

# Economic Classification Policy Committee

## Issues Papers

### Issues Paper No. 5 The Impact of Classification Revisions on Time Series

July 1993

Note to reader: This is the fifth in the series of Economic Classification Policy Committee issues papers. The first two, Issues Paper No. 1, "Conceptual Issues," and Issues Paper No. 2, "Aggregation Structures and Hierarchies," were published in the Federal Register, March 31, 1993, pp. 16990-17004. Copies are available by writing to Brenda M. Erickson, Economic Classification Policy Committee, Bureau of Economic Analysis (BE-42), U.S. Department of Commerce, Washington, D.C. 20230, or by telephone at (202) 606-9620.



# Economic Classification Policy Committee Issues Paper No. 5

## The Impact of Classification Revisions on Time Series

### Introduction

A number of the participants in the 1991 Conference on the Classification of Economic Activity held in Williamsburg, Virginia ([1], hereafter, "Williamsburg Conference"), referred to the inevitable disruptions to time series when classification systems are revised. None chose to pursue the topic in depth. The Economic Classification Policy Committee's (ECPC) charge to carry out a new "fresh slate" evaluation of economic classifications creates potentially significant repercussions for time series. This issues paper discusses trade-offs in maintaining time series continuity versus keeping the economic classification system modern. The paper documents the impact of past SIC revisions on time series and outlines why the impact of revisions cannot be overlooked during the critical planning stage for new economic classifications.

The issue of comparability of time series focuses on certain types of changes, namely those that regroup establishments at the 4-digit Standard Industrial Classification (SIC) level. Changes in the hierarchical structure will not affect historical comparability since the entire historical series can be retabulated. Retabulations of series do entail costs, but those costs are usually modest.

To simplify the discussion, the paper assumes that the relevant classification unit is the establishment, as it has been in the past. Whether the classification unit itself should be changed is also an issue, which is discussed in ECPC Issues Paper No. 1, "Conceptual Issues" [2]. Any change in the classification unit would, of course, have implications for time series comparability.

### 5.1 Background

The general concept of a time series is quite simple. A time series is a set of observations of a given variable in sequential order. Observations are made at consistent intervals (months, quarters, and so forth). Such chronologically-ordered data serve as a historical description of a phenomenon that can be counted or measured.

Time series continuity is understood to mean not only that the observations of a series are continuous, but also that they are collected using standard methods and definitions over time. If the standards for defining and observing a variable are inconsistent, the observations will have little relationship to one another and will not constitute a true series.

### 5.2 Need for Historical Continuity

Many statistical agencies, private analysts, and government officials use time series data for economic research and analysis. In addition to being used to analyze and interpret economic events and conditions, another application of time series is forecasting, where the data are used in specially-designed models to make projections or predictions about economic phenomena. Time series figure importantly in analyzing trends in employment and wages, calculating the leading and coincident indicators, and estimating and analyzing the national income and product accounts. Accurate analyses and forecasts require consistent time series data. Accuracy is essential for informed economic decision making by Congress, the Administration, the Federal Reserve, and State and local governments.

Harvey Monk ([7], p. 12) noted that U.S. statistical agencies "have competing views" on the importance of historical continuity, depending upon the type of data they produce. These views depend on the types of data users they serve. However, no documentation exists on the relative costs of time series disruptions or outdated classification systems to the nongovernment economic user community. The impact of any classification revision on time series must be weighed carefully, particularly when a new classification structure is under consideration.

### 5.3 The Problem of Time Series Continuity

Changes in industrial classifications interrupt the continuity of associated time series. But economic classification systems cannot remain unchanged indefinitely if they are to capture the full scope of constantly evolving industrial and business activities

in our economy. As Peter Struijs (Williamsburg Conference [10], p. 14) stated, "...changes cannot be measured appropriately when the measuring instrument is changing constantly, but not changing it reduces the significance of information on the industrial structure." Real-world practice compromises by minimizing the frequency and number of revisions and by subsequently attempting to reconstruct series.

The standard approach to preserving continuity after classification revisions is to create linkages where the series break. This is accomplished by producing the data series using both the old and new classifications for a given period of transition. With the dual classifications of data, the full impact of the revision can be assessed. Data producers then may measure the reallocation of the data at aggregate industry levels and develop a concordance between the new and old series for that given point in time. The concordance creates a crosswalk between the old and new classification systems which quantifies with "coefficients" the reduced or expanded scopes of all relevant industry codes in the classification system. For example, for employment series, the coefficients denote how much employment was reallocated from one industry group to another due to the revision.

Concordance coefficients do not constitute a real continuation of a series, but do provide a means to splice or link the break (Beekman [3], 6). Limited conversion of historical data, however, is possible. The coefficients can be applied to data classified under the old classification system to convert it to the new standard. This procedure only approximates what the earlier observations may have been and clearly must be applied with caution. A revision of the economic classification system encompasses broad changes to reflect innovations in industrial composition. Therefore, full conversion of the earlier series segment to the revised classification system is not only unsound because of the approximate nature of the coefficients, but also because the new classification principles do not necessarily reflect the economic reality of the historical data. While it is arithmetically possible to apply concordance coefficients to convert a historical series, certainly in practice it is rarely done.

Jacob Ryten (Williamsburg [9], pp. 475-6) recommends the use of microdata to compare time series derived from the old and new classification systems. The use of microdata would permit finer

comparisons and improve quality control for the concordance.

Research suggests advantages from using microdata to form linkages. However, only limited research on the topic has so far been carried out. In addition, adoption of microdata techniques for estimating concordances would be limited to those industries for which extensive product and employment data are collected and to those periods when detailed surveys are done.

In spite of efforts to create linkages for a transition from one classification standard to another, the utility of a time series still diminishes in application. If the basic framework of the classification system is completely revised, linking or reconstructing time series may not be possible, or possible only at great cost. With a distinct new classification structure and coding principles, it may be difficult to document coherent relationships between the new and old systems or to estimate concordance coefficients for users to link the new and old series.

It is also important to note that merely establishing a means to link time series from before and after a substantial revision may not be adequate to give consistent, accurate forecasts or analyses. Concordances for linking a series coded under two distinct classification systems may be complex and difficult to use. Major definitional changes, reclassification, and changes in reporting procedures resulting from a comprehensive classification revision may alter some time series, and introduce inconsistencies, biases, or distortions into the linked or reconstructed series.

When interviewed in the past, forecasters have indicated such changes in variable definitions and classifications, despite some overlap in the series, have caused serious problems with the output of their models and the related analysis (Bishop and Werbos [4], p. 8). There are substantial costs to data users in testing revised data in models and evaluating the output for reliability. Additional costs may be incurred if there is a need to develop alternative models or data.

However, changes to the classification structure are not the only changes that affect time series. Joel Popkin (Williamsburg Conference [8], p. 22) pointed out that maintaining consistent time series of establishments is problematic, given rapid technological and structural changes. Establishments may switch from one industry into another if at

various points in time they may have different primary products. Other changes that affect time series include changes to the questionnaires and changes in collection methodology.

Thus, time series may be disrupted even if an outdated classification structure is maintained. A complete break in time series may also be preferred to data series that are classified with outdated or inconsistent concepts.

#### 5.4 Activities and Costs of Previous Revisions

Producing a time series under two classification systems for however long a time period is neither an easy nor an inexpensive endeavor for the government agencies involved. The costs of converting the BLS business register to the 1987 SIC, for example, were described in detail by Brian MacDonald (Williamsburg Conference [6]). The total cost for implementing the 1987 revision to the BLS business register was \$9.8 million over a 5-year period beginning in fiscal 1987. At the time, the register consisted of 5.7 million reporting units, approximately 800,000 of which required a special refiling survey and labor-intensive analysis to reclassify to the current system. This business register is the primary sampling frame for BLS establishment surveys such as the Current Employment Statistics series and, as such, the updating of the business register constitutes the central cost and workload for implementing classification revisions to BLS data. Some of the major cost factors to BLS for the revision were:

- Hiring and training of additional staff;
- Computer systems enhancements and programming, including the writing, programming, testing, and documenting of software for installation at BLS and in the states;
- Designing, printing, and distributing survey forms for a special refiling survey; and
- Developing, printing, and distributing training and reference materials for staff.

The Census Bureau estimated that implementing the 1987 SIC revision cost \$4.0 million over 3 years. These costs included the following:

- Mailing and processing of classification cards to small establishments included

in Census files that were affected by the revision:

- Preparatory work for the 1987 Economic Censuses, which included designing Census forms, computer programming, and system enhancements needed to tabulate and publish data on both the old and new SIC basis; and
- Additional work needed to convert the many current Census surveys to the new SIC.

Census does not revise its historical series to incorporate changes to the SIC. Rather, it publishes data for the transition year on both the old basis and the new basis, providing the information needed by data users to adjust the series for prior years, if necessary. These tables, included in economic census publications, are referred to as "bridge tables."

As noted earlier, revisions also impose costs on the users of industrial statistics. These costs include the costs of learning and understanding the new system, and of adapting models and analyses to the new system. Though the ECPC is cognizant of such costs, no direct estimates exist. Costs may also be imposed on respondents to government surveys, who may have to adapt their responses to new survey forms and so forth that may be required by changes to the classification system.

Clearly, statistical agencies, which publish time series, require sufficient time to prepare for a revision, as they must continue to produce data throughout the transition from one classification standard to another. The experience of the 1987 SIC revision illustrates that advance planning is essential to implement even a modest revision of an existing system.

#### 5.5 Implications if a Fundamental SIC Revision is Implemented

Time series breaks confound historical analysis. Accordingly, they should be imposed only after careful consideration of the benefits against the analytical losses.

At one extreme, maintaining time series consistency for statistics that have limited analytical relevance confers little real benefit. The numerous "not elsewhere classified (nec)" industries distinguished in the present U.S. SIC provide examples. By definition, an "nec" industry is a

catchall, combining statistical units that have relatively little in common with each other. For example, SIC 7389, "Business Services, nec," contains such diverse activities as convention bureaus, telemarketing, and wig styling. They are grouped together precisely because in the last SIC these activities were too small to meet the size criterion for forming an industry and did not fit naturally with some other economic activity. Removing from the "Business Services, nec," category some economic activity that has expanded rapidly in the 1980's might diminish the historical continuity of SIC 7389. But because SIC 7389 does not provide, and was not anticipated to provide, useful economic data, loss of time series continuity has little if any cost in this case, and there is only analytic gain from recognizing the new economic activity.

Conversely, making changes to economic classifications for marginal reasons may confer very large costs because the loss of historical continuity outweighs the very small benefit accrued from increased analytical usefulness. Edward F. Denison [5] is a proponent of the view that past U.S. SIC revisions have too frequently made changes that have interrupted time series continuity without notable gains in analytic usefulness.

Most cases probably lie somewhere between the two extremes. There is clearly a trade-off between costs imposed by breaking historical continuity on the one hand and the gains from improving relevance and analytical usefulness on the other. When the gains provided by an improved classification system are high, users will be more willing to absorb the costs of broken historical continuity. And when the analytic gains from changing an industry classification are relatively low, one should give more weight to the cost of broken historical continuity. Because objective data for establishing these costs and benefits are not available to the ECPC, a great amount of judgment must be employed.

### 5.6 The Committee's Position

The rationale for a "fresh slate" review of classification systems implies that it may be possible to create a new system that better meets user needs. Conceptual issues and foundations for economic classification systems are discussed in ECPC Issues Paper No. 1, "Conceptual Issues," and Issues Paper

No. 2, "Aggregation Structures and Hierarchies." Because a new system implies a certain number of breaks in historical continuity, the discussion in this issues paper indicates that the Committee must consider the trade-offs between time series continuity and modernization of the economic classification system during the development of a new classification system. In addition to evaluating the gains and losses in time series utility, the costs of implementing revisions and maintaining continuity in time series must be taken into account as well.

The Committee recognizes, however, that it is unproductive to collect and maintain time series data that have questionable value. Thus, it may be preferable to accept a one-time break in historical continuity if the benefits of conversion to a new classification structure are apparent and accepted by users.

### 5.7 Request for Comment

To assist in weighing the advantages of a new economic classification system against the disadvantages of disruptions to series continuity, the Committee would like comments on how users of data view time series continuity. The Committee desires to learn how data users weigh the trade-offs between an improved classification structure and breaks in time series. The costs and benefits that users experienced following previous revisions would be helpful as part of this commentary, even though both the potential analytical gains from the ECPC's "fresh slate" examination and the potential disruption of time series may be larger than with previous U.S. SIC revisions.

### References

- [1] Bureau of the Census, *Proceedings*, International Conference on Classification of Economic Activities, Williamsburg, Virginia: U.S. Department of Commerce, November 6-8, 1991, 587 pages. (Referenced in the following as Williamsburg Conference.) Available from Bureau of the Census, Room 2069-3, Washington, D.C. 20233.
- [2] Economic Classification Policy Committee, Issues Paper No. 1, "Conceptual Issues," and Issues Paper No. 2, "Aggregation Structures and Hierarchies," *Federal Register*, March 31, 1993, pp. 16990-17004. Available from Economic Classification Policy Committee (BE-42), Bureau of

Economic Analysis, U.S. Department of Commerce, Washington, D.C. 20230.

[3] Beekman, Michael M., "International and National Standard Classifications," Third Independent Conference, International Association for Official Statistics, n.p., 1992.

[4] Bishop, Yvonne M., and Werbos, Paul J., "An Interagency Review of Time-Series Revision Policies," Report of the Subcommittee on Guidelines for Making and Publishing Revisions and Corrections to Time Series, Washington, D.C., n.d.

[5] Denison, Edward F., "Introductory Statement (Round Table of GNP Users)," in Murray F. Foss, ed. *The U.S. National Income and Product Accounts: Selected Topics*. Studies in Income and Wealth, Vol. 47. Chicago: The University of Chicago Press, for the National Bureau of Economic Research, 1983, pp. 313-5.

[6] MacDonald, Brian, "Administering and Implementing a Dynamic Classification System," Williamsburg Conference, pp. 549-57.

[7] Monk, C. Harvey, "Revising and Implementing the 1987 U.S. Standard Industrial Classification System," Seventh International Roundtable on Business Survey Frames, Copenhagen, 1992.

[8] Popkin, Joel, "Monitoring Economic Performance in the 21st Century: Measurement Needs and Issues," Williamsburg Conference, pp. 43-72.

[9] Ryten, Jacob, "Inter-Country Comparisons of Industry Statistics," Williamsburg Conference, pp. 472-83.

[10] Struijs, Peter, "The Concept of Industry in Transactor-Based Industrial Classifications," Williamsburg Conference, pp. 364-83.

