

Economic Classification Policy Committee

Issues Papers

Issues Paper No. 3, Collectibility of Data

May 1993

Note to reader: This is the third 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. 3

Collectibility of Data

Introduction

Economic Classification Policy Committee (ECPC) Issues Paper No. 1, "Conceptual Issues," and Issues Paper No. 2, "Aggregation Structures and Hierarchies," discuss structuring economic classifications according to the needs for use of the data. It is important to emphasize that in meeting the needs of data users, statistical agencies are limited by the information that can be obtained from data providers, the respondents to government surveys. Issues Paper No. 3 discusses the limitations on an economic classification system that arise from problems in collecting the data, and is a complement to Issues Papers Nos. 1 and 2.

Data collectibility limitations take two forms. First, the coding of business responding units (currently, the establishment) is limited by the information that is available to the coding agencies to determine the industry of the respondent. Section 3.1 discusses this limitation. Necessary background, provided in section 3.1, is a review of current coding practices followed by the major government statistical agencies that are responsible for industry coding under the present system.

A second collectibility problem goes beyond mere coding. To provide meaningful information on an "industry," the units placed in that industry must be able to report the information being sought in statistical agency programs, such as employment, sales, revenues, intermediate purchases, prices, and so forth. If a proposed industry does not meet the test of respondent reportability, this also places practical limits on useful definitions of "industry." This topic is discussed in section 3.2.

Finally, a classification system that is erected on an economic concept, as proposed in Issues Paper No. 1, may make demands on data collection that differ from those of the past. This matter is discussed in section 3.3.

3.1 Collectibility of Data - What Can the Statistical Agencies Code?

Four agencies, Bureau of Labor Statistics, Bureau of the Census, Internal Revenue Service, and Social Security Administration, are responsible for

most U.S. industry coding. These agencies code establishments into a Standard Industrial Classification (SIC) each year or update SIC codes of establishments on a regular basis. Besides providing internal classification information for use in their own programs, some agencies also provide classification information on small companies to the Census Bureau.

Following is a brief description of each agency's role in SIC coding, largely excerpted from Statistical Policy Working Paper 11, "A Review of Industry Coding Systems" [6].

Bureau of Labor Statistics (BLS)

The BLS produces major data series on the work force, wages, and prices. Detailed industry data are presented as part of many of these series, and information is collected and published at the 4-digit SIC level of detail as well as aggregated to the 2- and 3-digit level.

BLS updates each establishment's SIC code every 3 years based on information obtained through its Annual Refiling Survey (ARS). Each year one-third of its universe of employers (establishments) is asked to verify or update industrial, geographic, and ownership codes. Establishments list their principal products and activities and provide a percent of total sales or value of receipts for each. The states are actually responsible for this update and specially-trained state personnel assign 4-digit SIC codes to establishments based on the ARS.

Internal Revenue Service (IRS)

Industry coding by the IRS is based largely on self-coding by the taxpayers or coding from written descriptions of business activities on the tax returns. For example, partnership and corporate tax return forms carry a taxpayer assigned Principal Business Activity (PBA) code, which may be constructed from a combination of 4-digit SIC codes or 2- or 3-digit SIC codes. Sole-proprietor businesses select from codes prelisted on the tax form and also describe their business activities.

Tax returns are classified by PBA annually for both employers and nonemployers. The PBA code is the only source of SIC coding for many businesses and is the code used by Census for establishments not included in its samples. Since PBA codes are not as detailed as SIC codes, full coding to the present SIC system cannot be obtained.

Social Security Administration (SSA)

Currently, the SSA is responsible for assigning SIC codes to new businesses. IRS Form SS-4, Application for Employer Identification Number, requires the business to describe its primary business activity, from which an SIC code is assigned by SSA personnel. When additional information is needed to assign an SIC code, SSA sends a classification form which asks the business its principal business activities and if the business has two or more activities, the percent of business done in each. These SIC codes for new businesses are then provided to Census for its use in updating and maintaining a current list of businesses and their activities.

SSA does not, however, always code to a 4-digit SIC, but may classify some business activities to a partial 2- or 3-digit code or to "foldback" codes. Foldback codes are consolidations of two or more SIC codes in related areas and are used only in nonmanufacturing. An example is retail restaurants and bars, which are two separate industries in the SIC (5812 and 5813). SSA uses a foldback code 5811 for these businesses because these activities are commonly reported together on administrative sources with insufficient information to determine the primary activity. Foldback codes are used for some industries to expedite coding, when further correspondence with the company is not desired, or because the SSA does not need the full detail for its statistical purposes.

Bureau of the Census

The Census Bureau has several programs which assign or update industry codes--Economic Censuses, Annual Survey of Manufactures, and annual business register updating for the Standard Statistical Establishment List (SSEL), including the Company Organization Survey. The economic censuses, conducted every 5 years, include approximately 4.2

million establishments. All multi-establishment companies are included in the censuses, but only about 3 million single-establishment companies are included. Information on the remaining 10 million small single-establishment companies is obtained from the administrative records of the IRS and the SSA through the processes described in previous sections. Use of IRS and SSA records significantly reduces the reporting burden for small firms.

Detailed information on products, materials, and types of business activities is collected in the economic censuses. A 4-digit SIC code is derived from the detailed data reported by establishments and is compared against the SIC code in Census files. Information is used to update the SIC code if an SIC code exists or to assign a code if none resides in Census files.

For example, establishments producing knit shirts can be classified in industry 2253, Knit Outerwear Mills, or 2321, Men's and Boys' Shirts, Except Work Shirts, depending on whether the shirt is knit from yarn (2253) or cut and sewn from purchased knit fabric (2321). To code such establishments correctly, a check is first made to determine the materials used, i.e., whether the establishment reports using yarn or fabric. Next a check is made to determine the type of business and kind of operations reported by the business, i.e., making knit apparel from yarn or making apparel from purchased fabrics. If the business uses yarn to make the fabric and checks that it is a knitting mill, the business is coded into industry 2253. If it uses purchased fabric and checks that it is a cut and sew operation, the establishment is coded into industry 2321. If there is a conflict between the two items, further information is requested.

The Company Organization Survey, mailed annually to companies with 50 employees or more, asks companies to update, via a written description, the SIC code for each establishment they own. Companies with fewer than 50 employees are sampled and updated.

Information Required for SIC Coding

As Thomas Jabine stated some time ago, the present system requires a great deal of information for coding 4-digit SIC industries:

"In a broad sense, information is needed about the economic activities of an establishment: what

products it produces, processes, or sells, and what services it provides. With respect to products, it is sometimes necessary to know what materials are used and whether they are produced in the same establishment. It may also be necessary to know how products will be used and whether they are custom produced for particular clients. It is often necessary to know the process used to produce them and where they are produced. With respect to sales, it is essential to know the major class of customers, since that is the main basis for distinguishing wholesale and retail industries. It is also necessary in some cases to know whether the product is new or used, and what the method of selling is: from a store, by mail order, from vending machines, or door to door. For services, it is necessary to know whether they are for other establishments in the same enterprise or for external clients.... It may be necessary to know the particular product or service, the location from which it is leased or rented, whether the lessee is acquiring an equity and, for certain kinds of equipment, whether an operator is provided.

"Some information requirements are hard to fit into any general category. For example, drug stores are classified primarily by their 'trade designation,' i.e., whether the business name implies that the establishment is a drug store.... For banks, it is necessary to know if they are chartered by the National Bank Act or by one of the states or territories.

"When an establishment has activities in more than one SIC industry, it is always necessary to know the relative importance of its activities, based on whatever measure has been adopted for the coding system. The simplest principle is to assign the code for the industry with the largest proportion of total activity, but more complex rules apply in some cases (drug stores)." [4]

Only the Census Bureau and only economic censuses collect all of the required information for detailed SIC coding. IRS and SSA rely on written descriptions or check boxes which list, as explained earlier, principal business activity codes which may or may not equate directly to a 4-digit SIC code. Census codes 4.2 million establishments, while SSA and IRS code the remaining 10+ million establishments.

Implications of the Present System for SIC Coding

The present system requires detailed information on an establishment's activity for correct coding. This information is now only collected for 4.2 million establishments included in the economic censuses. Codes are assigned by BLS, SSA, and IRS on much less information. This may lead to erroneous coding or an increase in uncoded establishments.

For example, Census has found over the past few years that the number of uncoded new establishments received from administrative sources has increased significantly. Census recently compared SIC codes received from other agencies to codes derived from Census sources. The study found that Census and PBA codes matched only 72 percent of the time at the 2-digit level. This study underscores the need for a restructured SIC that either simplifies coding or allows for new procedures to assure that proper codes are assigned to establishments by all agencies.

In the past, proposals for 4-digit SIC industries have sometimes been rejected when non-Census statistical agencies were unable to code establishments accurately into 4-digit industries from the written or check box information used for coding. An example is a request to separate "fast food restaurants" from SIC 5812, Eating Places.

The current SIC 5812 includes all types of eating places, from ice cream and soft serve shops to cafeterias to fast food places to sit down restaurants. Fast food has not been established as a separate 4-digit industry because non-Census agencies find it difficult to code to this level of detail. A fast food restaurant is likely to describe its activity as "restaurant services" rather than as a "refreshment" or "fast food" restaurant. The cost to the other agencies to identify fast food restaurants is excessive and the information is not needed for agency programs.

Census, however, recognizes these distinctive restaurant activities in its census of retail trade and assigns a code that is more detailed than the 4-digit industry code. For industry 5812, businesses are asked to check a box which best describes their activity. There are seven possible restaurant activities: restaurants; social caterers; cafeterias; refreshment places (fast food restaurants); contract feeding; ice cream and soft serve shops; and frozen yogurt shops.

In the 1987 census, separate information on the number of establishments, sales, payroll, and number of employees was published on three of these kinds of activities: restaurants, cafeterias, and refreshment places (fast food). In 1987, there were 154,700 restaurants with a total value of sales of \$66.4 billion and 135,100 refreshment places with sales of \$56.9 billion. These activities meet the size criteria and have the specialization ratio necessary to be recognized as 4-digit industries, based on the guidelines used in past SIC revisions (see forthcoming ECPC Issues Paper No. 4, "Criteria for Determining Industries").

Possible Changes in Agency Coding Responsibilities

As part of a major effort to reduce the tax reporting requirements on employers, the IRS is currently studying changes in the employer tax reporting forms that provide the information for some current SIC coding. Simplified employer forms may have, as well, implications for SIC coding beyond the agencies that now make direct use of IRS coding for statistical purposes.

Though it is too early to determine the implications for SIC coding of any changes in employer tax reporting requirements that may eventually be put into effect, the ECPC is discussing the matter with the IRS Statistics of Income Division, and will take tax reporting changes into account in its review of economic classification systems.

3.2 Collectibility - What Data Can Business Provide?

Statistical agencies must assure that information collected is reportable by companies with a minimum of burden. Whether businesses can provide the data is a major consideration in determining whether or not 4-digit industries can be established. Requests to collect information that is not readily available from existing company records or that would impose excessive costs and burden on businesses are rejected.

For example, there has been increased interest in establishing a 4-digit industry for "tourism." This would include, in addition to travel agencies, theme parks, etc., parts of many industries currently recognized separately in the SIC, such as hotels and

airlines. To capture the activities of the "tourism" industry, it would be necessary for hotels, restaurants, and airlines to differentiate receipts between business and tourist customers. Travel agencies also would need to report whether or not receipts were for business or personal travel. To provide for a 4-digit tourism industry, or for an aggregation at a higher level, it also would be necessary to separate employment and expense data between those associated with tourism and those associated with business travel. Since it is highly improbable that a business (hotel or restaurant) could provide this kind of information, the data are not collectible because business does not maintain its records in this manner. The costs and burden to businesses to provide this kind of information are probably excessive. If there are compelling reasons for recognizing industries such as "tourism," alternative and innovative approaches to measuring such problem sectors should be explored.

Small businesses have been growing in number and importance in the past few years. The 1992 Census of Construction Industries pretest found that small companies in the construction industry generally maintain records based on tax reporting guidelines and often do not have the detailed kind of information needed to assign industry codes. Census experience with the economic censuses indicates this is true for most small businesses in all sectors. Small business classification must be simple, impose minimal burden, and be cost effective.

In 1989, the Census Bureau conducted a Recordkeeping Practices Survey [3] to determine the kinds of information maintained in company records. Questions were asked about the availability of detailed information such as employees by function, capital equipment expenditures, and sales of products or services by type. The survey found that recordkeeping practices vary widely, both within and across industries. Surprisingly, recordkeeping practices also vary within companies. Information such as employment and payroll is usually available by establishment, but detailed information on materials consumed or purchased and products and services produced or sold by type is much more difficult to report.

3.3 Data Needed for Classification (Restructured System)

The International Conference on the Classification of Economic Activities [2] identified the lack of a single, conceptual framework as a problem with the existing system. The conflicting views on the importance of a conceptual framework are outlined in ECPC Issues Paper No. 1.

Two primary concepts have been proposed for a restructured classification system: supply-based (production process) and demand-based. Regardless of the conceptual basis, information is required for classification. Let us briefly examine what information may be required under each concept.

A demand-based classification system requires detailed information on the output of the unit from which information is collected. Joel Popkin (Williamsburg Conference [5], p. 186) proposes the development of a demand-based system. He states that, "output aggregation would classify commodities on the basis of the markets in which they are sold. To a great extent this type of aggregation would also reflect end-user, i.e., whether the market serves final or intermediate consumption. The focus would be on the aggregation of commodities and services, wherever produced, by similarity in characteristics." Depending upon the kind of demand-based system developed, the following kinds of information might be required:

- Type and value of products sold or services rendered.
- Class of customer to which the product is sold or the service is rendered, e.g., other businesses, consumers, government, etc. For services, this may mean whether the service is provided for other establishments in the same enterprise or for external clients.
- End use of the product.
- Characteristics of the product.
- Boundaries of the market, or a listing of other products or services that are close substitutes or are marketed or used together.

For a supply-based or production process system, Joel Popkin (ibid, p. 187) states that "to group establishments by similarity of production structure, their production structure must first be identified." More specifically, for a supply-based system the following information might be required:

- Production process used to produce the resulting product or set of products. This may include whether or not the output is a product of a vertically integrated plant or an assembly plant or the method of manufacturing used.
- Capital equipment input quantities or costs.
- Materials and services input quantities or costs.
- Labor input quantities or costs.
- Energy input types, quantities, or costs.
- Characteristics of the labor force (especially for services).

Much of the information needed for classifying establishments into either a demand-based or a supply-based system is now collected in the economic censuses. The current system is, however, a multiple concept system (see Issues Paper No. 1), and it is not at yet clear if information could be collected to classify all establishments based on a single concept. It also is not clear if coding would be more or less difficult than it is in the present system. That depends on the availability of information in company records and the ability of agencies that code establishments to collect the information.

Discussion

What is collectible from businesses and what information agencies responsible for coding in the decentralized U.S. statistical system can collect are critical issues that the Committee must address in defining a framework for an economic classification system. As it considers these issues, there are important points to keep in mind.

- The economy is dominated by a relatively few, large companies that operate hundreds and thousands of locations.
- * There are 15 million+ businesses in the United States, 9 million of which have no paid employees and probably account for less than 5-10 percent of economic activity. IRS is responsible for assigning SIC codes to these 9 million companies with no paid employees.
- * 165,000 companies operate more than 1 million establishments and account for approximately 45 percent of economic activity. Census collects detailed

information, sufficient for SIC coding, from these establishments every 5 years, with most of them updated at least once between censuses.

- Detailed information on economic activities, necessary for the assignment of SIC codes, is only available in the economic censuses, which
 - * are conducted every 5 years;
 - * cover only 4.2 million establishments; and
 - * under existing legislation, the classification of these 4.2 million establishments cannot be shared with other agencies.

Depending upon the information needed, Census may be the only agency with the data necessary to classify statistical units using a conceptual basis.

Any decisions on the structure of an industry classification system must take into account the ability of the data provider to furnish the required information without undue burden. Business responding units must be able to report the information needed without undue burden and government agencies responsible for SIC coding must be able to code the activities without undue cost. Proposals for new or altered 4-digit industries that are presented for the Committee's review should include consideration of the collectibility of the data needed for establishing an industry before presenting proposals for the Committee's review. Each proposal will be closely examined to assure that the issue of collectibility of data has been addressed.

In the decentralized U.S. Government statistical system, it is extremely important that agencies such as IRS and SSA provide industry codes to statistical agencies for business entities through their administrative sources. Both cost and burden savings mandate this approach to industry coding. What agencies can collect within their resources is as important as what information businesses can provide to these agencies. Each of the agencies coding business entities has different legal and administrative responsibilities, and it is important to understand the different uses of SIC coded data for each of these agencies.

Request for Comment

The Committee is interested in the kinds of data that businesses can report and the most efficient methods for reporting that data. The Committee

invites comments on what kinds of information can be reported by businesses, including the availability of that data within company records and possible new avenues to explore in the realm of data collection.

Assigning industry codes to 15 million businesses in the decentralized statistical system of the United States is difficult. Are there ways not currently considered that would minimize cost and burden to both business and the government? How can agencies responsible for classification coding assure accurate coding with a minimum of cost and burden to both business and the government? The Committee invites comments on these questions.

References

- [1] Bureau of the Census, "Characteristics and Quality of the SSEL Standard Industrial Classification (SIC) Codes," U.S. Bureau of the Census Research and Evaluation Paper, February 1993.
- [2] 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.
- [3] Bureau of the Census, *1989 Recordkeeping Practices Survey*, U.S. Department of Commerce, December 1990. 27 pages. Available from Bureau of the Census, Room 2069-3, Washington, D.C. 20233.
- [4] Jabine, Thomas B., "The Comparability and Accuracy of Industry Codes in Different Data Systems," Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council, National Academy Press, Washington, D.C., 1984.
- [5] Popkin, Joel, "Recommendation and Description of the Principles Upon Which a Revised Industrial Classification System Should Be Built," Williamsburg Conference, pp. 157-216.
- [6] U.S., Executive Office of the President, Office of Management and Budget, *A Review of Industry Coding Systems*, Statistical Working Paper 11, Statistical Policy Office, Office of Information and Regulatory Affairs, March 1984.