METHODS USED TO DEVELOP RETAIL AND WHOLESALE TIME SERIES UNDER THE NORTH AMERICAN INDUSTRY CLASSIFICATION SYSTEM[1]

Michael Z. Shimberg, Ruth E. Detlefsen, William C. Davie, Jr., U.S. Bureau of the Census
Michael Z. Shimberg, U.S. Census Bureau, SSSD Rm. 2651-3, Washington, DC 20233-6500

KEY WORDS: Time series, North American Industry Classification, Benchmarking

1. Introduction

The U.S. Census Bureau has published estimates for the retail and wholesale sectors of the economy for many years. These estimates are produced from economic censuses conducted once every five years, from annual and monthly sample surveys, and from the tabulation of administrative data. Until recently, these estimates were based on the Standard Industrial Classification (SIC) system. The estimates are now based on the 1997 North American Industry Classification System (NAICS).

The U.S. Census Bureau was motivated to begin releasing estimates on a NAICS basis after the Federal governments of the U.S., Canada, and Mexico agreed to adopt the new classification system. In addition, the Census Bureau decided to restate its historical SIC estimates to a NAICS basis because of the importance of these estimates in the analysis of economic trends.

The Census Bureau's first NAICS estimates for the retail and wholesale sectors were from the 1997 Economic Census, the 1999 annual surveys, and the April 2001 monthly surveys. The initial releases of monthly retail and wholesale estimates also included NAICS time series for January 1992 through March 2001.

We open the paper with a brief description of the economic censuses and the retail and wholesale sample surveys. Most of the this paper is devoted to discussing the methods used to develop the historical NAICS time series. Throughout the discussion, we address changes to the original plans (Detlefsen 2000) for developing the series. We also discuss the development of a series for retail electronic-commerce (e-commerce) sales estimates. Near the end of the paper we describe seasonal, trading-day, and holiday adjustments made to the NAICS time series. We close with a brief summary of issues that arose after the initial release of the monthly NAICS estimates.

2. Program Descriptions

2.1 Economic Census

The U.S. Census Bureau conducts an economic census once every five years. Data from the 1997 census, the most recent, were tabulated on both a NAICS and an SIC basis with the primary focus on NAICS.

The census tabulates data obtained for establishments. Establishments are defined as business units at single geographic locations that produce or distribute goods or services.

Data for establishments with no paid employees (nonemployers) and for most of the smallest single-establishment companies with paid employees (employers), were obtained from administrative records. Data for all other establishments, including for a sample of the smallest single-establishment companies, were collected from census questionnaires.

Establishment data for the reference year were aggregated to provide estimates of sales, payroll, employment, and other data items for detailed industries, geographic levels, and size categories. Some data items were not available from administrative records. Hence, those items were estimated by multiplying the data collected for the sample of small single-establishment companies by their sample weights.

2.2 Annual and Monthly Surveys

The annual and monthly retail and merchant wholesale surveys use probability-based samples. The sampling units consist of establishments for single location firms and various aggregations of establishments for multi-location firms. The monthly survey samples are subsets of the annual survey samples. The samples are updated periodically to account for business births, deaths, and other organizational changes.

New NAICS-based samples were selected for the retail and wholesale sectors after data collection and most data

---

review were completed for the 1997 census. The sample surveys do not include all wholesalers as did the census. With one exception, the surveys include only those wholesalers taking title to goods bought for resale to other businesses (merchants). The exception is the inclusion of sales branches of ferrous metal service centers. Throughout the rest of this paper, "wholesale" refers to the sample survey coverage.

For these surveys, questionnaires are mailed to selected sampling units that were active during the reference period covered. Weighted data are summed to produce estimates for employer businesses. The annual retail survey employer estimates are supplemented by tabulations of administrative data for nonemployers. For the monthly retail estimates, nonemployers are accounted for by benchmarking to the annual survey results.

These surveys provide national estimates of level and change for various industries. Estimates of annual and monthly sales, annual purchases, end-of-year and end-of-month inventories, and annual e-commerce sales are computed for both the retail and wholesale industries. For retail, we also compute estimates of end-of-year accounts receivable and quarterly e-commerce sales.

3. Methods for Developing NAICS Time Series

Time series for the retail and wholesale sectors were developed for monthly sales, end-of-month inventories, annual sales, end-of-year inventories, and annual purchases. A series was also developed for quarterly retail e-commerce sales estimates. This paper excludes a discussion of historical estimates developed for any Accommodations and Food Services industries included in the monthly and annual retail survey publications.

Except for the e-commerce sales estimates (see Section 3.3), the time series were derived by first restating previously published SIC-based estimates on a NAICS basis. Then a sequence of benchmarking procedures was applied to the restated estimates. Benchmarking established NAICS level estimates that preserved, as much as possible, the year-to-year and month-to-month change estimates computed from the restated census and survey data. In addition, the annual estimates for 1992 and 1997 were constrained by census results on a NAICS basis.

3.1 Restatement Methods

To begin developing NAICS time series, we first restated the 1992 Economic Census sales totals and the SIC-based sample estimates on a NAICS basis. We also derived

1992 and 1997 purchases and end-of-year inventory estimates on a NAICS basis.

3.1.1 1992 Economic Census Sales

The first step in obtaining 1992 census sales for employers on a NAICS basis, was to assign NAICS codes to each employer establishment having an SIC that, by definition, was the same as a particular NAICS code. About 74% of the establishments were assigned NAICS codes in this way.

Employer establishments not assigned a NAICS code in this manner were matched by their identification number and 5-digit SIC to the 1997 census, which contained both SIC and NAICS codes. Matched establishments were assigned the 1997 census NAICS code. About 6% of the establishments were assigned NAICS codes in this way.

Remaining uncoded employer establishments that belonged to multi-establishment firms were assigned NAICS codes based on collective characteristics of the establishments within their firm, if the establishments of the firm were fairly homogeneous. Less than 0.1% of the establishments were assigned NAICS codes in this way.

All other employer establishments were assigned NAICS codes randomly. Probabilities for randomly assigning NAICS codes within the SICs were determined using a multinomial logistic regression model. Data available from the 1997 census were used in the computations. Details are available in Klimek 2000. About 20% of the establishments were assigned NAICS codes in this way.

The methods just described were not used to assign NAICS codes to establishments in the Warehouse Clubs and Superstores industry and in the discount part of the Department Stores industry. The composition and nature of the establishments now classified in these industries changed in dramatic and escalating ways between the 1992 and 1997 censuses. These changes made it necessary to conduct research on individual firms to determine the correct classification of their establishments in 1992.

After industry codes were assigned, establishment data were summed to give employer estimates for each NAICS industry.

Though not detailed in our initial plans, we also had to obtain 1992 census sales for retail nonemployers on a NAICS basis. To do this, we first distributed 1992 and 1997 nonemployer estimates for each SIC code to its corresponding NAICS code(s). Distributions were based on those observed for 1997 census employer sales. If a

NAICS code consisted of more than one SIC component, the distributed estimates from each SIC component were summed to the NAICS level.

To complete the restatement of the 1992 retail nonemployer estimates, the 1992 nonemployer estimates obtained via distribution were multiplied by a ratio. This was the ratio of 1997 nonemployer estimates tabulated on a NAICS basis to 1997 nonemployer estimates obtained by distributing SIC estimates to NAICS. The ratio adjusts the nonemployer estimates for errors introduced by the distribution process.

Restated 1992 census sales for retail were derived by summing the restated employer and nonemployer sales. These were used as constraints in the benchmarking process.

3.1.2 Monthly and Annual Survey Estimates

We restated the SIC-based monthly employer estimates for January 1992 through March 2001. We also restated the annual employer estimates and the annual total (employers plus nonemployers) retail estimates, as benchmarked to SIC-based census results, for 1992 through 1998. Estimates were restated for monthly sales, end-of-month inventories, annual sales, end-of-year inventories, and annual purchases.

To restate these estimates, we first distributed estimates for each SIC code to its corresponding NAICS code(s). Distributions for all data items were based on those observed in the 1997 census for sales. Monthly and annual estimates were distributed only for employers. Annual retail estimates for 1992 through 1997 were also distributed for the total of employers and nonemployers. When a NAICS code consisted of more than one SIC component, the distributed estimates from each SIC component were summed to give the restated NAICS estimate. Restated 1998 annual estimates for total retail were derived by summing the restated employer estimates and administrative nonemployer data.

Exceptions were made in distributing estimates from SIC to NAICS for the Office Supplies and Stationary Stores industry, for the Warehouse Clubs and Superstores industry, and for the discount part of the Department Stores industry.

Large firms within the Office Supplies and Stationary Stores industry that would have been classified as wholesale by the SIC system were found to have a different seasonal pattern than the rest of the firms within that industry. To prevent these large cases from having

undue influence on the monthly and annual estimates for that NAICS industry series, we developed separate distributions for these large firms and for the balance of that NAICS industry. Distributions were based on those observed in the 1997 census for sales. Estimates obtained via distribution for these two parts were summed to give estimates for the industry.

For firms having establishments in both the Warehouse Clubs and Superstores industry and in the discount part of the Department Stores industry, research was conducted on individual firms to determine how to distribute their annual sales. Rreasons for this are noted in Section 3.1.1.

The methods described to this point gave us the restated annual estimates for 1992 through 1998. One final step was needed to complete the restatement of monthly estimates. We multiplied the monthly employer NAICS estimates prior to March 2001 (those derived through distribution) by a geometric mean. The geometric mean adjusted the monthly NAICS estimates derived from our SIC-based monthly surveys to be consistent with estimates from the NAICS-based monthly surveys. Note that January, February, and March 2001 data were collected from both the SIC-based and the NAICS-based surveys.

Geometric means were computed separately for sales and inventory and were calculated as:

$$\left(\frac{X_t}{Y_t} \bullet \frac{X_{t-1}}{Y_{t-1}} \right)^{1/2} \text{, where}$$

X_t and Y_t represent the March 2001 NAICS estimates for employers. X_t was from the newly introduced NAICS-based sample surveys. Y_t was derived by distributing estimates from SIC to NAICS.

These restated estimates were used to compute annual and monthly change estimates for input to the benchmarking procedures.

The restatement method described above differs from our initial plans. Under the earlier plans, if 90% or more of the census sales for any NAICS industry was comprised of 90% or more of the census sales for an SIC industry, we would have used the change estimate for that SIC. For all other industries, we had planned to assign NAICS codes to the units in our samples, retabulate the estimates, and compute change estimates from the results. We abandoned the plan after observing that this added a great deal of nonsampling error to our estimates.

3.1.3 Annual Purchases and Inventory Estimates for Census Years

Before benchmarking could begin, we had to derive 1992 and 1997 purchases and inventory constraints. To derive these, for each of 1992 and 1997, we separately multiplied the NAICS census sales totals by ratios of the end-of-year inventory-to-sales and purchases-to-sales computed from the restated annual estimates for the corresponding years.

Earlier plans did not address the purchases and end-of-year inventory estimates. However, deriving these constraints was necessary because retail inventories and purchases were not collected in the census. While wholesale inventories were collected in the census, they did not include inventories from auxiliary and central administrative offices, as did those collected for the samples.

3.2 Benchmarking Procedures

Benchmarking procedures were applied twice to obtain NAICS level estimates. The first application gave us the annual estimates and the second the monthly estimates.

Each application of the benchmarking procedure minimized the function F:

$$F = \sum_{t=2}^{N} \left(\frac{O_t}{O_{t-1}} - \frac{I_t}{I_{t-1}} \right)^2 \quad , \text{where}$$

t= time period 1, 2,...,N.

O_t = NAICS level estimates for time period t. (Output series.)

I_t = level estimates restated from an SIC basis to a NAICS basis for time period t. (Input series.)

3.2.1 Annual NAICS Estimates

Inputs (i.e., I_t/I_{t-1}) to the first application of the benchmarking procedure were the 1993 through 1997 annual change estimates for sales, purchases, and end-of-year inventory derived from the restated NAICS level estimates. F was minimized subject to constraining O_1 and O_6 to be the 1992 and 1997 NAICS-based level estimates of sales, purchases, or end-of-year inventory derived from the 1992 and 1997 censuses (see Sections 3.1.1 and 3.1.3), respectively.

We collected data and published estimates for 1998 and 1999 on a NAICS basis from the 1999 annual surveys. Because a NAICS-based sample survey was not conducted for 1997, it was necessary to find an alternative way to link these estimates to the NAICS-based 1997 census results. At one time we had considered deriving a sample-based estimate by weighting 1997 census data for units in our NAICS-based samples, forming a ratio of the census to the sample-based estimate, and multiplying both the 1998 and 1999 NAICS-based estimates by the ratio.

Due to implementation complexities, we decided to take another approach to link the 1998 annual estimates to the 1997 NAICS census results. We multiplied NAICS-based employer estimates of sales, end-of-year inventories, and purchases derived from the 1997 census (see Sections 3.1.1 and 3.1.3) by ratios of the 1998 to 1997 NAICS-based estimates for employers derived from the SIC-based annual surveys (see Section 3.1.2). Estimates for retail nonemployers were derived from administrative records. The employer and nonemployer estimates for retail were summed to give totals. The procedures described to this point gave us the published annual sales estimates for 1992 through 1998.

Although not originally planned, one final adjustment was needed to derive the published end-of-year inventories and annual purchases estimates for the period. The adjustment was made to account for differences in the relationships of sales to inventories and sales to purchases under the SIC system and NAICS. The adjustment assumes that historical NAICS relationships are more like those observed from the NAICS-based samples than from the SIC-based samples.

To make this final adjustment, the published 1998 end-of-year inventory estimate was set to the published 1998 annual sales times the 1998 inventory-to-sales ratio estimated directly from the 1999 annual survey sample. To derive the published estimates for 1992 through 1997, each of the end-of-year inventory estimates from the benchmarking procedure was multiplied by the ratio of the published-to-linked estimates of 1998 end-of-year inventory. Published purchases estimates were derived similarly.

The 1999 employer estimates were derived by multiplying the 1998 published NAICS-based employer estimates by the ratio of the 1999 to 1998 estimates computed from the NAICS-based sample for employers. Estimates for retail nonemployers were derived from administrative data. Retail employer and nonemployer estimates were summed to give the published retail estimates.

3.2.2 Monthly NAICS Estimates

Inputs (i.e., I_t/I_{t-1}) to the second application of benchmarking were the monthly NAICS change estimates for sales and inventory for employer businesses for February 1992 through March 2001 as derived from restated NAICS level estimates. The published wholesale and retail NAICS estimates of annual sales and end-of-year inventory were used as constraining benchmarks for producing the NAICS level estimates of monthly sales and end-of-month inventory.

For sales, F was minimized subject to constraining the sum of the NAICS monthly sales estimates for employers for any particular year to equal the NAICS annual estimates, from the benchmarking described in Section 3.2.1, for that year.

For end-of-month inventory, F was minimized by constraining the December end-of-month inventory for employers for a particular year to equal the end-of-year inventory, from the benchmarking described in Section 3.2.1, for that year.

This resulted in the NAICS published estimates for January 1992 through March 2001.

3.3 Methods for Quarterly Retail E-commerce Sales Estimates

The Census Bureau began publishing quarterly retail e-commerce sales estimates on an SIC basis for the 4th quarter of 1999. The estimates that we published were are derived from e-commerce sales data collected on a monthly basis.

We collected monthly data on both an SIC and a NAICS basis for January through March 2001. Quarterly retail e-commerce estimates were first released on a NAICS basis for the 2nd quarter 2001. The release for the 2nd quarter included a NAICS historical time series of the quarterly estimates beginning with the 4th quarter of 1999.

To develop historical NAICS estimates for quarterly retail e-commerce sales, we first restated estimates for broad industry groupings for the 4th quarter 1999 through the 1st quarter 2001. To determine the broad industry groupings for which to develop historical estimates, consideration was given to the size of the contribution to total e-commerce sales made by particular NAICS industries. Industries with relatively large contributions were estimated separately, while those with relatively small contributions were collapsed into logical and convenient broad industry groupings.

For each quarter, we set the restated e-commerce sales estimates for broad NAICS industry groupings to be the e-commerce sales estimates from corresponding aggregated SIC levels. Historical wholesale e-commerce sales were obtained and added to the SIC estimates for NAICS groupings significantly affected by businesses classified as wholesale by the SIC.

After obtaining restated estimates for each quarter for broad industry groupings, we developed final NAICS estimates for the 1st quarter 2001. For each NAICS industry for both January and February 2001, we multiplied the monthly e-commerce sales estimates from the NAICS-based sample by the ratio of the restated monthly sales to the monthly sales from the NAICS-based sample. This adjusted the January and February NAICS e-commerce sales estimates to be consistent with the restated January and February sales estimates that were used in the benchmarking operations. (See the geometric mean discussion in Section 3.1.2.)

We summed the resulting January and February e-commerce sales estimates and the March 2001 e-commerce sales estimates obtained from the NAICS-based sample to give quarterly e-commerce sales estimates for each NAICS industry. We obtained published and restated 1st quarter sales estimates by summing corresponding monthly estimates. The e-commerce sales results were multiplied by the ratios of the published quarterly sales estimates to the restated quarterly sales estimates for each NAICS industry. This adjusted the e-commerce estimates for each industry to be consistent with the sales estimates resulting from the benchmarking operations, i.e. the published sales estimates.

These final NAICS estimates for each industry were aggregated to give final NAICS e-commerce sales estimates for the broad industry groupings and for total retail.

For each NAICS grouping, the restated estimates for the 4th quarter 1999 through the 4th quarter 2000 were multiplied by the ratio of the final NAICS estimate of 1st quarter 2001 e-commerce sales to the restated 1st quarter 2001 NAICS e-commerce estimate. The resulting quarterly NAICS estimates for the industry groupings were aggregated to quarterly retail totals.

Benchmarking methods were not used to derive a NAICS time series for e-commerce estimates because no benchmarks were available.

4. Seasonal and Other Adjustments

The NAICS historical series of monthly sales and inventory estimates for retail and wholesale were used as input to the X-12-ARIMA program. The program derived factors for adjusting data for seasonal variations for all the items, for adjusting data for trading-day differences for sales, and for adjusting data for holiday differences for retail sales.

We did not compute such factors for the quarterly retail e-commerce sales series since the series is not long enough to estimate the factors. We will begin to adjust the series after enough data becomes available to estimate the seasonal, trading-day, and holiday factors.

5. Subsequent Issues

Overall the historical series have performed well as indicated by the reasonable current to prior year ratios and the ability of seasonal, trading day, and holiday factors computed from the series to remove the effects of seasonality.

Since the release of the time series, some problems have arisen for two industries- the Warehouse Clubs and Superstores industry and one unpublished detail industry within the Building Materials and Garden Equipment and Supplies Dealers subsector.

Reporting errors for the NAICS-based sample caused problems in the Warehouse Clubs and Superstores industry series. Steps were taken to resolve and correct the errors.

Data collected on a NAICS basis for one unpublished detail industry within the Building Materials and Garden Equipment and Supplies Dealers subsector showed a different seasonal pattern than did the historical series developed for it. While the data for the unpublished industry did not affect published estimates which were not adjusted for seasonality, the data had a significant affect on seasonally adjusted estimates for the unpublished industry, its subsector, and total retail.

The problem caused by data from the unpublished industry was resolved by modifying the seasonal adjustment procedures for the unpublished industry. For that industry, more weight was given to estimates derived from the NAICS-based sample and less weight to NAICS estimates derived from the SIC-based samples. In addition, since the estimates in that industry are quite variable, we did not exclude outliers when computing seasonal factors for the industry.

REFERENCES

Detlefsen, R.E. and Shimberg, M. Z. (2000), "North American Industry Classification System (NAICS) Time Series of U.S. Retail and Wholesale Estimates," 2000 Proceeding of Survey Methods Section, Statistical Society of Canada.

Executive Office of the President (1998), *North American Industry Classification System, United States 1997*, Washington, D.C.: U.S. Office of Management and Budget.

Executive Office of the President (1987), *Standard Industrial Classification Manual, 1987*, Washington, D.C.: U.S. Office of Management and Budget.

Klimek, S.D. and Merell, D.R. (2000), "On Reclassifying Industries from the Standard Industrial Classification System to the North American Industry Classification System," Presented at the Second International Conference on Establishment Surveys, Buffalo, N.Y.

U.S. Census Bureau (1999), *Advance Report,* 1997 Economic Census, Washington, D.C.: U.S. Department of Commerce.

U.S. Census Bureau (2001), *Annual Benchmark Report for Retail Trade and Food Services: January 1992 to December 2000*, Washington, D.C.: U.S. Department of Commerce.

U.S. Census Bureau (2001), *Annual Benchmark Report for Wholesale Trade: January 1992 to March 2001*, Washington, D.C.: U.S. Department of Commerce.

U.S. Census Bureau (2001), *X-12-ARIMA Reference Manual, Final Version 0.2*, Washington, DC: U.S. Census Bureau.