate as to how to construct the relevant breakdown.

The absence of historical industry data that allowed for the construction of a bridge between SIC system and NAICS led us to construct such a bridge directly from plant-level data. The Center for Economic Studies at the Census Bureau maintains the Longitudinal Research Database (LRD), a database of plant-level records for all establishments that appeared in either a COM or an ASM since 1963. We used these microdata files to convert the industry code for each establishment in each COM from the SIC basis to NAICS.

Specifically, we rely on four information sources to convert the censuses to a NAICS basis: the Census Bureau's product level concordance table that shows a translation of 1987 SIC product codes to 1992 SIC product codes to 1997 NAICS product codes; the 1997 census that contains both SIC and NAICS codes for each establishment; plant-specific identifiers that do not change over time, and the existence of both current-year and the 1987 SIC equivalent industry codes for all establishments on census files prior to 1997. These last two pieces of information result from CES-specific work to standardize over time all files in the LRD. As for the industry information, because all LRD establishments have 1987 SIC information, the translation of SIC to NAICS is possible even when a single NAICS code corresponds to multiple SIC codes.

III. Methodology

Our general approach to assign establishments to NAICS industries takes place in reverse chronological order. The 1997 COM contains both SIC and NAICS industry codes and provides our initial link between SIC and NAICS classifications. As described in more detail below, we use the 1997 classifications as a base to help recode the 1992 COM to NAICS. Once all establishments in the 1992 COM are reclassified to NAICS, these industry codes are then used to help in the reclassification of the 1987 COM, and this process is repeated back to the 1963 COM.[3]

Within each COM year, we assign establishments to NAICS industries in a series of successive steps. Because the LRD contains a lot of information for each establishment, there are often many options for how we wish to go about the industry assignment process. We believe, however, that some methods are likely to be more reliable than others. We first attempt to assign each establishment to a NAICS industry by the most reliable method; any establishment that can't be matched is considered to be a residual and we attempt to match it in a subsequent step.

Briefly, we assign establishments to industries in one of two general ways: exact matching or statistical matching. The exact matching procedures include a "product method" where we classify the types of products made by the plant, an "industry method" where we classify a plant to a NAICS industry based on its SIC industry, an "establishment method" where we identify the SIC industry of an establishment at a certain point in time and use the plant's presence and industry code in earlier censuses to assign it to a NAICS industry. The statistical assignment procedure looks at a set of characteristics of each plant to help determine whether they are more consistent with assignment of that plant to one particular NAICS industry or another. There are a few "special cases," that is, plants that we cannot assign by one of the methods listed above, and we resort to either shipments-weighted or manual assignment.

Step 1: Product Method

The first step in reclassifying an establishment to a NAICS industry uses product level information collected by the Census Bureau for some establishments in the COM. The product code is essentially an extension of the SIC-based industry code where the first few digits are identical to the industry code and the last several digits specify a particular product within that industry. The plant is assigned to the industry that corresponds to the products that represent greatest value of the establishment's shipments. A plant may make several products, some of which may fall outside of the establishment's assigned industry (see below for an illustration).

We use a concordance that maps these SIC product codes to NAICS product codes. Sometimes, we can assign NAICS product codes to all of an

[3] COMs are generally conducted every five years. The LRD contains information from the COMs conducted in 1997, 1992, 1987, 1982, 1977, 1972, 1967, and 1963.

establishment's products; other times we can only assign a subset of them. When we assign NAICS product codes to the majority of the establishment's overall shipments, we then use these product codes as a guide to determine the NAICS industry associated with the greatest value of shipments for each establishment.

Consider the following hypothetical plant that produces three products and whose primary industry is 2282 (represented by the first four digits of the product code):

Product 1: SIC_{92} =2282231, shipments=4000 → $NAICS_{97}$ =313112xxxx
Product 2: SIC_{92} =2282241, shipments=3000 → $NAICS_{97}$ =313312xxxx
Product 3: SIC_{92} =2269021, shipments=3000 → $NAICS_{97}$ =313312xxxx

In this case, 70 percent of the plant's shipments in 1992 are in SIC 2282. After matching in NAICS product data, we see that the primary industry for the establishment derived from the product codes is NAICS 313312, which accounts for 60 percent of the establishment output. If we had assigned a NAICS industry to the establishment solely based on the primary four-digit SIC industry, then the plant would be classified into industry 313112.[4]

We can also assign a NAICS industry when only a subset of SIC products match to a unique NAICS product code. For example, for the hypothetical plant above, even if we had been able to match only product 2 and product 3 to NAICS products codes, we would have been able to assign the establishment to the proper industry. Other combinations of products matched to NAICS product codes are insufficient to assign the plant by this method. For all cases where we cannot assign more than half of the plant's shipments to a NAICS industry, we turn to another method of industry assignment.

Step 2: Industry Method

The next method of assigning NAICS industries uses an industry-level concordance file. Many four-digit SIC industries correspond to only one six-digit NAICS industry; we assign plants in these SIC industries to the relevant NAICS industry. Only establishments in those industries with a one-to-one correspondence can be assigned by this method. More than 80 percent of all SIC industries in manufacturing are represented by this one-to-one scenario. The remaining SIC industries are split into as many as seven NAICS industries.

[4] The NAICS manual shows that SIC 2282 matches only to NAICS 313112.

Step 3: Establishment Method

The next method of assigning NAICS industries to manufacturing establishments uses unchanging establishment identifiers to link plants across censuses. Recall that our assignment approach takes place in reverse chronological order. Therefore, once we have assigned a NAICS code to an establishment in, say, 1992, we can use this information to help inform the assignment of the plant in 1987. We use the plant identifiers to match establishments from the census in year t to the census in year t-5, and if the plant operates in the same four digit SIC in both censuses, then we assign the NAICS industry for the establishment in year t to the establishment in year t-5. Implicitly, this method assumes that there is no change within the establishment that would cause a change in classification between censuses.

Step 4: Multiple Characteristics Method ("Statistical Matching")

The next method of assigning establishments to NAICS industries uses a statistical model that examines characteristics of each plant to help determine whether the plant "looks like" other plants in one particular NAICS industry or another. The assignment techniques described in the first three steps work well for some establishments and industries, but are less useful for classifying those plants that are small so that Census does not collect data on their products, or are in SIC industries that map to multiple NAICS industries, or do not survive from census to census. As we illustrate below, the "statistical assignment" group accounts for a relatively small fraction of establishments and employment.

Recalling that our assignment procedure starts with the most recent census and works backwards, the basic idea behind our statistical assignment method is to look at the characteristics of the plants in the same SIC industry in year t, and, based on the distribution in the subsequent (t+5) census, determine the probability that each plant gets assigned to each of its corresponding NAICS industries.

Specifically, we run a multinomial logit regression for each SIC industry associated with more than one NAICS industry using the data for all observations in the industry in the census of year t+5. We run the following model:

$$NAICS = \beta_1 + \beta_2(d\_te) + \beta_3(d\_tvste) + \beta_4(d\_swte) + \beta_5(inmsa) + \varepsilon_i$$

The variable *te* is the total employment at the plant, *tvste* is the total value of shipments per worker, *swte* is the average wage, and *inmsa* is a dummy for whether or not the establishment is in an

metropolitan statistical area.[5] The *d_* prefix implies that all variables are measured as deviations from the SIC industry medians. We use the vector of coefficients to construct establishment level probabilities for all plants in these SIC industries.

We then draw from the resulting distribution for each establishment and assign the establishment to a NAICS industry. We perform this exercise 1,001 times and construct industry level aggregates for a small set of variables typically used in Census Bureau publications: number of establishments, total employment, total value of shipments, production worker hours, and salaries and wages. We then construct deviations of these totals from the industry median and aggregate across all the 474 NAICS manufacturing industries. To determine the "best" draw, ***j***, we pick the one draw that satisfies the following:

$$\min_{j \in J} \phi_j = \lambda_1(mdestabs_j) + \lambda_2(mdte_j) + \lambda_3(mdtvs_j) + \lambda_4(mdph_j) + \lambda_5(mdsw_j)$$

where $\sum_{k=1}^{5} \lambda_k = 1$ and $J = \{1,..,1001\}$.[6]

Step 5: Special Cases

Even after completing the exact and statistical matching, a few establishments escape classification. Some of these "special cases" are those in sparsely populated industries where there are too few observations to run a multinomial logit regression. In these cases, we use a shipments weighted probability that is constant for all establishments in the industry. Other special cases include the small set of establishments with industry codes on the data files that were never updated when there were revisions to the SIC system in 1972 and 1987. In these instances, we used information from the concordances in the SIC and NAICS manuals to manually assign NAICS industry codes to plant records.

[5] These variables contain few missing values in the censuses. Some of the multinomial regressions contain very few observations, so we wanted to avoid having to exclude plants whenever possible.

[6] All variables are measured as deviations from industry medians; "mdestabs" refers to the number of establishments, "mdte" refers to total employment, "mdtvs" refers to total value of shipments, "mdph" refers to production worker hours, and "mdsw" refers to salaries and wages. In this example we set the weights ($\lambda_{1 \text{ to}} \lambda_5$) equal, but the method allows for differential weighting.

IV. Results

In this section, we present some statistics about the industry assignment process described in the previous section. Table 2 shows the share of establishments assigned NAICS industries by each of the five methods for census years 1963 to 1992. Roughly 80 percent of all establishments are assigned NAICS industries by our exact methods. Although the share of establishments assigned using product data falls significantly over time, the percentage of establishments assigned by the industry and establishment methods rises over the same period. For 1992, we assign 46.2 percent of all establishments by the product method, but by 1967 that number falls to just below 3 percent. One reason is that seven-digit product codes are specific to the Census Bureau and need not be time consistent in the same manner as the SIC system. Also, over time (retrospectively), establishment assignment grows in importance from 24.6 percent in 1992 to 50.4 percent in 1963. Similarly, the share of establishments assigned by the industry method grows from 11.1 percent in 1992 to 21.1 percent in 1963. It is important to note that although the share of matches based on product assignment falls over time, these matches are replaced by one of our other exact matching techniques. The proportion of establishments assigned NAICS industries by statistical matching techniques remains relatively constant over time at roughly 20 percent. The share of special cases is negligible.

Table 2. Shares of Establishments by Method of Assignment

Type of assignment	1992	1987	1982	1977	1972	1967	1963
All "Exact" Matching	.820	.810	.794	.776	.809	.799	.813
product method	.462	.262	.164	.108	.111	.029	.098
estab method	.247	.379	.404	.388	.461	.519	.504
Industry method	.111	.169	.226	.280	.237	.251	.211
statistical assignment	.179	.190	.205	.219	.188	.201	.183
Other	.000	.000	.001	.005	.003	.001	.005

Table 3 shows the share of total manufacturing shipments that are assigned NAICS

industries based on each method.[7] This table shows that nearly all shipments are classified by one of the exact matching techniques. Exact methods account for more than 90 percent of all manufacturing shipments in every census year back to 1963. So although roughly 20 percent of *establishments* are assigned to NAICS industries by statistical methods, these establishments account for a much smaller share of overall shipments.

One concern is that although there are relatively few establishments that are statistically assigned, these establishments may represent a proportionately larger share of overall shipments. However, the data do not support this concern. For example, in 1992, 17.9 percent of all manufacturing establishments were assigned by the statistical method, but this group accounts for only 3.4 percent of total shipments in the sector. Although the share of establishments assigned by statistical techniques grows by just a few percentage points as we look back over earlier censuses, the share of shipments doubles, but is never more than 7 percent. The shares of shipments assigned by the product method (and establishment method) show even more dramatic patterns of decline (and increase) as compared to the shares of establishments. Also, the industry matching method, while relatively important in terms of establishments, is much less important from a shipments weighted perspective.

Table 3. Share of Shipments by Method of Assignment

Type of assignment	1992	1987	1982	1977	1972	1967	1963
All "exact" matching	.966	.965	.941	.938	.939	.925	.933
product data	.798	.611	.363	.305	.271	.029	.109
Estab. links	.149	.294	.441	.539	.563	.769	.718
Ind. data	.019	.060	.137	.094	.105	.127	.106
Stat. assignment	.034	.035	.058	.058	.056	.073	.064
other	.000	.000	.001	.005	.006	.002	.003

V. Evaluating data quality

Because there is no existing historical SIC-NAICS bridge table to check our data against, it is difficult to evaluate both the matching process and the quality of the data. One possibility, however, is to match all establishments by all possible methods to see how often each method yields the same result. Obviously, we cannot match all establishments by all methods (or we would have just assigned establishments to NAICS industries using the most reliable method). However, we can often match plants assigned to NAICS industries by one of the "exact" techniques by statistical assignment. To the extent that, for example, statistical assignment produces the same industry assignment as one of the exact techniques, we will gain confidence in the methodology.

Table 4 shows the percentage of matches, where calculable, that are assigned to the same industry by different techniques for 1992. The value in the upper left-hand corner of the table shows that more than 96 percent of the time that we assigned a plant by the (most reliable) product method, the industry method would have yielded the same result. The establishment method appears even more reliable – the "similarity rate" is more than 99 percent for establishments matched by both the product method and the industry method. As expected, the statistical matching method is less likely than other methods to yield a match identical to the industry identified by one of the "exact" methods. Although this "similarity rate" is lower for the statistical matching cases, it is still fairly high, which gives us reasonable confidence in the matches made by this method.[8]

Table 4: Percentage Matched to Same industry by different techniques

	Product method	Industry method	Estab. method
Industry method	96.6	X	X
Establishment method	99.1	99.6	X
Statistical method	70.8	N/A	74.8

Another way to evaluate the assignment methodology is to look at new industries, or industries we are familiar with, to see if they behave in predictable ways. For example, the SIC industry shown in Table 1, calculating & accounting machines, except electronic computers, corresponds to two different NAICS industries -- office machinery manufacturing and other computer peripheral equipment manufacturing. As noted above, we might expect the share of computer peripherals to be higher in recent years and much lower in earlier censuses. We compare the shares generated from the matching procedure, which we call "variable shares" because they are allowed to

[7] Constructing this table on an employment-weighted basis does not qualitatively change the results.

[8] We are currently exploring comparisons of these similarity rates with those constructed by random and shipments-weighted assignment.

change over time, with the shares implied by the 1997 Census data (the only publicly available data on shares), which we call "fixed shares." This comparison gives us a glimpse into whether the constructed data behave "reasonably."

Figure 1: Comparing Fixed and Variable Shares

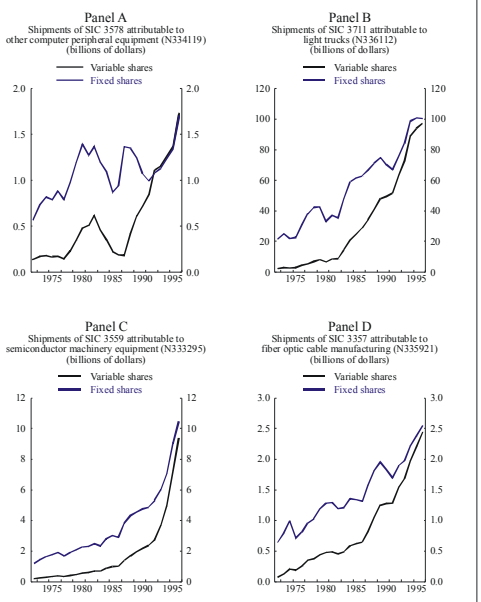

Figure 1 shows the comparison of variable and fixed shares in four industries. The top left panel, panel A, shows the total value of shipments of SIC 3578 attributable to the NAICS industry for computer peripherals calculated using variable shares and fixed shares. As expected, if we only had access to the 1997 public data and just applied those shares back in time, we would have likely overstated shipments by a substantial amount. The use of variable shares seems to provide a much more reasonable historical picture.

The other three panels in figure 1 present a similar story. Panel B, in the upper right corner, shows total shipments of motor vehicles (SIC 3711) attributable to light trucks under both the fixed and variable shares calculations. Panel C, in the lower left corner, shows for special industry machinery (SIC 3559), the shipments associated with semiconductor machinery equipment; and panel D, in the lower right corner, shows how much of the shipments of drawing and insulating of nonferrous wire (SIC 3357) should be attributed to fiber optic cable manufacturing under the two share methods. Each of these cases presents a similar scenario: if we had just taken the 1997 share and applied it backwards in time, we would likely have overestimated shipments for these new or growing industries.[9]

Panel D illustrates how the variable shares pick up more accurately real activity in an industry. Although first listed as an industry in the 1997 NAICS manual, fiber optic cable appeared as a product in the 1987 SIC manual, and there were no references in the 1972 manual. This coincides with the initial development of fiber optic cable as we know it in the early 1970s and more widespread production by the early 1980s (Hecht, 1999). The variable shares method seems to reflect this history more accurately than the fixed shares method.

Although these cases do not present a formal test or evaluation of our methodology, the consistency of these cases is reassuring to us that the data appear reasonable and provide a sense that our methodology generates the type of results one might expect.

V. Conclusion

This paper highlights the need for a historical bridge between the SIC and NAICS systems and briefly describes the method we use to generate a bridge for the manufacturing sector. We assigned each establishment in all COMs from 1963 to 1992 to a NAICS industry and examined the distribution of the different assignment methods. Although it is difficult to evaluate the quality of these assignments, we present two ways of looking at these matches. Once we were confident in the resulting data, we used the new historical concordances as an input into the reclassification to NAICS of the Federal Reserve's Index of Industrial Production. Our hope is that this work can eventually be more widely employed by the data using community.

[9] Graphs of employment shares present a similar picture.

References

Hecht, Jeff (1999). *City of Light: The Story of Fiber Optics.* Oxford University Press.

Executive Office of the President, Office of Management and Budget. "North American Industry Classification System, United States 1997". Lanham, MD: Bernan Press, 1998.

Executive Office of the President, Office of Management and Budget. "Standard Industrial Classification System, 1987".

Executive Office of the President, Office of Management and Budget. "Standard Industrial Classification System, 1972".