Doc. Class/05/41 1 July 2005

Methodological problems and options for SIC-NAICS conversion

(Statistics Canada, 2001)

EN

NACE/CPA Implementation Task Force meeting 4-6 July 2005
Agenda item 6

STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES

02 Statistical Governance, quality and evaluation
Tel. (352) 4301-32023, Fax (352) 4301-33899

Methodological Problems and Options for SIC-NAICS Conversion[1]

Mike Hidiroglou, Benoît Quenneville and Guy Huot, Statistics Canada,

11-RHC-BSMD, Ottawa, Ontario, K1A 0T6, 613-951-4767

Revised: Friday, October 5, 2001

1. Introduction

An industrial classification system must be periodically updated to reflect underlying changes in the economy. This creates the tasks of reclassifying all units on a frame, and then to create historical series with this new classification from the existing series with the previous classification.

There are two goals in the initial stages of implementing a revised classification. These are: (1) to update the classification code at a single point in time; and (2) to produce economic and statistical data under both old and new classification systems for at least one point in time. This overlap is crucial, for it allows the production of quantitative measures of concordance between the two classification systems, and enables the creation of *conversion tables* and *concordance coefficients* for assuring historical continuity.

2. Reclassification

Reclassification is the process by which a unit is reassigned a new industrial code. The assignment of these new codes can either be done automatically or manually.

An *automatic assignment* occurs when there is a one-to-one correspondence between the old and new classification, or if there is enough quantitative information to assign a new code with a high probability of obtaining the correct classification.

A *manual assignment* occurs when knowledgeable staff assigns a new classification code to a unit.

The relationships between the old and new classification codes can be *one-to-one*, *one-to-many*, *many-to-one*, or even *many-to-many*. The relationships one-to-one and many-to-one allow the automatic assignment of the new codes. However, when old codes have been split into two or more new ones (the one-to-many and many-to-many cases), the new codes have to be manually assigned.

[1] This note is inspired by McDonald (1995) and Morry (2000). Thanks are due to Pierre Cholette for his comments on previous versions of this note.

3. Conversion Tables and Concordance Coefficients

The first step in establishing concordance is to create *conversion tables* relating the old and new classification systems. These tables aid users in making the transition to the new system, in understanding the relationship between the old and new codes, in discerning the industrial scope of changes, and in understanding how the revision affects the historical continuity of estimates.

Major conversion tables provide a comparison of codes in the old and new systems. For example, a conversion table could have on the vertical axis the old codes, with newly created codes indented to the right. On the horizontal axis, the column descriptions indicate what has happened to the old codes. Examples of expanded column descriptions are:

- same code and industry scope;
- different code but same industry scope;
- entire old industry now part of a new industry;
- industry split into two or more industries;
- expanded industry scope;
- reduced industry scope;
- expanded and reduced industry scope;
- new code;
- deleted old code;
- old code reused but with different industry scope.

Subsidiary conversion tables show the movements of economic activities and/or the changes in industrial scope. They include (1) lists of all industries within each category, and (2) changes in industrial scope (additions and subtractions) from old to new system and from new to old. Relationships from old to new and from new to old are exhibited, and crosswalks between the two systems are provided. These tables provide the basis for comparing the two classification systems and for converting data between them. By summarizing the revision, the tables assist in validating code changes in the split industries and in converting codes mechanically for other industries.

Concordance coefficients are conversion factors based on a measured reallocation of data at aggregate industry levels that reflect the changes between the old and new classification systems. Given the subsidiary conversion tables, the concordance coefficients can be calculated for each classification based on the number of units. They can also be calculated based on variables such as employment, earnings, sales, etc. They show how much each industry has changed (either in terms of units or in terms of a variable), where the movements took place, and between which industries the movements occurred and in which direction. These concordance coefficients can be computed at a single time point or at several time points. The advantage of measuring them at several time points is that we can determine whether the concordance coefficients at a single point in time are appropriate. Ideally, concordance coefficients are calculated at every time points and for all the variables of interest.

4. Backcasting Time Series on the New Classification Basis

The most significant impact of a classification update is its effect on the continuity and use of time series.

The time span for the new series can be broken up into three time segments:

(1) The historical time segment where only the old classification existed.
(2) The transitory time segment where the old and new classifications are present.
(3) The final time segment where only the new classification will be used.

There are two problems connected to the construction of a new time series spanning the three time segments. The first one is to construct estimates using the new classification for the time points within the historical segment, and, if not already available, in the transitory segment as well. The second problem is to link the estimates of those three segments.

4.1 Construction of Estimates Using the New Classification for the Time Points within the Historical and Transitory Segments

The construction of estimates using the new classification for the time points within the historical and transitory segments can either be done at a *micro* or a *macro* level.

At a micro level, domain estimation can be carried out using the re-coded records. This first requires assigning new codes to all the sampled units in the historical span. Next the production needs to be re-run with domain estimation for all the time points in the historical segment. (Domain estimation involves computing the estimates under the new classification with the survey weights from the old classification.) This can be costly and result in lower quality data (higher coefficients of variation). Furthermore, domain estimation may result in undercoverage of the estimates because previously out of scope units (now in scope) may not have been surveyed.

At the macro level, estimates are generated with a weighted linear combination of estimates from the old classification. The weights of this linear combination use concordance coefficients. As already mentioned, one would require concordance coefficients for every time point within the historical and transitory segments, and for every variable. It is quite often the case that only a single set of concordance coefficients is available. In this case, application of conversion coefficients works well for short periods of time; however, the assumptions underlying the coefficients may become invalid over longer periods where the new system's economic structure differs substantially from that of the old.

The following errors can be introduced with the macro level approach:

1. An error due to using concordance coefficients calculated at a given time point and applying them at different time points.
2. An error introduced due to using concordance coefficients that were calculated based on one variable (say employment) and applied to another variable (say earnings).

4.2 Linkage of the Estimates from the Three Time-Segments

Regardless of the method used to obtain estimates over the historical segment, a jump will typically occur between the historical and the transitory segment. However the jump should now be mainly due to the survey redesign and to the errors introduced by the historical conversion. The purpose of linking is to alleviate the jump. One approach is to raise the converted historical segment to the level of the transitory segment, which eliminates the jump; another is to "wedge" the jump, i.e. to spread it over a number of months or years. A variant of these two approaches consists of satisfying converted annual benchmarks, by hovering around the jump correction (Helfand, Monsour and Trager, 1977).

4.3 Multiple Series Consistency

Once the new table of continuous time series is produced, contemporaneous additivity must be restored. For example, the component series of an industry group may not add to the industry total at every time point. This reconciliation method used to is based on a generalization of "raking" procedure (Byron, 1978) applied to tables of time series.

The final stage is that subject-matter analysts review the series and adjust them to agree with their prior knowledge.

5. Statistics Canada Experiences[2]

Several working papers exist on the SIC-NAICS conversion within Statistics Canada. The procedure for MRCTS is described in Cholette (1998). The conversion of the SEPH historical series to NAICS is described in Laflotte et al. (1999) using a macro level adjustment procedure. Morry (2000) covers the recoding of the Business Register, an example of backcasting using micro records (Labour Force Survey) as well as examples of macro level backcasting (IOFD, ASM-MSM). The work to date on MWRTS and AWRTS is described in Meyer (2001).

[2] This is a partial list that does not cover all of Statistics Canada divisions.

References

Byron, R.P. (1978), "The Estimation of Large Social Account Matrices," Journal of the Royal Statistical Society, Series A, Vol 141, pp. 359-367.

Cholette, P. (1998), "Linking to NAICS, the Case of the MRCTS Series," Unpublished Statistics Canada paper, Time Series Research and Analysis Centre, October 14, 1998.

Helfand, S.D., Monsour, J.J. and Trager, M.L. (1977), "Historical Revision of Current Business Survey Estimates," Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp. 246-250.

Laflotte, R., Lavallée, S. and Lavallée, P., (1999), "Converting the SEPH Historical Series to NAICS," Unpublished Statistics Canada paper, February 22, 1999.

McDonald, B. (1995), “Implementing a Standard Industrial Classification (SIC) System Revision,” In Business Survey Methods, Edited by Cox et al., New York: John Wiley.

Meyer, B. (2001), "1980 SIC/NAICS Conversion Wholesale and Retail Trade. An Overview from a Survey Area Perspective," Unpublished Statistics Canada, Distributive Trades Division, Statistics Canada, September 28, 2001.

Morry, M. (2000), “Backcasting Time Series at Statistics Canada under NAICS,” Power Point presentation.