

CONVERTING HISTORICAL DATA TO A NEW CLASSIFICATION SYSTEM: THE MONTHLY WHOLESALE AND RETAIL TRADE SURVEY

1. INTRODUCTION

The January 1994 signing of the North American Free Trade Agreement between Canada, the United States and Mexico created the need for an industry classification common to the three signatories. The statistics agencies of the three countries developed the North American Industry Classification System (NAICS) in 1997. NAICS is based on a production-oriented or supply-based conceptual framework in that establishments are grouped into industries according to similarity in the processes used to produce goods and services. Use of the new system provides greater comparability of industry statistics between the three countries. However, it also breaks the historical continuity of national series based on other classification systems. Since 1948, Statistics Canada had been using primarily the Standard Industrial Classification (SIC) and its revisions (1960, 1970 and 1980).

Since 1988, Monthly Wholesale and Retail Trade Survey (MWRTS) data have been produced using SIC 1980 definitions. Following the development of NAICS, the survey was redesigned, and its first NAICS-based estimates were released for the April 2004 reference month in June 2004. To maintain the historical continuity of the estimates, it was necessary to produce NAICS-based historical monthly estimates by converting the SIC-based historical data.

2. DEFINITION OF WHOLESALE AND RETAIL TRADE

Under both SIC and NAICS, the wholesale and retail trade sectors have the same principal function: the purchase of goods for resale. However, the two classification systems differ in how they distinguish between the two sectors. In SIC, the distinction is based on the class of customer. Retailer merchandise is intended for the public, for personal or household use. Wholesalers resell merchandise to retailers, manufacturers, dealers, public institutions, farmers, professionals and other wholesalers. In NAICS, the difference between wholesale and retail lies in the production process: that is, whether the merchandise is sold in a store or not. This concept change results in a sector change for some stores. For example, computer stores, sellers of building materials (including home centres) and office supply and stationery stores, which were wholesalers under SIC, are retailers under NAICS. Establishments whose principal activity is installation and repair, which belonged under retail trade in SIC, are now part of the service sector in NAICS. Table 1 provides an overview of the relationship between SIC and NAICS. It contains a cross-tabulation of wholesale and retail sales under SIC and NAICS.

Table 1 – Sales by SIC industrial sector under NAICS (1998-2001)

Sector by classification	NAICS:	NAICS:	NAICS:	
	Retail trade	Wholesale trade	Other	Total
SIC: Retail trade	96.3%	-	3.7%	100%
SIC: Wholesale trade	4.2%	94.3%	1.5%	100%

The bulk of the activity remains in the same sector. However, it is important to note that the MWRTS publishes its data at a more detailed level than sectors - it uses trade groups. The SIC-based survey has 11 wholesale trade groups and 18 retail trade groups, while the redesigned NAICS-based survey has 15 and 19 respectively. The changes between the old and new trade groups are many and varied. To analyze the 34 new NAICS-based series, it is necessary to estimate their historical values by backcasting SIC-based data.

3. METHODS CONSIDERED

Statistics Canada's Business Register (BR) is a database of the population of Canadian businesses. Their establishments are catalogued in the BR and assigned an industry code and a geographic code. The MWRTS was using a simple random sampling (without replacement) design, which had been in place since 1988. The sample frame was the BR. The industry stratification was based on SIC. Since 1998, establishments in the BR have been classified by both SIC and NAICS. This dual classification makes it easy to obtain NAICS-based estimates from SIC-based estimates. Sample units are assigned a NAICS classification based on the information available in the BR, and domain estimation is performed.

Since every unit in the population is classified under both systems, post-stratification could also have been used. This option was rejected because the NAICS classification of non-sample units is of poorer quality than the classification of sample units, particularly in 1998. Domain estimation was used to produce NAICS-based estimates for the period from 1998 to the date the redesigned survey was introduced. To meet users' need for historical series, backcasting began in January 1991 for retail trade and January 1993 for wholesale trade.

A "macro" approach was used in backcasting MWRTS estimates. Under this approach, sample units for the SIC-based estimates do not have to be individually reclassified. For each NAICS group, a weighted linear combination of the total for all SIC groups is used.

The total $X_j^{(t)(geo)}$ for NAICS trade group j for period t for geographic region geo is given by

$$X_j^{(t)(geo)} = \sum_i \alpha_{ij}^{(t)(geo)} * X_i^{(t)(geo)}$$

where $X_i^{(t)(geo)}$ represents the sales of trade group i on a SIC basis, for period t for geographic region geo and $\alpha_{ij}^{(t)(geo)}$ represents the conversion coefficient between trade group i under SIC and trade group j under NAICS for period t for geographic region geo . They are estimated using data for which the dual classification is known.

The main advantage of this approach is that no additional processing is required for aggregate-level adjustments in the SIC series (smoothing, for example). Its principal disadvantage is that the process of estimating the conversion coefficients introduces errors.

4. ESTIMATING THE CONVERSION COEFFICIENTS

Between January 1998 and March 2004, all establishments covered by the MWRTS were coded under both SIC and NAICS. Thanks to this dual classification, we can produce NAICS estimates from SIC estimates and compute conversion factors when dual classification is impossible.

A conversion coefficient represents the percentage of the total for a SIC trade group that was assigned to a NAICS trade group. Coefficients that were judged invalid by wholesale or retail experts or had an absolute value of less than 0.3% were eliminated and reallocated. The series of non-zero coefficients were plotted on graphs so that any regional differences, seasonal patterns or outliers could be detected. We found very small differences at the provincial/territorial level and at the seasonal level.

As a result, we decided to use *average* monthly conversion coefficients for the 1991-1997 period. These conversion factors were estimated using the average of the coefficients computed for the corresponding months in the 1998-2001 period for each province and territory. No backcasting was done for the Montréal, Toronto and Vancouver census metropolitan areas (CMAs), for which retail data are available from January 2004 on.

5. QUALITY ASSURANCE

The effectiveness of the “macro” approach used to backcast MWRTS estimates is affected by various factors. One factor is the sample frame itself. An error in the SIC or NAICS classification for a given month between 1998 and 2001 obviously affects the month in question and all corresponding months in the 1991-1997 period. To reduce the impact of misclassified units, major contributors were checked manually and recoded if necessary. Corrections made in the frame since 1998 were catalogued so that their impact could be assessed; where necessary, the NAICS series were adjusted.

Another factor is the use of conversion coefficients based on recent years (1998-2001) to estimate conversion coefficients for previous years. This method is effective if the NAICS-based distribution is stable from year to year; if not, we believe that the risk of error is lower in 1997 than in 1991. The stability hypothesis was confirmed in the

majority of cases. However, conversion coefficients based solely on averages must not be used when an industry has undergone a significant structural change. One example is the home centres industry (NAICS 444110). Since the renovation sector has grown substantially in the last few years and big-box stores are a recent development in the Canadian economy, their contribution to the SIC trade group in 1998 is not representative of their contribution in the early 1990s. Consequently, the associated conversion coefficients must be adjusted downward for 1991 to 1997. The size of the adjustment is based on expert analysis and the results of a partial NAICS classification at the micro level. This type of adjustment makes it possible to model the coefficients' variations over time. Other factors related to seasonality, such as working days and the Easter effect, were not considered because few observations were available at the time of analysis.

6. WHOLESALE TRADE - INVENTORIES

The MWRTS has two variables of interest: sales and inventories. Statistics for the latter are released only for wholesale trade by trade group for Canada. The conversion coefficient analysis was performed entirely at the sales level. Instead of converting the SIC inventories series to NAICS, we converted the backcasted sales series to inventories using ratios. Inventories-to-sales ratios were computed for each month since 1993 at the SIC trade group level and applied to the corresponding NAICS trade groups. Correspondence was based on the following rule: a NAICS trade group corresponds to a SIC trade group if the latter contributes over 99% of the former's total. Seven NAICS wholesale trade groups have a corresponding SIC group. The NAICS groups that have no corresponding SIC group were processed using a general inventories-to-sales ratio.

7. CONTINUITY OF THE NAICS SERIES

The NAICS series can be divided into three parts. The first part runs from January 1991 to December 1997. The estimates for this part are based on *estimated* conversion coefficients. The second part covers the period from January 1998 to March 2004. In this part, the NAICS series are computed by domain estimation or using *observed* conversion coefficients. The third part begins with the introduction of the redesigned survey. The second and third parts overlapped by a few months, during which time the old and new surveys were both in production.

The transition from the SIC-based survey to the NAICS-based survey caused a break in the series. This discontinuity is due to both the classification change and other methodological changes. The results of the overlap period were used to adjust the level of the backcasted series. A time-constant multiplicative adjustment was used to adjust the historical series to the published levels from the redesigned survey.

This benchmarking process strips the system of series of its additivity because the series are benchmarked individually. For example, the process results in discrepancies between the trade group total and the provincial/territorial total. To restore additivity, the benchmarked series are subjected to a reconciliation process.

The reconciliation process is such that (a) the system's components are altered as little as possible in percentage terms, and (b) the additivity of the system of series is maintained for each month.

Hence, every series is reconciled in such a way that, for each month, (1) the totals for the trade groups by province and territory equal the provincial/territorial totals; (2) the totals for the provinces and territories by trade group equal the trade group totals; and (3) the sum of the provincial/territorial totals equals the sum of the trade group totals.

When seasonality was considered, breaks in the NAICS series could also be seen in January 1998, when the transition was made from estimated coefficients to observed coefficients. To mitigate the effect, all backcasted data for 1998 were recalculated using estimated coefficients; that is, the average values for the four years including 1998 were used. Note that the 1998 coefficients deviate from the average more often than the coefficients for the other three years, and that highly aberrant coefficients were not included in the average. By extending the first part of the series to December 1998, we eliminate any breaks between the first two parts.