The Comparability and Accuracy of Industry Codes in Different Data Systems (1984)

Pages
152

Size
8.5 x 10

ISBN
0309323665

Jabine, Thomas B.; Committee on National Statistics; Commission on Behavioral and Social Sciences and Education; National Research Council

The Comparability and Accuracy of Industry Codes in Different Data Systems

Thomas B. Jabine

Committee on National Statistics
Commission on Behavioral and Social Sciences and Education
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1984

E4-0064
2.1

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Research Council was established by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and of advising the federal government. The Council operates in accordance with general policies determined by the Academy under the authority of its congressional charter of 1863, which established the Academy as a private, nonprofit, self-governing membership corporation. The Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both Academies and the Institute of Medicine. The National Academy of Engineering and the Institute of Medicine were established in 1964 and 1970, respectively, under the charter of the National Academy of Sciences.

Available from

Committee on National Statistics
National Research Council
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Printed in the United States of America

Order from
National Technical
Information Service,
Springfield, Va.
22161 PB 84-215607
Order No.

COMMITTEE ON NATIONAL STATISTICS

CONTENTS

FIGURES, TABLES, AND EXHIBITS

FIGURES

TABLES

EXHIBITS

ABBREVIATIONS

BEA	Bureau of Economic Analysis
BES	Bureau of Employment Security
BIR	Bureau of Internal Revenue
BLS	Bureau of Labor Statistics
CWHS	Continuous Work History Sample
D&B	Dun & Bradstreet Corporation
EIA	Energy Information Administration
EIN	Employer Identification Number
ERP	Establishment Reporting Plan
ESIC	Enterprise Standard Industrial Classification
FTC	Federal Trade Commission
IPC	Industry and Product Classification
IRS	Internal Revenue Service
OASI	Old Age and Survivors Insurance
OMB	Office of Management and Budget
PIA	Principal Industrial Activity
QFR	Quarterly Financial Report
SEC	Securities and Exchange Commission
SIC	Standard Industrial Classification
SOI	Statistics of Income
SSA	Social Security Administration
SSB	Social Security Board
SSEL	Standard Statistical Establishment List
UI	Unemployment Insurance

CHAPTER 1

INTRODUCTION

Industry classifications are assigned to business enterprises for both statistical and nonstatistical purposes. In the United States, industry classification is done by many federal and state government agencies and, for commercial and research purposes, by several organizations in the private sector.[1]

HISTORICAL BACKGROUND

In 1939, the U.S. Central Statistical Board (the predecessor, several times removed, of the Statistical Policy Office of the Office of Management and Budget), observing that several federal agencies were engaged in the industrial classification of business enterprises and that these agencies were using various classification systems, proclaimed the need for a standard classification of industries and, to complement it, ". . . a United States Business Directory or Official Mailing List which will show the name, address, and industrial classification of each important business enterprise" (cited in Bureau of the Budget, 1961:1).

The Board's first objective was met fairly quickly through the efforts of an Interdepartmental Committee on Industrial Classification and a Technical Subcommittee on Industrial Classification. Draft lists and descriptions of industries and alphabetical indices were produced separately for manufacturing and nonmanufacturing industries during 1938-1940. After review, printed editions of the Standard Industrial Classification (SIC) were published for manufacturing industries in 1941 and for nonmanufacturing industries in 1942 (Bureau of the Budget, 1957). Subsequently, there have been several revisions of the SIC to reflect both real changes in the structure of U.S. industry and changes in the framework viewed by industry analysts as being most appropriate for their purposes; the last major revision was published in 1972. A revision scheduled for 1982 was postponed (Federal Statistics Users Conference, 1982), and it is now expected that there will be a revision in 1987, in conjunction with the next round of economic censuses.

The SIC is used to classify establishments: basically, business activities carried out at a single location under one ownership (for

further discussion, see Chapter 2, "Definition of Basic Coding Units"). In 1963, a separate Enterprise Standard Industrial Classification (ESIC) was developed and published by the Bureau of the Budget. Revised versions were published in 1968 and 1974. The ESIC is intended for use in classifying enterprise units, consisting of "all establishments under common direct or indirect ownership." Its structure closely follows that of the SIC; with one exception, all ESIC categories can be defined in terms of complete SIC categories.[2]

The SIC and the ESIC are the basis for all industrial classification systems used by federal agencies for their statistical activities and have been widely adopted by other organizations for both statistical and other purposes. However, use of the SIC and the ESIC has by no means eliminated all differences in industry codes assigned to identical units by different agencies. Nevertheless, it can be fairly said that the Central Statistical Board's first objective was accomplished.

The Board's second objective, to establish a central business directory, has been met only to a limited extent. The Bureau of the Census has succeeded in developing a directory of economic establishments, the Standard Statistical Establishment List (SSEL), which became operational in the early 1970s. The SSEL is the list used by the Census Bureau to select units for inclusion in all of its economic censuses and surveys. It includes, for each establishment, an SIC code based on the most complete and available information. As a result, Census Bureau data classified by industry are much more comparable between programs than in pre-SSEL times, when identical establishments were sometimes coded independently in censuses and surveys.

The same level of comparability has not been established between the economic data systems of the Census Bureau and those of other agencies, such as the Bureau of Labor Statistics (BLS), the Internal Revenue Service (IRS), and the Social Security Administration (SSA). The SSEL was funded and developed on the premise that it would be available to other federal and state agencies for statistical purposes. Such availability, however, requires legislation to remove existing legal barriers to interagency disclosures of information for individual business enterprises. In the past 10 years, several drafts of such legislation have been prepared and circulated to executive branch agencies for review; however, none of those drafts has been formally submitted to Congress. Thus, after more than 40 years, the Central Statistical Board's second objective has still not been achieved.

There are now several federal agencies, some with a large annual volume of coding, that classify business enterprises by industry. There is considerable, although not complete, overlap in the various universes or samples of business enterprises coded. For the most part, coding in different agencies is carried out independently, and the fact that all coding systems are based on the SIC or the ESIC does not guarantee that the same code will be assigned to a given unit in different systems. As a result, data by industry produced from these systems are not fully comparable, a fact that has been frequently

noted by users of the data (for recent examples, see National Research Council, 1979:178-179; General Accounting Office, 1979:10-11).

ASSUMPTIONS OF THE REPORT

A basic assumption of this report is that greater comparability of industry coding between systems is desirable. Because the U.S. federal statistical system is decentralized, many users of economic data find it necessary to combine data by industry from two or more agency sources in their analyses: for example, the Bureau of Economic Analysis (BEA) in preparing the national economic accounts and the BLS in analyzing productivity by industry (Office of Federal Statistical Policy and Standards, 1977a:54-56; National Research Council, 1979:177-179). Classification differences for establishments or other economic units are an obvious source of error for these kinds of analyses, and these errors can cause serious difficulties when large units are involved. In addition, in the limited number of cases in which industry codes for individual units are now being transferred from one agency to another for statistical purposes (see Chapter 6), the value of the codes to the agencies receiving them is limited by differences in the classification principles and coding procedures used by the agencies involved.

A further assumption is that more code sharing between agencies is acceptable, both as a means of improving comparability and as a way of reducing the overall cost and respondent burden of producing economic statistics. The primary disagreement with this assumption comes from the agencies that would be asked to disclose industry codes and associated information for individually identifiable units to other agencies. They argue that public awareness of such disclosures might impair their ability to collect complete and accurate information from businesses, whether intended for statistical or other purposes. This is a valid concern and suggests that procedures for increased code sharing should be developed carefully and with full discussion between the agencies and the original providers of the information.

Full comparability of industrial classification systems would not automatically result from the passage of legislation making the SSEL available to other agencies for statistical purposes. As is discussed below, there are many other problems, both technical and administrative, that would have to be overcome. In addition, the question would remain as to what could be done to improve comparability with data from systems used, at least in part, for nonstatistical purposes and therefore not eligible under any versions of proposed legislation to receive industry codes from the SSEL. Nevertheless, the major thesis of this report is that only modest improvements in comparability and efficiency can be achieved by changes to individual systems; significant gains must await removal of some of the existing barriers to code sharing between agencies.

For the most part, this report takes the SIC as given. The classification principles and procedures discussed are those that are used by different agencies to assign codes to particular units in

accordance with the current SIC. The term "classification principles," as used here, refers to agency practices with respect to grouping or splitting SIC categories, choice of reference periods, treatment of changes over time, treatment of multiple activities in a single unit, and other similar matters. It does not refer to the principles used in establishing and revising the SIC, i.e., the guidelines for determining what economic activities should be recognized as separate industries and how those industries should be placed in a hierarchical structure.

ORGANIZATION OF THE REPORT

Chapters 2, 3, and 4 discuss in detail the factors that affect the comparability of industry codes in different systems. The discussion follows the sequence of the flow diagram in Figure 1. The diagram displays the reasons for differences between industry codes in different data systems and between aggregate data classified by industry from these systems.

Some differences result from user requirements: differences in system coverage, definition of reporting units, and classification principles occur because the systems have differing objectives. For example, the industry classification systems used by IRS were designed primarily to support tax compliance activities; statistical uses of industry codes are a secondary consideration. Differences in system requirements are discussed in Chapter 2.

Other differences occur largely because of limitations in the resources available for industry coding. In particular, the source data available for coding depend very heavily on each agency's judgment as to how much detail it can afford to collect and process in the attempt to assign the correct industry code to each unit and whether the same information is also needed for other purposes. These judgments depend both on the monetary costs of various alternatives and on the potential burden on respondents who are asked to provide the source data. Variations in the source data used for industry coding are covered in Chapter 3.

Even if systems were formally fully comparable, some differences would occur because of errors, such as incomplete or incorrect information provided on source documents or mistakes in the execution of coding procedures. The level of error in each system is influenced by the choice of coding procedures and the amount of checking done at each stage of the process. Differences in coding procedures, including quality control methods, are covered in Chapter 4.

To develop effective ways of improving the comparability and quality of industry coding, merely identifying each of the factors that causes differences or errors is not enough; some information is needed on the size of differences and errors associated with each of the factors. The quantitative data available that bear on the comparability and quality of industry coding in major federal data systems are reviewed in Chapter 5.

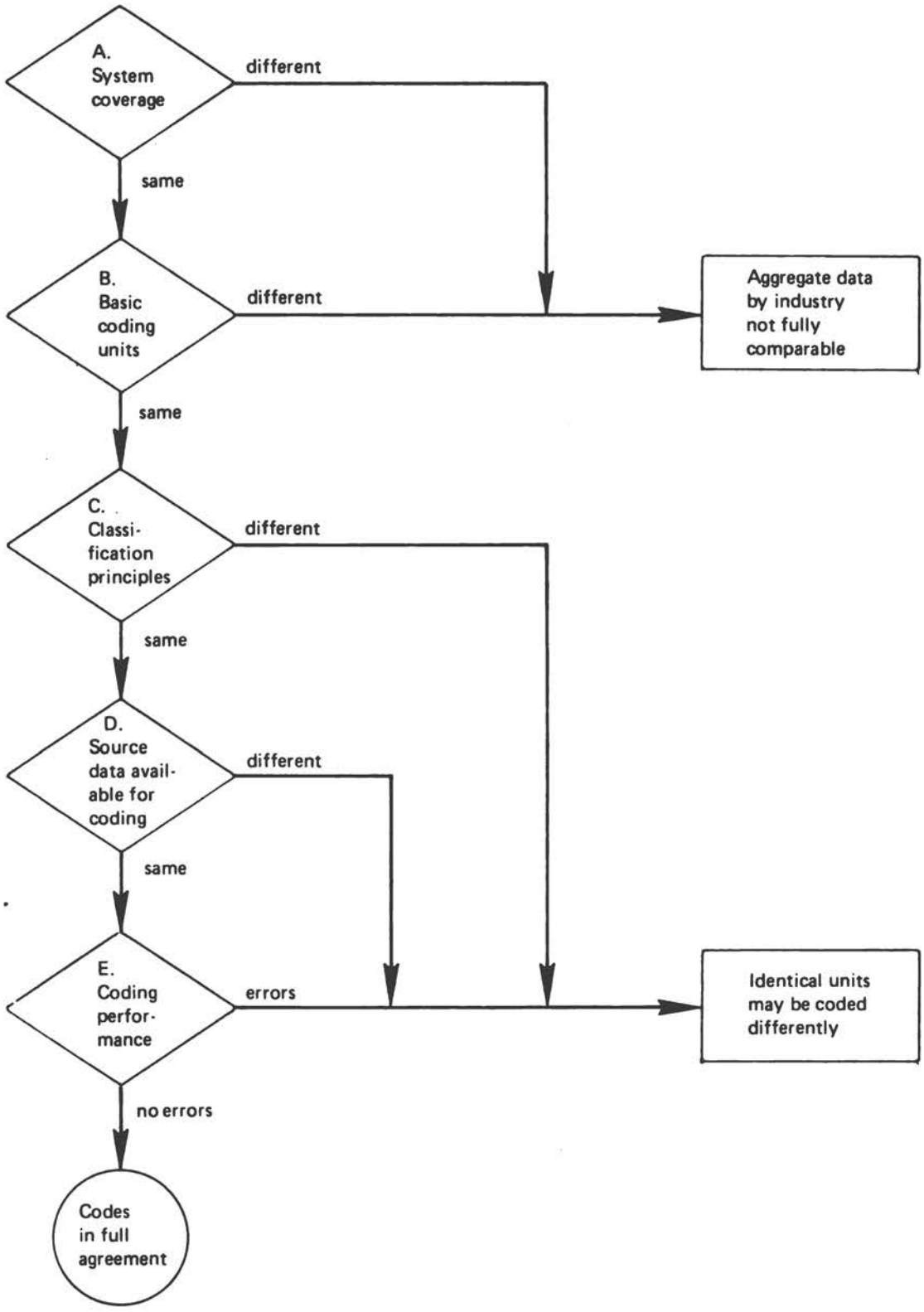

FIGURE 1 Factors that Affect the Comparability of Data from Different Systems Classified by Industry

Chapter 6 summarizes the present status of major industry classification and coding activities and presents recommendations for improving intersystem comparability and efficiency.

SOURCES OF INFORMATION

The discussion of specific coding systems in this report relates largely to 16 major coding systems in 6 federal agencies, which account for a very large proportion of the total industry coding activity by federal agencies (and in the case of the Bureau of Labor Statistics's employment and wage system, by state agencies under a federal-state cooperative program). The discussion of these systems is based on the work of the Industry Coding Working Group, which began in late 1981 under the auspices of the Federal Committee on Statistical Methodology, to review and document those coding systems with a view toward ultimately improving the comparability and quality of data classified by industry.[3] For each of the 16 systems, a detailed system description, following a standard format adopted by the working group, has been or is being prepared. Most of the examples and illustrations in this report are taken from the system descriptions and the supplementary documentation acquired by the working group for the 16 systems.

Two serious difficulties face anyone who attempts a review of this kind. First, it is difficult to get a clear and full understanding of the coding principles and procedures used in a system, and of the reasons for differences between systems, without reviewing some "live data," i.e., looking at the source information for some individual units and the codes that were assigned to them. For the most part, this option was not available to the author. There are few opportunities to review codes assigned to identical units in different systems because of the confidentiality requirements of the agencies involved. Restrictions on access to records of individual establishments also made it impossible for the author to review the instruction materials used by coders in one particular system because the materials included confidential information used in examples.

Second, documentation of significant aspects of some systems is not available or is very hard to obtain. For example, several agencies said that in some systems they used resistance coding (a procedure that takes a unit's prior classification into account in determining the current classification; see Chapter 2), especially for large units, but were unable to provide written instructions or guidelines covering this aspect of the coding process. One would also expect that much could be learned from reviewing statistical summaries showing the kinds of errors detected in manual verification and computer edits of industry codes; however, this kind of information was often not available in convenient form. Evidently, the results of quality control activities are not being used to any great extent to explore possibilities for long-range improvement of the processes. Finally, some older materials of considerable potential interest have apparently not been preserved. A useful report of the Bureau of the

Budget (1961) provides a summary description of several studies that were undertaken in the early 1950s, under the general direction of the Office of Statistical Standards, to examine the relationships between reporting unit definitions and industry classification practices in different agency systems. However, detailed reports of the methods used and results of these studies are apparently no longer available.

Clearly, improved documentation and better access to individual records for methodological research and analysis could do much to support basic improvements in the quality and efficiency of industry coding systems.

NOTES

1. The term "business enterprise" is used in this report in a generic sense to cover all types of business units, including establishments, employers, companies, enterprises, and others.
2. The exception is SIC major group 13, oil and gas extraction. Enterprises are split between two ESIC industry groups depending on whether or not their extractive activities are associated with refining.
3. For a detailed report of the status of this undertaking as of mid-1982, see Farrell et al. (1982). Publication of the working group's final report is scheduled for 1984.

CHAPTER 2

FACTORS AFFECTING COMPARABILITY

SYSTEM COVERAGE

Following the scheme presented in Figure 1 (in Chapter 1), the first factor considered in comparing systems is their respective coverage. Table 1 (adapted from Farrell et al., 1982), compares the coding of 12 major industry systems by SIC division (the 10 primary groups) and legal form of organization. The 12 systems shown in Table 1 include all of the 16 systems documented by the Industry Coding Working Group; however, the 6 IRS systems have been grouped to form 2 systems. The IRS systems are the most complete, covering all divisions except J, public administration, and all forms of organization except "government establishments."

Evaluation of the coverage of division J has been made more difficult by the 1972 revision of the SIC, which changed the principles for the classification of government establishments. Previously most of them had been classified under division J, government; since 1972, each one is to be classified by its primary economic activity, with only those not classified in other divisions to be assigned to the new division J, public administration. One result of this change is that the IRS systems, which do not include any government establishments (since they are not taxed), can no longer be expected to have full coverage in all of the other SIC divisions. The most complete coverage of division J is by the employment and wages system of the Bureau of Labor Statistics, since most public employers are covered by the unemployment insurance system.

For employers, i.e., businesses with one or more paid employees, the BLS employment and wages and the single-unit employer identification systems of the Social Security Administration between them should have virtually complete coverage of all SIC divisions. The BLS system excludes railroads and some "small" agricultural employers (the cutoff varies by state); the SSA single-unit system has only partial coverage of federal, state, and local government employers and of tax-exempt nonprofit organizations. These exclusions reflect the coverage of the social insurance programs in support of which these two data systems were established. For example, railroad employers were initally covered by the unemployment insurance system, but this segment was removed in 1939 and joined with other social

TABLE 1 Coverage of Industry Coding Systems by SIC Division

Agency and Name of System		SIC Division[a] A	B	C	D	E	F	G	H	I	J	Remarks
Systems concerning all forms of organization and including zero employee units												
Census:	Agricultural Census	Pt.										Major groups 01 and 02 only.
IRS:	Statistics of Income	X	X	X	X	X	X	X	X	X		
	Revenue Processing	X	X	X	X	X	X	X	X	X		
Systems concerning all forms of organization, employers only												
BLS:	Employment and wages	Pt.	X	X	X	Pt.	X	X	X	X	X	"Small" agricultural employers and railroads are excluded.
Census:	Business Births						X	X		Pt.		Selected services.
	Company Organization Survey	X	X	X	X	X	X	X	X	X		
	County Business Patterns	Pt.	X	X	X	Pt.	X	X	X	X		Farms and railroads are excluded.
	Economic Census		X	X	X	Pt.	X	X		Pt.	Pt.	Selected categories in the group marked "Pt." IRS assigned codes are used for Censuses zero employee units.

SSA:	Single-Unit Employers	X	X	X	X	X	X	X	X	X	Pt.	Includes federal, state, and local government units that elect social security coverage.
	Multiunit Employers	X	X	X	X	X	X	X	X	X		
Systems covering corporations only												
BEA:	Direct Investment	X	X	X	X	X	X	X	X	X	Pt.	Foreign government parents included for "inward investment" part of program.
FTC:	Quarterly Financial Report		X		X		X	X				

[a]SIC divisions:

A--Agriculture, forestry, mining
B--Mining
C--Construction
D--Manufacturing
E--Transportation
F--Wholesale trade
G--Retail trade
H--Finance, insurance, and real estate
I--Services
J--Public administration

Source: Adapted from Farrell et al. (1982).

insurance provisions for the railroad industry in an independent program. Coverage under the social security system is optional for state and local government employers and for tax-exempt nonprofit organizations. These examples serve to illustrate the general point that the coverage and content of administrative record systems seldom conform completely to the data requirements for statistical systems.

A distinction needs to be made between intended or theoretical coverage, which has been discussed in the preceding paragraphs, and actual coverage, which is what is actually in the systems. Not all businesses required by law to file tax returns do file them, and not all employers who should be covered by social security and unemployment insurance systems actually enroll and pay taxes. Recent concern about the "underground economy" suggests there may be significant numbers of business enterprises that escape these administrative systems; however, such units are likely to be small and to have only minor effects on aggregate data in most industry categories.

DEFINITION OF BASIC CODING UNITS

Coverage of exactly the same industrial activities by two different data systems does not ensure that data classified by industry from the two systems will be comparable. One factor that can prevent full comparability is the use of different units of observation: the basic coding units or simply units, i.e., the units of observation to which industry codes are applied.

Such lack of comparability can be illustrated by the simple example shown in Table 2. Consider two data systems: system 1, which contains data only at the enterprise level, and system 2, which contains data at the subunit (e.g., establishment) level and also has the identification information necessary to group subunits to the enterprise level. In system 1, following the usual practice of assigning an SIC code based on the principal activity, the enterprise would be coded to SIC category Y. In system 2, if data were tabulated by subunit, the activities of subunits A and B would be assigned to SIC category X and those of subunit C to category Y. Clearly, the tabulations by industry from the two systems would not be comparable, even if their overall coverage of activities were identical.[1]

In actual practice, business enterprises consisting of a single establishment, as defined for purposes of the SIC, are identified in essentially the same way in all of the 16 systems covered in this report. There are, to be sure, some elements of judgment in the SIC definition, especially in those instances in which ". . . distinct and separate economic activities are performed at a single physical location . . ." (Office of Management and Budget, SIC Manual, 1972:10). The SIC Manual states that these activities shall be treated as separate establishments if the employment in each is "significant" and "reports can be prepared" on employment, payrolls, sales or receipts, and other establishment type data separately for each activity. These criteria clearly allow some latitude for

TABLE 2 A Business Enterprise with Three Subunits

Subunit	Level of Activity[a]		
	SIC Category X	SIC Category Y	Total
A	65	35	100
B	55	45	100
C	20	80	100
Enterprise Total	140	160	300

[a]Aggregate amounts, in unspecified units.

judgment by the agency collecting the data, and one could expect to find some cases in which establishments were defined differently by different agencies. This might occur, for example, when different activities are carried on by the same company in different buildings at the same general location.

Nevertheless, the major differences between systems with regard to definitions of basic coding units are those affecting only multi-establishment enterprises. Here the systems reviewed use a variety of units, including those with a legal or administrative basis, such as employers, taxpayers, corporations, consolidated corporations, etc., and those with a statistical basis, such as the "reporting units" defined by BLS for the system maintained in connection with its Employment and Wage Statistics Program (ES-202) and by SSA for the multiunit employer identification system.

The "reporting units" used by BLS and SSA deserve special attention. Although they have the same name and have been established for similar purposes, their operating definitions are <u>not</u> identical for multiestablishment employers. Basically, the reporting unit in each case can be a single establishment or a group of two or more establishments under the same employer (same employer identification number, EIN) in the same county and four-digit industry. The procedure for grouping establishments in some cases is intended for the convenience of employers who might find it difficult or burdensome to file separate administrative returns to SSA and to state employment security agencies for each establishment. Since a breakdown by establishment within county-industry groups is not deemed essential to meet the statistical objectives of either BLS or SSA, this compromise between full establishment reporting and reporting at the EIN level has been adopted.

There are three main reasons that the BLS and the SSA multiestablishment reporting units are not fully comparable. First, the definitions are not identical. Second, in the BLS system, the definitions are applied separately in each state only to the establishments operated by the employer in that state, while in the SSA system for multiunit employers all establishments in every state are considered as a group in applying the eligibility test and reporting unit definition. Third, compliance with the definitions is, in a certain sense, voluntary in both systems.

The definitions of reporting units used in the two systems, as provided by the agencies in their systems descriptions, are reproduced in Appendix A.[2] A careful comparison shows the differences clearly; an illustration may be helpful. Suppose an employer has 180 employees in a primary industry and 20 in a secondary industry, all in the same county. Under the SSA definition this employer will be eligible to participate in SSA's establishment reporting plan (ERP) and will be requested to report separately for all establishments or at least separately for those in each of the two industries. The ERP, however, is voluntary and if the employer does not want to participate, the SSA will code all of the employees under the primary industry. In the BLS system, an employer with fewer than 50 employees in all secondary industries need not submit an industry breakdown, in which case all of the employees will be coded to the primary industry. However, states are encouraged to come as close as possible to full establishment reporting, so the treatment of this case would not be the same in every state. The BLS system is mandatory in the sense that all states are required to observe certain minimum standards of detail in defining reporting units, and employers are required to report on the basis established by the states in which they operate. The voluntary aspect of the BLS system consists of the extent to which individual states exceed the minimum standards.

The net effects of these three kinds of differences on the numbers and distribution of reporting units in the two systems are difficult to judge. The BLS system cannot readily distinguish single and multiestablishment reporting units and provide separate counts for the latter. Furthermore, BLS cannot always link reports for the same employer in different states, since not all states include the EIN in the records forwarded to BLS. The SSA systems are capable of providing separate counts of single and multiunit employers and counts of reporting units for the latter, but the only available counts of active units, i.e., those with current-year employment, are about 4 years old.

In summary, lack of comparability due to differences in definitions of basic coding units primarily affects multiestablishment enterprises. As of 1979, about 4,231,000 of 5,181,000 establishments included in the Census Bureau's SSEL were independent or single-unit enterprises, i.e., were not directly associated with any other establishment. However, about 54 percent of all employment was accounted for by the remaining 950,000 establishments associated with multiestablishment (multiunit) enterprises (Bureau of the Census, 1982a). It is probable, therefore, that intersystem differences in

basic coding unit definition will affect aggregate data classified by industry more than they will affect counts of establishments by industry.

CLASSIFICATION PRINCIPLES

Having the same coverage and basic coding units is not enough to guarantee full comparability beteen industry coding systems. Other systemic or procedural features, which come under the general heading of principles for industry classification, can lead to differences in codes assigned to identical units. These principles are discussed in this section. The identification of these principles and their role in an industrial classification system owes much to an article by Simmons (1953).

Adherence to SIC Categories

It is of no particular importance if the numeric codes assigned in a system to specific industry categories differ from those used in the SIC Manual (or the ESIC Manual, where appropriate) provided that a one-to-one transformation exists. The question discussed in this section is the extent to which all of the four-digit SIC categories and only those categories are used in different systems.

Examination of the classification structure for each of the industry coding systems reviewed shows that while each classification system is based on the 1972 SIC or the 1974 ESIC (which in turn is derived from the 1972 SIC), each system departs from it in one or more respects. These departures fall into three categories: (1) grouping of SIC categories; (2) subdivision of four-digit SIC categories; and (3) addition of categories not covered by the SIC. For the systems reviewed, grouping of SIC categories is more common than subdivision of categories.

Differences by Agency

Among the major systems, IRS uses a much smaller number of categories than the others, currently about 200 for each of its 6 systems. The groupings vary by type of organization. There are different groupings for sole proprietors, partnerships, and corporations. For each organization type, the groupings for the revenue processing and statistics of income (SOI) systems are essentially the same. There are a few instances where IRS has subdivided SIC industries. For example, in the partnership systems, SIC industry 7011, hotels, motels, and tourist courts, has been divided into two categories: (1) hotels and (2) motels, motor hotels, and tourist courts.

The BLS and SSA systems use most but not all of the SIC four-digit categories. On the grounds that adequate employer records are often

not available for full four-digit coding in certain industry groups, BLS permits the use of three-digit codes, followed by 0, for establishments in these industry groups. This "exception" is optional; state agencies that wish to may use full four-digit coding throughout. As a result of these provisions, the BLS system carries full detail for 971 of the 1,005 SIC industry codes. For establishments in the remaining 34 industries, representing 9 of the 421 industry groups (three-digit SIC categories), classification is complete only at the industry group level.

In the SSA systems, the full four-digit SIC code is the preferred code, except for major groups 01 (agricultural production--crops) and 02 (agricultural production--livestock) and division J, public administration, for which no further detail is provided. For example, the code 9011 is used for all of the 27 four-digit industries in division J. The codes used for these groups are called "foldback" codes. Thus, 63 of the 1,005 SIC industry codes are not used at all. For some other categories, which account for 115 additional industries, foldback codes are used only if the employer does not furnish enough information to code to the four-digit level; follow-ups for additional information are not attempted by SSA. The use of these foldback codes is said to have been especially heavy during a period in the early 1970s when SSA was doing dual coding--assigning two codes to each employer, one based on the 1967 SIC and one based on the 1972 SIC--in preparation for conversion of its systems to the 1972 SIC. In summary, it seems fair to say that full SIC detail is lacking for 178 of the 1,005 industries in the 1972 SIC.

The Census Bureau's current industry classification system is described in its 1977 Industry and Product Classification Manual (hereafter, IPC Manual, 1977b). The Census Bureau establishment codes carry full SIC four-digit industry detail except when information available for classification is incomplete or when publication of establishment data for a particular industry would disclose individual company operations. Three industries are affected by the latter restriction: (1) mercury, 1092, carried as 1099 (metal ores, not elsewhere classified); (2) typewriters, 3572, carried as 3579 (office machines, not elsewhere classified); (3) electronic tubes, 3671 to 3673, carried as 3671. In addition, for economic census purposes, the IPC Manual provides for subdivision of selected industries in SIC major groups 41, 42, 47, 50-59 and 70-89, i.e., in the areas of transportation, wholesale and retail trade, and services. The "subindustries" are identified by adding two digits (plus a check digit) to the four-digit SIC code. For the 1977 economic census, 83 four-digit industries in these major groups were subdivided to form 256 six-digit subindustries. Two different patterns have been followed in subdividing four-digit industries. In most cases there is only one level of disaggregation for an industry, i.e., the six-digit codes differ only in the fifth digit, and the sixth digit is 0. In a few cases, however, there are two levels of disaggregation, i.e., one or more of the five-digit codes will be subdivided by using different digits in the sixth position. In past revisions of the SIC,

subindustries already established by the Census Bureau have often been upgraded to the level of separate four-digit industries.

Reasons for Differences

Several factors account for the grouping and subdivision described above. Users of data from the systems generally want more detail by industry; this explains the Census Bureau's use of the industry and product classification and the gradual increase in the amount of detail shown in the other systems. Several factors tend to limit the amount of detail included. Some systems are based on samples that are too small to support more industry detail. This factor is probably the main reason that the IRS SOI systems carry less industry detail.

Another factor is cost and respondent burden. This factor is especially important for the IRS partnership and corporation revenue processing systems, which use self-coding by taxpayers, based on a listing of categories included with tax-return instructions. Expanding this list, which now fits on a single page, to the full four-digit SIC detail would have a significant cost just for printing and would place some additional burden on taxpayers. The question of how the cost of coding from an open-ended activity description would be affected by coding to a longer list is more difficult to evaluate, but it seems likely there would be some increase in the unit cost of coding.

The difficulty in distinguishing between similar SIC categories is a factor that may explain, at least in part, why the BLS and SSA systems combine certain SIC categories. For example, the SSA system uses only a single category for SIC major groups 01 (agricultural production--crops) and 02 (agricultural production--livestock) while the SIC has 36 separate four-digit categories in these groups. To code accurately to the full SIC industry detail would require either the addition of one and possibly more items to the Application for Employer Identification Number (Form SS-4), which is the source document for coding, or a substantial amount of follow-up activity. For another example, both BLS and SSA combine "eating places" (5812) and "drinking places" (5813). The distinction hinges on whether the greater portion of receipts comes from the sale of alcoholic beverages for consumption on the premises or from the sale of food. This information is often not readily available, and some employers may be reluctant to provide it, especially to state officials, for fear that they may be adversely affected under various laws and regulations governing the sale of alcoholic beverages. The IRS systems for sole proprietorships and partnerships, on the other hand, keep these two industries separate even though the IRS systems have a much smaller total number of categories than do the BLS and SSA systems.

The addition of categories not included in the SIC is of fairly minor importance as far as comparability is concerned. For example, the SSA systems have separate codes for foreign governments and international organizations, since these units may be subject to social security taxes for U.S. citizens working in their U.S.

installations. For another example, the industry classifications established by IRS for revenue processing of sole proprietorships include the following groups:

9100 Government positions, public officials, sheriffs, postmasters, tax collectors, notaries, and related occupations.
9910 Ministers
9920 Sextons, chaplains

None of these categories covers establishments in the usual sense; they are included in the IRS system to cover occupations receiving special treatment under the tax laws. Since these kinds of units are put in separate categories, they do not cause any insurmountable problems in making comparisons between systems.

Effects of Grouping and Splitting

The actual and potential effects of grouping and splitting SIC four-digit categories deserve careful examination. For this examination it is necessary to understand first that the SIC is a hierarchical classification system. Establishments may be classified at any of four levels, as shown in Table 3. The categories at any level are defined by subdividing each of the categories at the next higher level: each industry division consists of a group of major industries and so on. The first two digits of a four-digit code identify a major group, and the first three digits identify an industry group within that major group. This principle does not apply to the first digit of the four-digit code, however. Some divisions include major groups with different first digits; conversely, some major groups with the same first digit are in different divisions.

Another important point is that some establishments have more than one economic activity and that those activities do not always belong to the same industry. In fact, an establishment may have activities in more than one SIC division. The existence of multiple activities is a fundamental problem in industry classification. It is also treated more fully below, but is relevant in this discussion of grouping, i.e., classification at a higher than four-digit level.

It is frequently not realized, although explicitly stated in the SIC Manual (Office of Management and Budget, 1972:12), that the classification of an establishment based on its primary activities at the division, major group, or industry group level can differ from that assigned on the basis of its primary activity at the four-digit industry level. This can be readily seen from the example shown in Table 4 of an establishment with multiple activities (the codes are arbitrary and not intended to represent any real situation). A system classifying establishments to the full four-digit level would assign this establishment to industry 7654. A system that assigned only three-digit codes and did not have any information about the breakdown of activities within industry group would, perforce, assign the

TABLE 3 Structure of the 1972 Standard Industrial Classification[a]

Level	Identification	Number of Categories
Division	Capital letter (A through K)[b]	11
Major group	Two-digit code	84
Industry group	Three-digit code	421
Industry	Four-digit code	1,005

[a]As modified in 1977 (Office of Federal Statistical Policy and Standards, 1976).
[b]Divisions are not identified by the first digit of the four-digit SIC code. Each division consists of one or more major groups identified by two-digit codes.

establishment to industry group 123. Clearly the codes in the two systems would not be fully comparable. This is one reason the SIC Manual recommends (1972:12) that:

> Even though a data collecting organization may have no immediate need to analyze or publish establishment data at the four-digit industry level of classification, it may nevertheless be useful to assign four-digit codes to each establishment report wherever the information is available and the incremental cost of such classification is not excessive.

The situation is not hypothetical. The relationship between IRS and other coding systems is similar to that of the two systems described in the example above. However, the importance of this lack of uniformity is not known; an analysis might show that this factor has only trivial effects on comparability. Empirical analysis, using a data base with full information about each establishment's activities at the four-digit level, would be necessary to determine the importance of this factor.

Similar considerations exist for four-digit industries that are subdivided. To maintain comparability in such cases, the SIC Manual recommends (1972:12,n.1) that the establishment first be assigned to a four-digit industry in the usual way and the subindustry code then be

TABLE 4 An Establishment with Multiple Activities

SIC Code	Proportion of Total Activity
1234	.25
1235	.35
7654	.40

assigned to the primary activity within that four-digit industry. The alternative, not recommended in the SIC Manual, would be to assign the establishment to the largest subindustry without regard to four-digit industry categories.

Most agencies that establish subindustries do so by splitting four-digit industries. However, in the IRS systems for coding sole proprietorships, some four-digit industries are split and combined with other four-digit industries, for example:

IRS Codes	SIC Codes
0753	0751 (part)
0754	0751 (part), 0752
8048	8049 (part)
8098	8049 (part), 8081, 8091

For IRS codes 0753 and 0754, comparability can still be achieved at the industry group (three-digit) level, but for IRS codes 8048 and 8098 the arrangement destroys comparability below the major group (two-digit) level. A similar instance exists in the IRS partnership coding systems: investment clubs have been separated from SIC industry 6799 (investors, not elsewhere classified), and the rest of this industry has been combined with other industries in the same major group.

The SIC Manual also notes (1972:13) that grouping can cause problems when revisions are made in the SIC. An industry group (three-digit level) might be revised by shifting some four-digit industries into it and others out of it. If the data base contains full four-digit industry detail, it will be relatively easy to develop historical data for that industry group as defined after the revision; otherwise it may be impossible or very difficult and costly.

Failure to classify establishments with full SIC detail could also eliminate some flexibility for users who want to combine industries into groups that differ from the standard SIC groupings. An example is an allocation scheme developed by Singelmann (1978) in which SIC

major industries, sometimes from more than one SIC division, are grouped to form 37 "industries" (not SIC) that in turn are grouped to form 6 sectors: extractive, transformative, distributive services, producer services, social services, and personal services. An analyst wanting to use Singelmann's classification would only be able to draw on data from publications or systems with full detail at the two-digit SIC level.

In summary, an industry classification system that does not classify to the full four-digit SIC level or that creates subindustries that cross four-digit industries will pay a price in terms of comparability with other systems and historical comparability within the system. Such a system may also be less capable of meeting a wide variety of user needs. These losses should carefully be weighed against any cost savings resulting from the use of a less detailed classification system.

Finally, there is one federal agency (not among those whose industry coding systems were reviewed by the Industry Coding Working Group) that departs significantly from the concepts embodied in the SIC. The Energy Information Administration (1983:Part 9, table footnotes), in its data on marketing of petroleum products, treats all sales to "ultimate consumers" as retail, whether the consumers are in the utility, industrial, commercial, or residential sector. By extension, establishments selling petroleum products are classified as retail if most of their sales are to ultimate consumers. This classification is much broader than the SIC concept of retail sales, which covers only sales for personal or household consumption. Clearly, one could not use Energy Information Agency data directly with data from other sources in an analysis of the retail sector, as defined in the SIC.

Reference Period for Classification

Whatever the set of categories used to classify establishments or other units by industry, the classification of each unit is usually based on that unit's activities during some specific time period or as of a specified date. Since a unit's activities can change over time, the choice of reference period or date is one of the factors determining comparability of industry codes between systems.

When a system is used to produce aggregate data such as employment, payroll, and receipts classified by industry, the reference period on which the industry code is based may or may not be the same as the period to which the data refer. Thus data by industry for the same reference period from two systems may differ because the industry codes are not for the same reference period. In fact, the major industry coding systems reviewed for this report do differ considerably in this respect. The rest of this section presents a broad outline of the practices followed by each of the four agencies.

IRS Tax returns are classified by industry annually, based on either self-coding by taxpayers or coding from an activity description on the

tax return. Thus, for data by industry from the IRS systems, the reference periods for the data and the industry classification always coincide.

BLS Each reporting unit is classified initially when an employer enters the unemployment insurance system. It is BLS policy that codes should be reviewed and updated on a fixed time schedule, as follows:

Type of Unit	Frequency
Units with 500 or more employees, except government	Annually
All other units, except government	Every 3 years
Government units	Every 5 years
Nonclassifiable units	Annually

The timing of the 3-year cycles varies by SIC division so that review and updating is done for reporting units in certain divisions each year. Information leading to code changes may come from other sources between regular updates; the extent of such changes and how well they track actual changes is not known. The source documents used for initial coding and updates request relevant information on activities for the most recent calendar year.

SSA Every employer is classified initially at the time an application for an EIN is filed. The application form asks for information about the nature of the business at the time of filing; there is no defined reference period. Shortly thereafter, eligible multiunit employers are asked to submit activity information for each of their reporting units; instructions on the report form call for information on gross dollar volume for each activity "during a recent period." For single-unit employers, the last general update was based on a comparison with codes assigned in the 1972 economic censuses. For multiunit employers, changes are based either on reports filed voluntarily by employers or on correspondence initiated by SSA when the units for which current wage reports are submitted do not match those in the file. The agency's resources for such correspondence are limited. (Since both the single- and multiunit employer files carry date codes indicating the most recent update of an employer's industry classification, it would be possible to tabulate each file to obtain a distribution of employers by years elapsed since last update.)

Census Bureau For the Census Bureau systems, the reference periods vary by coding system. For units covered by mail (or interview) in economic censuses, the industry classification has the same reference period as the data, which is also true in some but not all current surveys. The industry codes in the SSEL, which provides the frame (except for zero-employee units) for all censuses and surveys and for the annual County Business Patterns Program, do not all have the same reference period. For large multiunit companies, industry codes for their establishments are updated annually in the Company Organization

Survey. Smaller multiunit companies are updated once between 5-year economic censuses. At the other end of the spectrum, industry codes for single-unit employers outside the scope of the economic censuses--such as those included in division H (finance, insurance, and real estate) and some industries in other divisions--and for those small employers who are within the scope but not included in the mail portion of the census will in most cases be the original codes assigned to them by SSA when they applied for EINs.

Like the SSA files, the SSEL includes codes indicating the source and date of the most recent update of the industry classification for each establishment. When codes are available to the Census Bureau from two or more sources, one is selected on the basis of priority rankings that have been assigned to different sources. Periodic tabulations of the SSEL showing the distribution of establishments and one or two key aggregates (such as employment and payroll) by date and source code can provide an indication of the potential effects of not having current industry classifications for all units. A tabulation of single-unit establishments by SIC division and source code appears in Table 20 (in Chapter 5).

In summary, most agencies use a 1-year reference period for the activity data on which industry classification is based, the exception being SSA, which asks for current activities with no defined reference period. Updating practices vary widely, both within and between agencies. As a result, differences in reference period are a factor in the lack of comparability between systems, except in those few cases where the codes of one agency are transferred to another and there are no subsequent unilateral changes. However, no studies have been done on the actual effects of this factor or on the optimum schedules for updating.

Treatment of Changes Over Time

When codes are reviewed and updated, another factor affecting comparability in industry classification is the use of rules that take into account, in addition to the activity data for the current reference period, the corresponding data and the codes assigned for one or more prior reference periods. These rules are incorporated into resistance coding procedures, which in general involve three steps:

(1) A tentative code is assigned, based solely on activities for the current reference period;

(2) The tentative code is compared with the most recent prior code for the same unit;

   (a) If the tentative code and prior code agree, the tenative code is accepted.

(b) If the tentative code differs from the prior code, the activity patterns for the current and one or more prior periods are examined. The most recent prior code is retained unless the changes from the prior period exceed some specified threshold, in which case the new code is accepted.

There are two kinds of arguments for the use of resistance coding. The first is that erratic shifts back and forth from one industry to another should be avoided, especially in those cases where a unit has roughly equal levels of activity in two or more SIC categories. This argument seems based on the assumption that shifts would be more misleading than informative to data users. A second and perhaps more defensible argument is that, in sample surveys of establishments or other types of business enterprises, such shifts can result in substantial increases in sampling variance for the industry groups affected. In such cases a classification bias may be preferable to a large increase in variance. Based on this second argument, industry codes for sample establishments (but not for those in the group from which all establishments are included) in the Annual Survey of Manufacturers are frozen between economic census years (Bureau of the Census, 1971).

Of course, resistance coding can only be used when the necessary codes or activity data for a prior period or periods are available during the current coding process. In general, it has not been used in the IRS coding systems, in which units are coded annually without reference to prior year data. An exception is the industry coding of large corporations in the SOI system; for these units a historical file is maintained manually, and unspecified resistance rules are applied.

Resistance coding principles are applied in several of the systems reviewed for this study. It was not possible to obtain detailed descriptions of the specific rules used in every system, partly because they are sometimes embedded in complex computer processing specifications[3] and partly because some aspects are left to the judgment of agency analysts. There appears to have been no serious effort to develop government-wide standards for resistance coding. The Census Bureau has stated that it will incorporate some capability for resistance-type coding in the SSEL system, but will treat the SIC code based on the most current data available for each establishment as the primary code for the system (Bureau of the Census, 1979:35).

Clearly this lack of uniformity in the use or nonuse of resistance principles and in the exact rules used when they are applied is another source of differences between systems.

A different method of moderating the effects of activity changes within an establishment or enterprise is to have a transition period during which the data (e.g., receipts, employment, payroll) for the unit are allocated, in tabulations, partly to the old industry and partly to the new industry. The proportion allocated to the new industry is gradually increased during the transition period. This procedure, which is called "wedging," is not currently used in any of

the systems studied; it was formerly used in the BLS unemployment insurance employment and wage (ES-202) units for which a gradual but expected permanent change had occurred. The length of the transition period used depended on the employment size of the reporting unit.

Treatment of Multiple Activities

If each business enterprise had only one activity or if all of its activities came under only one four-digit SIC industry, industry classification would be a much simpler process. This is not the case, however, and the principles and procedures followed in classifying units with more than one activity are a key factor in determining comparability between systems. Establishments and higher-level units pose somewhat different problems and are discussed separately.

Establishments

This discussion assumes that different systems have defined specific establishments in the same way and is concerned with differences in industry classification for identical establishments in two or more systems.

For establishments, the main question is what measure of the relative importance of different activities should be used? The SIC Manual (1972:12) is clear on this point: "Ideally, the principal product or service should be determined by its relative share of 'value added' at the establishment." Recognizing, however, that data for value added for each product or service are difficult to obtain, it recommends that the following data measures be used (SIC Manual, 1972:12):

Division	Data Measure
Agriculture, forestry, and fishing, hunting, and trapping (except agricultural services)	Value of production
Mining	Value of production
Construction	Value of production
Manufacturing	Value of production
Transportation, communication, electric, gas, and sanitary services	Value of receipts or revenues
Wholesale trade	Value of sales
Retail trade	Value of sales
Finance, insurance, and real estate	Value of receipts
Services (including agricultural services)	Value of receipts or revenues
Public administration	Employment or payroll

The recommendation is qualified in two ways. First, the SIC Manual states that these measures should be used "when available." Second,

it states: "In some instances, an industry classification based upon the recommended output measure will not represent adequately the relative economic importance of each of the varied activities carried on at such establishments. In such cases, employment or payroll information should be used to determine the primary activity of the establishments."

Once relative (or absolute) values of the measures have been obtained for each product or service, they are supposed to be aggregated by four-digit industry and the establishment coded to the industry with the largest share of the total, without regard to the shares of higher-level SIC categories (industry groups, major industries, or divisions).

To what extent are the SIC Manual recommendations followed? A review of the practices of the four major agencies whose systems were studied showed that none of them follows the SIC Manual in every respect.

BLS For all SIC divisions except division J (public administration), the source documents for industry coding ask for sales or receipts. The source document for government reporting units asks for employment or payroll.

Census Bureau According to the official description of industry coding procedures for the SSEL (Bureau of the Census, 1979), the recommended measures are used except in division C (construction), in which value of receipts is used in place of value of production and division D (manufacturing), in which value of shipments is used in place of value of production. It should be recognized, however, that the specified measures are not available on a current basis for some units in the SSEL, in particular, those that are out of scope of the economic censuses or are not included in the mail portion of the censuses.

IRS Taxpayers are asked to provide short descriptions of their "principal activity," which is generally defined in the instructions as the one accounting for the greatest proportion of sales or receipts. There are two exceptions to this general rule. First, the tax schedule for farm sole proprietors (Schedule F) contains entries for income (receipts) for each of several distinct crop and livestock items, so that a more objective basis is available for coding to industries within this division. Second, starting in tax year 1977, the instructions for the partnership tax return (Form 1065) have stated that the principal activity should be the one accounting for the largest proportion of assets. Before then, the standard instruction was to base principal activity on sales or receipts.

SSA Until recently, employers applying for an EIN were asked to report their "nature of business" without any reference to the treatment of multiple activities. The latest version of the application form asks for "nature of principal business activity." Multiunit employers who provide data for their separate establishments

or reporting units are asked to provide percentages of "gross dollar volume" corresponding to the principal activities of each one, listed in order of importance. The report form also asks for the number of employees engaged in each activity. In the coding process based on these reports, a manufacturing industry code is preferred over all others if the associated percentage is 20 percent or more.

Except for the SSA's special treatment of manufacturing just noted, all agencies assign the industry code for the category with the greatest share of activity, using data by four-digit SIC industry or the most detailed level contained in the system.[4]

One solution that has been proposed for the problem of coding multiple-activity establishments is to assign more than one industry code to establishments with more than one activity. The Census Bureau (1979) has developed but not yet implemented a proposal that the SSEL include secondary activity codes for each four-digit SIC activity with sales or receipts (depending on the industry division) of $100,000 or more. The record for the establishment would also carry a sales or receipts size class code corresponding to each activity code.

In the commercial sector, the Dun & Bradstreet Corporation maintains a large file of business enterprises for credit reporting and marketing purposes. The units in this file are coded by industry using the full SIC. Each unit with multiple activities may be assigned up to 6 SIC four-digit codes, subject to the constraint that the activity should account for 10 percent or more of total sales or "produce a large enough sales volume that it could stand alone as a separate business" (Dun & Bradstreet, n.d.). The main file includes the codes and the narrative descriptions of activities, with percentage of sales figures, from the source document.

Higher-Level Units

Business enterprises consisting of two or more establishments are more likely than single-establishment enterprises to have multiple activities. A large conglomerate, i.e., a corporation or family of corporations and their subsidiaries, may indeed have such a variety of activities that it becomes questionable as to whether there is any reasonable basis for assigning a single SIC or ESIC code to it. Morgenstern (1963) discussed this problem, using as an illustration the General Motors Corporation, which "produces motor cars, airplane engines, diesel locomotives, electrical appliances, heating equipment, etc." He concluded that some form of classification is necessary, but there is little hope of uniformity because no firm theoretical basis for such classification exists. In the more than 20 years since Morgenstern wrote, the situation has not changed.

There are two basic approaches to industry coding of business enterprises with two or more establishments. The first is to obtain activity information for the entity as a whole and to assign codes to the activity with the largest share (however measured). The ESIC Manual recommends that enterprises be classified at the four-digit

(most detailed) ESIC level, even if the full detail will not be used in analysis or publication. (The reasons for this recommendation are similar to those explained above in connection with the classification of establishments.) The second approach is to use the establishment SIC codes and measures of size, according to specific rules, to arrive at an enterprise code. This approach treats each component establishment as though it were engaged in only one SIC activity.

The two approaches lead to different results, as can be seen easily by turning to the example based on Table 2 above and identifying the subunits more specifically as establishments. (This example assumes that the SIC and ESIC categories are identical in this case.) If the first approach were used, i.e., classification based on enterprise totals, the enterprise would be assigned to category Y. If the second approach were used, establishments A and B would be assigned to category X and establishment C to category Y. Since the three establishments have equal measures of size, the enterprise would be assigned to category X.

This example has some basis in reality. The second approach described above has been proposed by the Census Bureau for use in the SSEL in assigning SIC codes to groups of establishments under a single EIN and in assigning ESIC codes to multiestablishment enterprises (Bureau of the Census, 1979:33-35). In both cases payroll would be used as a measure of size for establishments. The same method has already been used by the Census Bureau (1981b) for assigning industry codes to enterprises in connection with the enterprise statistics program which was part of the 1977 economic censuses. However, the classification of identical units in other major systems is based on aggregate activity data for the entity being classified. When this method is used, it is possible (although not very likely) for an enterprise to receive an industry code different from those assigned to any of its component establishments.

In coding enterprises, there are some departures from the ESIC recommendation to assign the enterprise to the single four-digit ESIC category accounting for the largest share of total activity (however measured). Some agencies use instead a top-down or filter-down approach. In general terms, this approach means first assigning the unit to the ESIC division accounting for the largest proportion of total activity, then assigning it to the next lower-level (two-digit) category accounting for the largest proportion of activity within that division, and so on. This approach can clearly lead to a different result than the ESIC Manual recommended procedure. Examples of this general approach are the Census Bureau's system for coding enterprises in the 1977 Enterprise Statistics Program and the Federal Trade Commission's system for coding consolidated corporations included in the sample for its Quarterly Financial Report Program.

The examples cited in this section demonstrate that there are many differences between systems in the treatment of units with multiple activities, both at the establishment and higher levels. Clearly, this is a factor that can lead to lack of comparability between systems.

Other Classification Principles

The above factors do not exhaust those aspects of classification principles on which agencies may differ. The SIC and the ESIC were developed to accommodate the full spectrum of U.S. economic activities, but there are many specific kinds of activities that present special classification problems: for example, the classification of central administrative offices and auxiliary establishments and the classification of units operated by governments. The SIC Manual provides some general guidelines for dealing with such problems, but several agencies have prepared procedures manuals and instructions that expand on these guidelines, as needed, to handle specific problems that arise in practice. Other agencies appear to rely almost entirely on the SIC Manual to resolve technical problems; in such cases, the more detailed principles needed to resolve specific cases must come from the individuals who code or provide technical assistance to coders, and those principles are not necessarily documented.

Resources and time available for this project did not allow a full review of procedures and instruction manuals used by different agencies. Probably some differences between agencies would turn up in such a review. One possible source of differences can be cited. As a general rule, central administrative offices and auxiliaries are to be classified to the four-digit industry representing the primary activity of the establishments they serve. In the BLS system, however, the state agency that does the coding will only have information in its own files about the activities of an employer that are carried on in that state. Contacts with one or more other state agencies may therefore be necessary to obtain the information needed to assign the correct code to a central administrative office or auxiliary. To the extent that this is not done, the code assigned could differ from that assigned by another agency, such as SSA, which has, in its multiunit employer system, information about an employer's activities in every state.

THE EFFECT OF SIC REVISIONS ON COMPARABILITY

The structure of U.S. industry is dynamic: new industries continually emerge and some existing ones may disappear or decline in importance. And new products, processes, and methods of doing business are introduced into existing industries. Consequently, revisions of the SIC Manual are considered necessary and are undertaken periodically so that it can continue to reflect the existing industrial structure of the economy. Major revisions were published in 1957 and 1972. Most of the work was completed for a revision originally scheduled for 1982, but the revision was postponed because funds were not available for its implementation by the federal and state statistical units most directly affected (Federal Statistics User's Conference, 1982). A revision is now scheduled for 1987.

The principles and procedures for review and revision of the SIC Manual are described in detail elsewhere (see, e.g., Office of Federal Statistical Policy and Standards, 1981) and so are not discussed at length here. The changes that are made in a revision are mostly of the following kinds:

Splitting an existing (four-digit) industry to form two or more new ones;
Combining two or more industries to form a new one;
Shifting an industry, without changing its description, to a different (three-digit) industry group;
Transferring specific activities from one industry to another.

These changes affect comparability of industry data in two ways: they affect comparability over time from the same system, and they affect comparability between systems.

Comparability Over Time

Some loss of comparability over time as a result of SIC revisions is inevitable and is the price of keeping the classification system up to date. Opinions differ, of course, on the optimum frequency and scope for revisions and on how to minimize their effects on comparability.

It is possible for abrupt and somewhat artificial-appearing changes to result from a revision. To take, for example, what may be a rare case, suppose that an establishment has 60 percent of its receipts in major group A and 40 percent in major group B and that the receipts come from a single industry in each major group. If the industry in major group A is split into two industries and the establishment has 30 percent of its receipts in each, it would then be classified in the industry in major group B if that industry had not been touched by the revision.[5]

The agencies that publish periodic statistics by industry try to assist users who are interested in time-series analyses by using various bridging procedures and by publishing special analyses of the effects of SIC revisions. (For an example of the latter, see Bureau of the Census, 1972b.) To the extent possible, data are published for at least a short period based on both the old and new classifications. In its description of plans for the SSEL, the Census Bureau (1979:36) has indicated its intention to retain both old and new industry codes for establishments during transition periods.

Although useful, bridging procedures are expensive and in some systems are severely limited by the data available. Consider first the situation when a system classifies units to full four-digit SIC detail. Shifts and combinations are easy to deal with, but splits and transfers are not. The correct new codes for prior periods can be assigned only if the necessary data on products, services and other relevant factors are already available for those prior periods in the system. The correct assignment of old and new codes for periods

following the revision can be made only if the same kinds of data are obtained on a current basis. In general, these requirements cannot be satisfied if the only information obtained on activities concerns each establishment's principal or main activity, as is the case for the SSA single-unit and some IRS systems, or when self-coding is used, as in the other IRS systems.

Grouping of SIC industries in a system (see above) further complicates matters. Conversion to the new codes when combinations or shifts occur is no longer a simple transformation in all cases: if they involve more than one of the groups used in the system, additional information will be needed.

Comparability Between Systems

All of the agencies whose systems were reviewed have conformed to past SIC revisions and presumably expect to do so in the future. However, revisions are likely to lessen the comparability between systems, especially during transition periods. The timing of the conversion from old to new codes cannot be exactly the same in all systems. Revisions, as a rule, are timed to coincide with quinquennial economic censuses. Agencies that code annually, such as IRS, have relatively little difficulty in making the conversion at about the same time. The BLS, which uses a 3-year updating cycle staggered by SIC division, would have to undertake a special update, at least for those industries affected by the revision, in order to do a full conversion at about the same time. The changes resulting from the 1972 SIC revision were not made in the BLS employment and wage system until 1975. The SSA has a special problem, since its only comprehensive updating procedure depends on a match against Census Bureau records. The matching procedure is costly, and it is not fully effective since there is always a significant residue of unmatched cases in the files of both agencies. The procedure was last carried out for single-unit employers following the 1972 economic censuses and was carried out successfully only once, following the 1957 economic censuses, for multiunit employers. Thus, the only fully satisfactory alternative available to SSA for those categories involved in splits or transfers would be to contact the employers at the time of the revision to get the information needed to assign the new codes.

There can be little doubt that revisions do diminish the comparability of industry classification in different systems. The extent of problems created for users depends on what kinds of resources are made available to the statistical agencies to expedite their conversions to the new codes and to do special tabulations showing the effects of the revision on major time series based on data by industry. Users like the Bureau of Economic Analysis, which uses data by industry from many different agencies in compiling the national income and product accounts, are likely to have the greatest difficulties. One BEA employee told the author that he had been through five SIC revisions and that there had never been sufficient

resources allocated to minimize their effects on the continuity of time series.

INTERPRETATION OF THE SIC MANUAL

The 1972 SIC Manual is largely self-contained in the sense that most units supplying a reasonable amount of information about their activities can be classified unambiguously to an industry (four-digit) by reference to the industry descriptions and other materials included in the manual. However, it would be unreasonable to expect the SIC Manual to be sufficiently detailed to cover every possible situation. And even if it were possible at the time of each revision, new activities would develop between revisions. Therefore, the agencies that base their coding systems on the SIC Manual have developed procedures and supplementary instruction materials to deal with situations that are not unambiguously covered in it. For manual coding systems, some of the agencies provide instruction manuals to supplement the SIC Manual, and these in turn may be supplemented from time to time by memoranda covering the assignment of codes to activities encountered for the first time. In automated coding systems, the specifications for computer programming may sometimes deal with situations not specifically covered in the SIC Manual.

The interpretations and extensions of the SIC Manual are being made by several different agencies and sometimes for different systems within an agency. Although some efforts have been made in the past to coordinate what the agencies do in this respect, there is no mechanism that guarantees uniformity. In recent years, the Statistical Policy Office of the Office of Management and Budget (OMB) and its predecessors have not had the resources for a systematic coordination effort.

It has not been possible, in preparing this report, to do a thorough study of how different agencies handle the classification of units for which the SIC Manual does not provide sufficient guidance. (Ideally, such a study would be based on a match of units in different systems and comparison of their SIC codes.) However, conversations with some of the classification experts in the agencies have made it clear that much more could be done to promote uniformity in the treatment of those units. One example noted was that of dinner theaters, an activity not specifically mentioned in the 1972 SIC Manual. One major system classifies them in retail trade, major group 58 (eating and drinking places); another puts them in services, major group 79 (amusement and recreation services). A similar problem arose prior to the 1972 SIC revision with respect to bowling alleys that also served food or beverages. The Census Bureau wanted to classify such establishments on the basis of the SIC activity that accounted for the largest share of receipts, but the Social Security Administration wanted to call them all bowling alleys. The SSA won out in this case, and the definition of bowling alleys (industry 7933) in the 1972 SIC Manual notes that such establishments also frequently sell meals and refreshments.

NOTES

1. This illustration is used again below in connection with the treatment of multiple activities.
2. The reporting unit definitions provided in Appendix A are the ones that were available from the two agencies when this report was written. The BLS is considering changes in its definition, which would bring it closer to the SIC establishment concept.
3. For an example of specific rules used in the Annual Survey of Manufactures, see Bureau of the Census (1971:60-61); those rules are not necessarily identical to the ones now used.
4. Some exceptions to this general rule are built into the SIC for industries in which certain kinds of related activities are frequently carried on in the same establishment. One example occurs in major industry 56, retail apparel and accessory stores. An establishment that sells men's clothing and furnishings (5611), women's clothing (5621), and children's clothing (5641) is classified as a family clothing store (5651) if none of the 3 categories accounts for 50 percent or more of total sales. Other industries with similar definitions are wholesalers of groceries, general line (5141) and general auto repair (7538). Some major industries have been established to cover establishments that sell a wide variety of products or provide several related services, for example, major industry 53, general merchandise stores, and major industry 66, combination of real estate, insurance, loans, law offices.
5. This example assumes classification is based directly on the principal four-digit activity; if a top-down approach were used, the establishment would remain in major group A.

CHAPTER 3

SOURCE DATA

INTRODUCTION--ACCURACY OF CODES

The previous chapter dealt mostly with differences resulting from system features deliberately adopted to meet the particular requirements of an agency. The agencies recognize (in most cases) and accept that those features will cause some of their industry codes to differ from those in other systems. The differences are considered necessary to meet their particular data system requirements. It cannot be said that such differences result from codes in one system being more or less accurate than those in another system.

Accuracy of industry codes can be defined in various ways. In this report, a code is considered inaccurate or incorrect if it is not the code that should have been assigned to a unit given the definitions and classification principles adopted for the particular system. Operationally, the accuracy of particular codes in a particular system may often be difficult or impossible to determine. This is especially true in those systems for which the documentation of these definitions and principles is inadequate. In such cases the SIC Manual (or the ESIC Manual) is the only basis for judging what is correct.

By this definition, a code is not inaccurate, nor are the data to which it is applied inaccurate, simply because the code and the data have different reference periods. If the system specifications for updating have been observed and if respondents have provided activity and other data for the correct reference period, the code is not considered incorrect even though activity patterns may change in subsequent periods.

A major cause of inaccuracy in industry coding is the use of incomplete or incorrect source data to determine the code. To understand this situation fully, it is first necessary to know what data are needed to assign correct codes and then what data are actually available and used in different systems.

DATA NEEDED FOR CODING

What information is needed to assign the correct SIC four-digit industry code to a particular establishment or other unit? One simple

answer would be the correct code for that unit taken from some other data system. This is not meant to be facetious; almost all agencies do obtain some of their codes from other agency sources or from commercial lists. The accuracy of those codes depends first on whether the unit to be coded has been correctly identified in the other system and secondly on whether the code assigned to the unit in the other system is "correct," as defined for the system to which the code is to be transferred. However, all of the codes taken from other systems must have been based at some point on direct information about the units coded, so the question remains, what specific kinds of information are needed?

To answer the question, it is necessary to turn to the 1972 SIC Manual. While a full analysis of all of the kinds of information needed would require a lengthy dissertation, certain general patterns can be discerned.

It is assumed here that all questions about how many establishments exist at a particular location have been resolved and that the only issue is what information to obtain for each establishment in order to assign a four-digit SIC code. It is also assumed that the information will apply to the reference period required by the coding system except that, if resistance coding is to be used (see above), the same information for one or more prior periods (or at least the code based on that information) will also be needed as input to the coding process.

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. In the former case, the establishment is classified as a central administrative office or auxiliary and it is necessary to know the primary activity of the establishments it serves. Some products or facilities are leased or rented rather than sold, in which case 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. Or, for feed lots, if they do not buy,

sell, or auction livestock, it is necessary to know for how long the cattle are fed. For banks, it is necessary to know whether they are members of the Federal Reserve System and whether their deposits are insured by the Federal Deposit Insurance Corporation. For other financial institutions, classification depends on their type of charter and whether they are members of the Federal Home Loan Bank Corporation.

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. Activity measures must be reported as or converted to percents or proportions to apply the various coding principles that apply to such establishments. 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 (for an example, see the comment on family clothing stores in footnote 3, Chapter 2, above).

To give a better idea of the kinds of questions about products and services that must be answered to assign the correct industry classification, the rest of this section presents selected examples taken primarily from the BLS Handbook (Bureau of Labor Statistics, 1974).

What Raw Materials Are Used?

The manufacture of gloves is classified in different industries depending on whether they are made of leather, plastic, or rubber.

The manufacture of food containers is classified in several different industries depending on whether the raw materials are glass, plastic, metal, pressed or molded pulp, or special food board.

The classification of the manufacture of butadiene depends on whether it is produced from petroleum or from alcohol.

Where Do the Raw Materials Come From?

The classification of the manufacture of asphalt board depends on whether the raw materials are purchased or made in the same establishment.

The classification of the manufacture of glass products made from purchased glass differs from that of products made from glass produced in the same establishment.

Who Uses the Product?

The classification for many types of electrical equipment, such as cooking equipment, dishwashers, and refrigerators, depends on whether they are intended for household or commercial/industrial use.

Is the Product Mass Produced or Custom Made?

The production of ready-to-wear clothing is classified in manufacturing while custom tailoring of clothing is classified in retail trade or services, depending on who supplies the materials.

What Process is Used?

The classification of the "production" of fish depends on whether they are raised commercially or caught from their natural environments.

The classification of printing establishments depends on the process used, e.g., letterpress, lithography, gravure, or screen.

Where Does Production Occur?

The production of vegetables grown in the open air is classified differently than the production of vegetables grown under cover.

The classification of the production of several petroleum products, such as aromatic chemicals, benzene, and naphtha, depends on whether they are produced in petroleum refineries or as a product of coal-tar distillation.

Is the Product Sold New or Used?

The retail sales of automobile parts and accessories is classified in different industries depending on whether they are new or used.

The retail sale of new household goods, such as furniture and home furnishings, is generally included in SIC major group 57 (furniture, home furnishings, and equipment stores). The sale of similar items when used or antique is generally included in major group 59 (miscellaneous retail).

What Is the Method of Selling?

In the retail trade division, a separate industry group has been established for nonstore retailers so that sales of products by mail order, vending machine or door to door are classified differently than sales of the same products from stores.

Producers of baked goods who sell primarily on the premises at retail are classified as retail (5462); those who produce for home service delivery or sale at one or more nonbaking outlets are classified in manufacturing (2051).

What Type of Rental or Leasing Arrangement is Used?

The leasing, rental and time-sharing of computers are classified in several industries: leasing of computer time (7374); leasing of equipment directly by manufacturer (3573); leasing by sales office of manufacturer (5081); finance (equity) leasing of computers (6159); and other rental of computer equipment (7379).

SOURCES OF DATA

General Considerations

The previous section enumerated the kinds of information needed in order to code establishments to the SIC four-digit industry level. Systems that do not code to this level of detail do not necessarily require all of this information; some of the Census Bureau systems, which are more detailed than the SIC, require more information.

It has already been pointed out that most systems obtain some of their codes from other agencies or commercial sources. Lacking access to the SSEL, most agencies prefer to do their own coding with current information obtained directly from the establishments (or other units) involved. The rationale for this may be that codes that are available from other systems are not sufficiently accurate or up to date or that the coverage, definitions, and coding principles of the other systems are not fully compatible. And in many cases, code transfers from one agency to another are prohibited by law.

Thus, direct transfer of codes from another system is used at present mainly as a fallback procedure or as a first step in a more elaborate coding process. In the 5-year economic censuses, for example, for small establishments that are not sent questionnaires (in general, those with no employees or a small number of employees), the codes used are those obtained from the SSA or IRS systems. For larger establishments not previously contacted by the Census Bureau, the SSA or IRS code (if available) may be used for a preliminary classification that determines the type of economic census questionnaire to be mailed. The responses to these questionnaires (with follow-ups in many cases) provide the information that the Census Bureau believes it needs to

assign a definitive Industry and Product Classification (IPC) code. In many systems, when information from source documents is missing or incomplete, reference is made to commercial directories, such as those available from Dun & Bradstreet, Moody's, and Standard and Poor's. The potential for additional use of code transfers is discussed further in Chapter 6. The rest of this section discusses the kinds of source documents and other direct approaches used to obtain classification information directly from business enterprises.

Concepts of incomplete data and nonresponse for industry coding are not very clear cut. The absence of complete data on a source document can be remedied by follow-up procedures. If these fail or cannot be afforded, recourse to commercial lists is possible as a last resort. In updating codes, if none of these sources is available, an alternative may be simply to retain the code already in the system for that unit.

It is quite difficult and has not proved feasible in this review to obtain quantitative data on the relative frequencies with which codes in various systems are assigned from an original source document or by the other methods described above. It is sometimes possible, often from publications, to obtain data on the number and relative importance of units in a system that either could not be classified at all or could not be classified to the level of detail called for by that system. Some data of this kind are presented in the last section of Chapter 5.

The source documents used for the different coding systems reviewed for this report show wide variation. Aside from the fact that no standards for such documents have ever been established, there are many reasons for this. In designing source documents, an agency must consider:

The kinds of information needed to code to the level of detail called for by the system. (An agency may decide to code to less detail because it does not feel it can reasonably burden respondents with the task of completing a source document that would provide all of the information needed to code to the SIC four-digit industry level.)

Whether a single document will suffice for all sectors or whether there should be variations for different SIC divisions or other groupings. Source documents tailored to the requirements of particular groups of industries should, in general, lead to more accurate classification, although errors may occur when the wrong version of a document is sent to a particular unit. The use of tailored documents will, of course, increase the cost of document design, printing, and processing.

Whether all information should be explicitly requested on the initial source document(s) or whether some details, especially those needed less frequently, should be obtained through follow-up inquiries.

The particular combination of manual and computer operations by which the codes will be determined from the source information.

How much and what kinds of instruction materials should accompany the source documents. (Virtually all source documents are designed for mailing and self-administration by business enterprises; direct interview is reserved for a few very large units and for some follow-ups.)

In summary, the design of source documents cannot be properly done in isolation; each one must be regarded as a component part of a system for industry coding.

EXAMPLES OF SOURCE DOCUMENTS

A somewhat arbitrary classification of source documents into four categories has been made, and this section presents examples for each category. The categories are shown below:

Category	Coding by	Level of Source Information Detail
A	Respondent	Not applicable
B	Agency	Low
C	Agency	Medium
D	Agency	High

The forms and, in some cases, the relevant instructions for them appear in Appendix B.

Category A (self-coded) The only systems that use self-coding, i.e., coding by respondents, are the IRS revenue processing systems for partnerships and corporations.[1] The source documents are the appropriate tax return forms for those two categories of taxpayers. The relevant data items and instructions from the partnership return (IRS Form 1065) for tax year 1981 are shown as Exhibit B-1. The "Business Code Number" is to be entered by the taxpayer in item C on the first page, using the instructions and code list on page 12 of the instructions. The code list provides a short description of the industry or group of industries corresponding to each code used by IRS. Taxpayers are also asked to give a brief description of their principal business activity and principal product or service in items A and B, respectively. This information is used very little in revenue processing, but is used for the industry coding for the Statistics of Income Program, which is based on a sample of returns. A peculiar feature of self-coding is the potential for a high proportion of incorrect codes immediately following a revision of the SIC; some evidence on this topic is presented in Chapter 5.

Category B (agency coded, low detail) The example for this category is also taken from IRS. Exhibit B-2 shows the relevant data items and instructions from the 1981 tax return schedule used for nonfarm sole proprietorships (IRS Form 1040, Schedule C). The primary data items used for coding are item A, a two-part item calling for brief descriptions of the "main business activity" and its "product" and item B, the business name. The instruction for item A is to "Report the business activity that accounted for the most income. . . . Give the general field as well as the product or service. For example, 'wholesale--groceries' or 'retail--hardware'."

For some returns, additional clues to the correct classification may be found by examining other parts of the return, e.g., the kinds of expenses (deductions) reported in Part II and the kinds of property listed in Schedule C-2, depreciation. Note, however, that taxpayers are not required to show a breakdown of receipts or sales by source, so there is no way even to check that the main activity has been properly identified, let alone to apply the more complex rules that apply to some combinations of activities. It should be noted that IRS Form 1040, Schedule F and Form 4385, which are used for farm sole proprietorships, do require a breakdown of sales or income from different kinds of crop and livestock production; this is sufficient, in the author's judgment, to put these source documents in Category D.

Other source documents that provide a low level of input detail are certain ones used by the Census Bureau as a preliminary to more precise coding based on the economic censuses or current surveys.

Category C (agency coded, medium detail) The main example for this category is Form SS-4, application for an EIN, which is used by SSA to classify all employers for the single-unit employer file.[2] Codes for establishments or reporting units of multiunit employers are based on a more detailed form that is sent to eligible employers following receipt of the initial application. The complete Form SS-4 and the relevant section of the instructions for it appear as Exhibit B-3. The primary data item used for industry classification is item 14, nature of principal business activity. The instructions for this item give examples of the kinds of descriptions desired for various SIC divisions. Several other items may assist in classification:

Item 1, name
Item 4, trade name
Item 10, type of organization
Item 16, breakdown of employees by type
Item 17, for manufacturers, principal product and raw material used
Item 18, to whom does the employer sell most of his or her products or services.

These items, especially 17 and 18, cover certain of the key data requirements needed for classification that are not covered in the Category B example.

The Form SS-4 was classified in the medium rather than high-detail category primarily because it does not provide any breakdown of

multiple activities. Until recently, the data item and instructions did not even specify that the main or principal activity should be reported, although it might have been regarded as implicit. Several earlier versions of the SS-4 did include an item asking manufacturers to list their three principal products and to give the percentage of total value of products represented by each of these.

Category D (agency coded, high detail) Within this category, the amount of detail and the general approaches used vary, so it is useful to give more than one example.

Probably the source documents that provide the most information for industry coding are the mail questionnaires used in the quinquennial economic censuses. These questionnaires are tailored to different groups of SIC industries and so can and do include the specialized inquiries needed to assign industry codes within those groups. Special procedures are, of course, needed to process questionnaires that turn out to have been inappropriate for the establishments to which they were sent.

Exhibit B-4 shows the questionnaire for the 1982 census of retail trade: tires, batteries, parts, accessories, (Form CB-5502). This questionnaire was mailed to establishments believed to be in Census Bureau IPC categories 553110 (tire, battery, and accessory dealers) and 553120 (other auto and home supply stores). The "mailout" code, i.e., the latest IPC code for that unit from the SSEL, is imprinted on the mailing label. When the filled questionnaire is received, a "self-designated" code is determined on the basis of the respondent's entry in item 9, kind of business. Normally, the final IPC code is assigned by computer, based primarily on the merchandise lines data (item 11), but also taking into account other relevant items on the form, including dollar volume of business (item 5), class of customer (item 7), method of selling (item 10) and a specific inquiry on sales and receipts from retreading tires (item 12a). The mailout and self-designated codes enter into the final IPC code determination only if the data for the items normally used are incomplete or ambiguous.

Another form that provides a high level of detailed information for industry coding is the BLS Form 3023, which is used for updating industry codes (also area and type of organization code) of units covered by the Unemployment Insurance Employment and Wages (ES-202) Program. There are separate versions of this form for most SIC divisions; wholesale and retail trade have the same version, and there are separate forms within manufacturing for durable and nondurable goods. There is also an "all industries" version.

Exhibit B-5 shows BLS Form 3023-A7 (Rev. Dec. 1982), used to update industry codes for reporting units currently classified in wholesale or retail trade. Unlike other examples discussed in this section, this form is designed primarily to get the information needed for industry classification of the reporting unit. The key items on the form for this purpose are items 1, 4, and 5 in part 1. Item 1 covers the identification of multiple products or activities of the reporting unit and the percent of total sales (value or receipts) accounted for by each during the most recent calendar year. The instruction page

includes a lengthy list of examples of the kinds of detail wanted in item 1, with each example covering activities within a single SIC four-digit industry. Item 4 identifies central administrative offices and auxiliary units, and item 5 asks for the principal class of customer, as an aid to determining whether the unit is wholesale or retail.[3]

Recently, the BLS has developed and tested a verification method of updating industry codes (Hostetter, 1983). A form similar to the Form 3023 is sent to each reporting unit; however, each form includes a brief preprinted description, adapted from the SIC Manual, of the main activities covered by the four-digit industry in which the unit was previously classified. If respondents consider the description correct, they may so indicate and are not required to provide current information on principal products or activities. Tests have shown that use of the new verification form significantly reduces respondent burden and coding costs, but limited data on quality effects reported by Hofstetter suggest that a substantial proportion--perhaps one-third--of the actual changes in classification may be missed when this method is used. Further tests are under way, and various refinements to the general approach are being developed.

A final example in this category comes from the Federal Trade Commission (FTC) Quarterly Financial Report (QFR) Program.[4] Exhibit A-6 shows FTC Form 59-103 (rev. Oct. 1979), nature of business report. The FTC uses two versions of this form: the one shown, which is for the manufacturing division, and a second version that is for the other SIC divisions included in the QFR Program (mining, wholesale trade, and retail trade). The nature of business report is sent to all corporations that are about to enter the QFR sample for initial determination of status, and, for updating purposes, to certain corporations reentering or remaining in the sample. Like the BLS Form 3023, its primary purpose is to classify reporting units by industry. In addition, several questions are asked to determine the current corporate structure of the reporting unit.

The key item on the form is item 3, in which the respondent is asked to list products made, processed, or assembled and/or sold, with the percent share of gross receipts accounted for by each. In addition, information is requested on kinds of raw materials and processes used in production. Unlike the BLS form, this form does not provide any illustration of the level of detail desired in distinguishing different product categories.

This review of the sources of information used for industry classification, particularly the source documents, while by no means exhaustive, illustrates the wide variability of the inputs used in different systems. Because of this variability, one would expect to find differences in the codes assigned to the same units in different systems. Other things being equal, one would expect to find a positive association between the amount of input information available and the accuracy of the codes assigned. Many of the differences in source documents can be justified by differences in system requirements and the resources available for industry classification;

nevertheless, it should be possible to achieve a greater degree of standardization than now exists, especially for systems that have comparable requirements and resources. To complete the analysis of why differences occur, one more aspect of the industry classification and coding process must be examined, namely, how the source data for each unit are converted to a code appearing in a computer record for that unit (this step is represented by box E in Figure 1, in Chapter 1); this final phase is discussed in the next chapter.

NOTES

1. Some of the forms used in the Bureau of Economic Analysis (BEA) direct investment statistics program ask respondents to enter sales and associated three-digit codes for up to eight "direct investment industry classifications" under which they have sales. The codes are listed on the form, and additional detail is provided in the Direct Investment Industry and Foreign Trade Classification Booklet. However, the overall codes for the units are determined and entered on the forms by BEA coders.
2. The SS-4 is an IRS form, but the industry coding is done by SSA.
3. Comparing item 5 on this form with item 10, class of customer, of the economic census form shown in Exhibit B-4, the latter would appear to provide a more reliable basis for distinguishing between wholesale and retail activity.
4. The QFR Program was transferred to the Bureau of the Census in October 1982.

CHAPTER 4

CODING PROCEDURES

INTRODUCTION

The description in this chapter of coding procedures, i.e., the methods used to produce final codes from source data, follows and expands somewhat on the treatment of this topic by Farrell et al. (1982). It does not include a detailed description of the procedures used in each major industry coding system, which would be impossible to do without much more documentation than is now available to the author and without observing actual coding operations in each system. Rather, the purpose of this chapter is to enumerate the kinds of procedures available and currently in use for industry coding, and to give some examples. This enumeration will make it clear that procedures vary widely and that this diversity is likely to contribute significantly to lack of comparability between systems. The reader is also invited to speculate on the potential relative accuracy of the different approaches described; such limited evidence as is available on this topic is presented in Chapter 5.

The next section of this chapter covers manual procedures, such as manual entry of industry codes on source documents and manual data entry operations. The third section covers automated procedures, such as initial determination of industry codes from source data and computer consistency checks. Actually, the distinction is rather arbitrary. "Manual" procedures may be computer assisted, and a certain proportion of units processed by computer may require human intervention to correct errors that have been detected or to deal with problems resulting from incomplete or inconsistent data. Most systems are, in fact, mixed. The treatment of incomplete data, which has already been discussed to some extent, is reviewed in the last section.

MANUAL PROCEDURES

Coding

Most of the systems reviewed for this report use manual coding: i.e., the code for each unit is determined by a coder, based on information available for that unit from the source document and

(sometimes) from other sources, and is then recorded, usually on the source document, for data entry. In some systems, when a code is being updated, prior data or the prior code for the unit may be part of the information available to the coder. Such prior information may be used in a formal sense in the determination of the current code (as in the case of resistance coding, discussed in Chapter 2), or it may serve merely to alert the coder to be especially careful when the proposed current code differs from the prior code.

The nature and complexity of the coding process will, of course, depend on the kind of information available to the coder. As described in Chapter 3, that information can vary from a brief description of each unit's principal activity to a quantitative breakdown of the unit's receipts from various products or services, supplemented by other relevant information on the nature of its activities. In the former case, the coder's task is, in a sense, simpler, but the potential for accurate coding is much lower.

In recent years, the Census Bureau has introduced computer-assisted manual coding procedures in some of its industry coding operations (Farrell et al., 1983). For the 1982 economic censuses, nearly 1 million reports and unclassified administrative records were assigned IPC industry codes by this method. The coder in this type of operation works at an interactive computer terminal and is presented with name information or a brief activity description for each unit to be coded. The coder attempts to determine a keyword that describes the unit's probable primary activity from the name or activity information. When only a name is available, the keyword may be part of the name, e.g., "record" from Smith's Record Shop, or it may be imputed from the name, e.g., "automobile" from Davis Motors, Inc. The coder keys the keyword and a code for the SIC division in which he or she believes the unit belongs. The terminal then displays a code and a short description for each IPC category, in the SIC division selected, that contains a match to the keyword. The coder decides which of these categories is appropriate and keys the line number for the selected category.

The process may fail at any of several stages: the coder may not be able to identify a keyword from the name or activity information or may not be able to decide which SIC division is appropriate; the reference file may not contain the keyword chosen; it may not contain any IPC categories that appear to apply to the unit; etc. In such cases, other methods of classification must be used.

Factors Affecting the Quality of Manual Coding

At the beginning of Chapter 3, "accuracy" was defined by assuming the existence of a correct industry code for each unit given the definitions and classification principles adopted for the particular coding system. Errors occur for two reasons. First, they may be inherent in the system, because, e.g., certain necessary information is not made available to the coders or because the coders are given ambiguous, incomplete, or incorrect instructions for assigning codes

from the information available to them. Second, coders may commit errors by failing to carry out some aspect of the coding procedure correctly, or they may follow the instructions fully but make careless errors in recording codes.

With respect to the first class of errors, which could be called system errors, Chapter 3 described the wide variation in the completeness of the source information available to coders in different systems, which results primarily from differences in the basic source documents. There is also wide variation in the kinds of instructions and reference materials given to coders. In nearly all systems, the coder has access to the latest SIC Manual and a list of the code categories to be used for that system. Generally, there are also some index lists, i.e., alphabetical lists of activities with the corresponding system codes. Beyond these, some manual systems have virtually no additional instruction materials (IRS is one example), while others provide a substantial amount.

An example of a system with additional materials is the BLS ES-202 system, which provides a separate Handbook on Standard Industrial Classification (Bureau of Labor Statistics, 1974) as well as various program memoranda dealing with specific issues, such as code changes. The Bureau of Labor Statistics (n.d.) has also recently developed an SIC Coding Workbook for use by employees (generally those in state employment security agencies who are responsible for industry coding) "with little or no experience in assigning SIC codes" or "who wish to review the basic concepts of coding." The Workbook has numerous exercises and can be used for independent study or in formal training sessions. In the private sector, Dun & Bradstreet (n.d.) has developed a self-instruction programmed-learning course, Assigning SIC Numbers, for the "reporters" who assign SIC codes to the units included in their system.

With respect to the second class of errors, which could be called coder errors, several factors can influence the level of error in a system, including the methods of selecting and training coders, the grade level associated with the work, and the extent of specialization: i.e., do the coders work full-time on industry coding, or is it a sporadic activity or one step in a more extensive manual processing operation? For the SSA systems, industry coding is a full-time job; in the IRS systems, industry coding is generally only one element of a fairly extensive tax-return processing operation.[1]

A key factor in coder errors is the kind of formal or informal quality assurance systems used to measure and control them. The large-volume manual coding systems, such as those of SSA, IRS, and some Census Bureau systems, have formal quality-control systems, usually involving 100 percent verification of the work of inexperienced coders and sample verification for others. SSA, for example, uses two types of reviews of industry coding for its single-unit employer identification system. Within the coding branch, a sample of each coder's work is reviewed by other coders (peer review), and the differences found are reviewed by technicians before the work is returned to the original coders for correction. In addition, a weekly subsample of 1,000 cases already subjected to peer

review is audited by the SSA Office of Research and Statistics. Based on the results of these audits, technical memoranda are prepared and training sessions conducted to improve the quality and consistency of coding. Smaller systems, involving only a few coders, generally do spot checking on a less systematic basis.

All of the quality control systems reported on for this report use dependent verification, i.e., the verifier or reviewer has access to the code entered on the source document by the original coder. It has been demonstrated in various contexts that dependent, in contrast to independent, verification often tends to understate the actual level of error.

The BLS has a special problem with regard to coder errors because most of the coding is done in the 53 state employment security agencies (SESAs), and they have a certain amount of leeway in the structuring of their SIC coding operations. Some information about these operations is collected annually from each SESA on an "ES-202 State Operations Review" questionnaire. One matter of special concern to BLS now is to ensure that the ES-202 coding system makes full use of industry coding information collected in related BLS programs, such as current employment statistics, and occupational safety and health statistics. However, no information was provided to the Industry Coding Working Group about the nature or results of quality control systems used by the SESA's in the ES-202 manual coding operations.

Data Entry

The final manual operation--not counting manual interventions required for some units in automated processing steps--is data entry. Although the system descriptions available for this study do not specify the particular equipment used, it seems likely that key-to-disk or key-to-tape data entry systems are being used in nearly all cases, so that some simple edits can be made as part of the process.

If the industry code for the unit has first been assigned manually, then a primary concern will be that it has been correctly keyed. This is especially true when few or no consistency checks involving the industry code are possible. If the industry code has not been assigned manually, it is important that all of the items that will be used for automated coding (mostly codes identifying specific activities and data on their shares of total sales and receipts) be correctly keyed. As is usually the case in multistage data-processing activities, there are interesting trade-offs to be considered with respect to those alternatives: Does the presumptive greater reliability of a computer in applying industry coding rules to a data set offset the higher probability of error in keying a set of items as opposed to a single industry code? There are no simple answers to this and similar questions and little relevant empirical data. System designers appear to operate mostly by intuition in evaluating these trade-offs.

AUTOMATED PROCEDURES

Coding

Among the major agency systems, the only ones currently using primarily automated industry coding procedures are the IRS's SOI systems and some of the Census Bureau systems, in particular, those used for the mail portion of the economic censuses, for farms with sales of $2,500 or more in the 1974 census of agriculture, and for certain periodic surveys, such as the annual survey of manufactures.

Statistics of Income

Since 1981, SOI industry coding has been largely automated, with manual coding on an exception basis (Powell and Stubbs, 1981; Cys et al., 1982). For sole proprietorships, the current year revenue processing code is accepted as the SOI code if it is a valid industry code (other than "not allocable"). If there is no revenue processing code or an invalid or "not allocable" code, the SOI code is determined manually. The automated coding process for partnerships and corporations makes use of the prior year's SOI and revenue processing industry codes as well as the current year revenue processing code. If the current and prior year revenue processing codes agree, the prior year SOI industry code is accepted for the current year. If they differ, the SOI code for the current year is determined manually. If prior year codes are not available, a valid current year revenue processing code is accepted, except for taxpayers in certain industries and large corporations.

Economic Censuses

A detailed description of the Census Bureau's automated coding systems is beyond the scope of this report. They are complex, and there are many variations, depending on the particular SIC division and industry groups involved. However, a look at the sample form from the 1982 economic censuses (Exhibit A-4) gives a general idea of how it works. Item 9 asks the respondent to identify the principal kind of business for the establishment in 1982. This particular form is mailed to establishments classified, prior to the Census Bureau mailing, in SIC industry 5531, auto and home supply stores, which are covered by the first three categories in item 9. The remaining specified activities in item 9 are those known to be commonly associated with this industry. Those not specified in item 9 may be written in the space at the bottom of item 9. The IPC code based on item 9 is known as the self-designated industry code.

A key item in determining the industry code is the merchandise lines inquiry, item 11, in which the respondent is asked to enter amounts and percents of sales for each merchandise line specified. The merchandise lines in parts 1a to 1g are those associated with SIC

industry 5531. Other items that may enter into the final IPC code determination are the mailout code (which appears on the mailing label), item 5 (dollar volume), item 7 (class of customer), item 10 (method of selling), and item 12a, sales and receipts from retreading tires.

The computerized determination of the industry code follows a generalized edit to ensure that the data for the establishment are reasonably complete and internally consistent, e.g., that sales by merchandise lines sum to the reported total. The specifications for coding are complex, consisting of several pages of decision logic tables. They start with some general routines to determine whether the establishment is, in fact, retail and if so, whether it is a department store, other store, or nonstore retail operation. If it is in the other store category, it is sent to an edit routine specific to the particular questionnaire used (in this case, the one for industry 5531). The routine examines the merchandise line and other relevant data. If appropriate, it assigns one of the two IPC codes included in this SIC industry; if not, the record is routed to a generalized industry coding routine that assigns a code based on these same items. The final code is usually based on this computerized determination. The mailing and self-designated codes come into play only when the merchandise line or other data are incomplete or, for larger establishments, when the corresponding activities differ substantially. In the latter case, the establishment record will be flagged for review by an industry analyst.

This example illustrates the use of redundancy to improve the quality of coding. Not all of the items would have been necessary to meet the basic data requirements of the Census Bureau; in particular, item 9 might be disposed of. However, its use provides an additional check on the accuracy of the code assignment and a basis for coding if other items are incompletely reported.

Analysts specializing in particular groups of industries usually have an important role in preparing the specifications for coding. For example, they may specify at what stages of computer processing and under what conditions cases should be referred to them. One example of this is described for the annual survey of manufactures (Bureau of the Census, 1971). Codes for a few selected industries involving heavy capitalization in specialized equipment, such as blast furnaces and steel mills, are frozen: i.e., no establishments can be recoded into or out of the industry without the express decision of the responsible analyst. In addition, the analysts often make key decisions as to what variables will be considered and how they will be used for classification of more or less ambiguous cases in the automated coding operation.

Automated coding ensures consistent decisions when the computer-assigned codes are accepted; however, it does not ensure that all of those codes are correct. If the specifications provided are incorrect or if the written computer programs do not conform fully to the specifications, then the codes assigned to units with certain characteristics can be consistently wrong. Gross errors are usually noticed quickly and remedied, but more subtle errors may escape notice.

Bureau of Economic Analysis

The industry coding system for BEA's direct investment (DI) statistics program provides an example of deliberately redundant manual and automated coding. The relevant section of form BE-12, Benchmark Survey of Foreign Direct Investment in the U.S., 1980, is shown as Exhibit A-7. The questionnaire is used for business enterprises that are U. S. affiliates of foreign persons. Items 23 to 33 are used to assign an industry code to the enterprise. The respondent is asked to enter a DI industry code in column (1), using the three-digit codes listed at the bottom of the page, for each category in which the business enterprise had sales or revenues,[2] and to enter the sales or gross operating revenues for each of the categories in column (2). After checking to see that the detail and total items in column (2) are consistent, an editor uses the entries to assign a DI industry code to the enterprise and enters it in column (1) of the line marked "BEA use only." Subsequently, all of the entries are keyed into the computer record for the enterprise, and an industry code is generated, using the rules followed by the editor. This is compared with the editor's code and differences are resolved manually.

Computerized Dictionary Systems

A completely different approach to automated industry coding could be used when the coding is based solely on a brief description of the major activity. The basic method would be to key the description for each unit and match it against a computerized "lexicon," i.e., a collection of word groups associated with specific codes. When there is an acceptable match, the industry code would be taken from the lexicon. When there is not, some other method of coding would have to be used. The lexicon must be developed from a manually coded data set, preferably one with a minimum of coding error. None of the systems reviewed for this study uses this method. The Census Bureau did some developmental work on it in the late 1960s, and the system was used operationally to code industry of sole proprietors based on activity descriptions from IRS Form 1040, Schedule C (O'Reagan, 1972), but this application was soon abandoned because the listings from IRS became available in computerized form, with industry codes included.

For about 5 years prior to the 1980 census of population, the Census Bureau undertook a substantial research and development effort directed at possible application of the same general approach to coding occupation and industry entries from census forms (Lakatos, n.d.). It was judged that the results obtained did not warrant use of the method in that census, but they were promising enough for the effort to be continued. The Statistics of Income Division of IRS (reported by Sailer et al., 1980; Sailer et al., 1983), is developing a dictionary for use in automated coding of occupations reported by taxpayers on individual income tax returns.

There are several good reasons that this method of industry coding has not been developed further for use in coding business enterprises. A key consideration is that a brief description of the unit's principal activity is simply not enough information for accurate coding. Significant cost savings are unlikely; the cost of keying the descriptions is substantial, and even after investing substantial resources in developing a computerized lexicon, it is likely that 10 percent or more of the cases will require manual intervention. However, there are potential gains in the quality of coding.

Consistency Checks

For those systems in which an industry code is first assigned manually, the most obvious automated check is to compare each code with a list of valid codes for that system. Most systems make this check, some at the time of data entry and others in computer edits.

In nearly all systems, many kinds of inter-item consistency checks are possible, but most systems do not exploit these possibilities very much to check industry codes. As described above, the largely automated coding systems used by the Census Bureau for the mail part of the economic censuses and for some surveys make use of several different items from the source document and examine their relationships in the process of assigning an industry code and in deciding whether the code should be reviewed by an analyst. These systems also make use of check digits to minimize data entry errors in the product codes used for industry coding in the manufacturing division.

TREATMENT OF INCOMPLETE DATA

Ideally, in each industry coding system the source document is expected to provide the information needed to classify each unit to the level of detail required by that system; frequently, it does not. A common problem is that respondent descriptions of activities, products, or services are not sufficiently precise. In addition, some relevant items on the source document may be left blank. In updating operations, some units may fail to return source documents (questionnaires) at all. Several of the options that are available to deal with this problem are discussed in the rest of this chapter.

Further Contact with the Unit

The most obvious solution to the problem of incomplete data, if time and resources are available, is to contact the business enterprise and request more information. Such contacts are normally done by telephone or correspondence. Most source documents--except various income tax returns--request the telephone number of the person

to contact for further information. Data on the frequency of follow-ups in different systems were not obtained for this study, nor are data available, for the most part, on the extent to which most of the other methods of dealing with incomplete data are used, the exception being the assignment of less detailed codes.

Use of Reference Materials

Several of the systems reviewed use commercially available lists or directories that identify specific business enterprises and give their four-digit SIC codes, along with other information. The specific sources reported for one or more systems were Dun & Bradstreet, Moody's, Standard and Poor's, and state industrial directories. The coverage of these reference sources varies. Many state directories cover only manufacturing. In general, corporations are more likely to be included than partnerships and sole proprietorships. Although it was not mentioned for any of the systems, the Securities and Exchange Commission (1981) issues a directory, with SIC codes, of companies required to file annual reports.

As mentioned above, the accuracy of codes derived from reference materials depends on two factors: first, the ability to correctly identify, in the reference list, the unit for which the code is needed; second, whether the code for that unit from the reference list is "correct" or can be transformed into a correct code, as defined for the system in which it is to be entered. The latter will depend, in turn, on the coding principles used to derive the code in the reference source and the extent to which those principles were fully and correctly followed.

Codes from Prior Years in the Same System

Codes from prior years are often used in conjunction with full current information, either to check the accuracy of the code based on current information or to apply resistance principles when activity changes occur. They may also be used in place of the current code when the information needed to assign a current code is incomplete.

A special case occurs in longitudinal files, such as those maintained for the continuous work history sample. When industry codes are missing for sample persons in one or more years, they can be imputed on the basis of codes assigned for both prior and subsequent years; a procedure for doing this is described by Levine (1980).

Codes from Other Federal Systems

Some federal agencies have access to data, including industry codes, for identifiable establishments or other units from data systems maintained by other federal agencies (for a current listing of transfers involving industry codes see Farrell et al., 1982:Table 2).

The Census Bureau, in particular, relies largely on transfers from IRS and SSA systems for early identification of new single-unit establishments and for data for zero-employee and certain other small establishments. These interagency transfers also provide a basis for industry coding when the normal source data are not obtained. An obvious problem is that the codes transferred may not be as detailed as those which the receiving agency is attempting to assign. For example, Census Bureau systems use full SIC detail (and sometimes more) while IRS systems use only about one-fifth as many categories. In addition, the codes assigned by IRS and SSA are often based on less complete information than is available from Census Bureau source documents.

Another similar example comes from the FTC's industry coding system for its quarterly financial report. Most of its coding is based on source document data, reference lists, and follow-up contacts. However, in rare instances, the code will be based on the industry code in the IRS listing of corporations, which is obtained by the FTC to use as a sampling frame for the QFR Program.

Partial Coding

When all of the above options fail or resources do not permit their full use to obtain information needed to code at the level of detail required by the system, some systems rely on partial coding, i.e., coding to two or three digits and filling in the remaining digits with values that signify "not allocable" within the designated SIC division, major group, or industry group.

The types of partial codes permitted are not the same in all systems. Table 5 gives some examples of partial codes used in different systems reviewed for this report. The BEA and FTC systems do not permit any partial codes. This is feasible because reporting on the source documents used is mandatory and because the number of units included in each system is relatively small, and coverage is limited mostly to large corporations, for which some information is usually available from reference lists and other public sources.

The industry classifications used by IRS to assign codes to corporation tax returns (not included in Table 5) follow a practice that is not recommended: they combine "not elsewhere classified" (n.e.c.) and "not allocable" units within certain SIC divisions, e.g., manufacturing, in the same category. The n.e.c. codes have been established to cover residual activities within an SIC division or major group that, according to the SIC Manual (Office of Management and Budget, 1972:10-11), ". . . do not usually constitute homogeneous primary activity groups; for purposes of this classification system they are grouped together and treated as a separate industry to retain the homogeneity of other industries in the group." For a unit to be properly classified in an n.e.c. industry or industry group, its specific activities must be known; this is quite different from the situation in which the activities are not clearly described and cannot

TABLE 5 Partial Codes Used in Selected Industry Coding Systems

Agency	Industry Coding System	Completely Unclassified (SIC Division K)	Partial Codes Used
BEA	Direct investment	Not used	Not used
BLS	ES-202	Used	Not used
Census Bureau	Economic censuses	Used	Any valid two- or three-digit code, filled out with 0s
D&B	Business activity coding	Used	Not used
FTC	Quarterly financial report	Not used	Not used
IRS	Sole proprietors, statistics of income	Used	1. Not allocable codes for SIC divisions: construction, manufacturing, wholesale, retail. 2. Not allocable, wholesale and retail combined. 3. Not allocable, farms (SIC major groups 01 and 02).
SSA	Single and multiunit	Used	Any valid two- or three-digit code, filled out with 0s

be allocated to a particular category within the division or industry group.

If a partial code in one system agrees with a complete code in another system at the level of detail included in the partial code, the two codes are, in a sense, comparable. However, the use of partial codes creates some obvious problems for the comparability of aggregate data by industry from other systems.

Force Coding

Under certain conditions it may be considered desirable to avoid the use of partial codes, especially at the three-digit level. The use of not allocable categories (at the fourth digit level) for a large number of industry groups complicates the presentation of data

and their use by analysts, and it can also cause problems in doing the prepublication analyses and adjustments necessary to avoid disclosure of data for individually identifiable establishments.

An imputation method that has been used by the Census Bureau to eliminate two-digit not allocable codes is called "force coding." For partially coded establishments in each industry group, third and fourth digits are assigned by a random process designed to generate a distribution of these establishments by industry similar to that observed for completely coded establishments. This process was used in the 1980 county business patterns program to code approximately 44,000 units that could not be classified beyond the two-digit level by other methods.

NOTES

1. Specialization does not guarantee better quality. It reduces the number of coders for a given workload, and if there is a significant correlated coder variance (i.e., some coders consistently make certain errors), the reduced number of coders is a disadvantage.
2. Due to an oversight, the boxes in column (1) for items 31, 32, and 33 were not shaded. It is clearly inappropriate to ask respondents to enter DI industry codes in these boxes.

CHAPTER 5

QUANTITATIVE INFORMATION ON COMPARABILITY AND ACCURACY

INTRODUCTION

The discussion of comparability and accuracy of industry coding so far has been largely in qualitative terms. The factors that lead to differences between systems have been identified. Some of these factors, such as coverage, definition of units, and classification principles, depend primarily on the particular purposes for which each data system has been developed. Other factors, such as the kinds of source data and the procedures used for coding, depend on the resources available and on the judgments and preferences of system designers. Differences also arise from errors in carrying out the coding procedures. Several examples of features of different coding systems have been presented, and readers, on the basis of these, may have already begun to form some intuitive conclusions as to the relative accuracy of codes in different systems.

This chapter presents some quantitative data bearing on the comparability and accuracy of industry coding in different systems. The data come from both published and unpublished sources, the latter consisting largely of items supplied to the Industry Coding Working Group by the participating agencies. The next section covers intersystem macrocomparisons, i.e., comparisons of aggregate data by industry from different systems. The third section presents results from intersystem microcomparisons, i.e., comparisons of industry codes from different systems for identical units. The final section presents information on components of error in individual systems.

INTERSYSTEM MACROCOMPARISONS

It is fairly routine for an agency to compare aggregate data for such items as employment, payroll, and receipts, by industry, with similar data produced by other agencies or other systems within the agency. Generally, the data sets compared cannot be expected to agree fully because there are differences in coverage, concepts, and definitions, but comparisons are sometimes useful as a means of detecting gross errors in one or both data sets. Such a comparison may be regarded as a rough diagnostic device. The location and

correction of specific errors require a more detailed examination of the cells in which large differences occur.

Observed differences in aggregates do not provide any direct information about the accuracy of industry codes in the systems compared; however, differences in industry codes for identical units may explain some proportion of the differences in the aggregates and often have done so when individual unit comparisons have been made (see the next section). Ideally, the sequence of investigation for such comparisons has four steps:

(1) Compare the available documentation of the coverage, concepts, and definitions associated with the data sets;

(2) Compare the data sets at a broad level, e.g., national totals by SIC division or major group;

(3) When large differences are found, make comparisons at a lower level of aggregation, e.g., by state and industry group or industry;

(4) For the cells with large differences, match individual units from the two systems and compare the data items and industry codes.[1]

This ideal approach runs into practical difficulties. Analysis of results obtained by matching individual units is often technically difficult and costly, and the ability to match may be limited by agency confidentiality requirements.

One example of the general approach is described in a report from the Bureau of the Budget (1961). The Census Bureau's 1947 Census of Manufactures produced employment figures about 7 percent below those of BLS's current employment statistics. The Budget Bureau's Division of Statistical Standards established an interagency working group to explore the reasons for the difference. The working group undertook case studies of how 60 of the largest companies in manufacturing were reporting employment data to the Census Bureau and BLS. These studies eventually led to several clarifications of and changes in the establishment definition, the treatment of administrative offices and auxiliary units, and the structure of SIC categories within the manufacturing division. About 35 of the 60 companies studied agreed at the time to report on a uniform basis for the same list of establishments to all the agencies. The 1954 Census of Manufactures produced employment figures that differed from those of BLS by only 182,000 (about 1 percent). The Budget Bureau report took this result as a demonstration that "the work over the years had not been in vain."

Another comparison that led to a matching study is reported by the Bureau of the Census (1965b). Payroll statistics from the retail portion of the economic censuses for 1958 and 1963 were compared with data from the Bureau of Employment Security (BES) for 19 states in which coverage rules in the two systems were believed to be the same. The BES payroll totals exceeded those from the Census Bureau by 5.8

percent in 1958 and by 7.2 percent in 1963; this result led to a matching study for the state of Delaware, which is discussed in the next section.

The Bureau of Economic Analysis (1972) made extensive comparisons of aggregate data on employment and wages by industry from several sources in connection with a study for the Department of Labor on the usefulness of SSA's Continuous Work History Sample (CWHS). These comparisons, which involved data from the CWHS (both the 1 percent and 10 percent samples), population censuses, the County Business Patterns (CBP) Program, and the Unemployment Insurance system, are also summarized in another BEA report (Bureau of Economic Analysis, 1976:chapter VII). The observed differences are the result of several factors, so it is impossible to draw any firm conclusions from the data about differences in industry coding. There are very large differences between systems in the number of persons employed in service industries. The report says (1976:92):

> CWHS services employment tends to be higher because of the inclusion of many public service workers (for example, in educational institutions or hospitals) who are either classified as government workers in the CBP and UI data or are excluded.

Government establishments are, in fact, excluded from CBP data, so the main implication is that the SSA and BLS systems may have been classifying some government establishments differently during the period covered by these comparisons (mainly 1971 and 1973).

Other more recent aggregate or macrocomparisons are available in both published and unpublished form: for examples of published comparisons, see Office of Federal Statistical Policy and Standards (1977a:29, 1980); Metropolitan Washington Council of Governments (1977); and Harris (1981). However, none of these comparisons offers any additional enlightenment on comparability and accuracy of industry coding in different systems.

INTERSYSTEM MICROCOMPARISONS

Introduction

This section covers the comparison of industry codes for individual units in different systems that cover, at least in part, the same business establishments or enterprises. Such comparisons may involve two different data bases or coding systems in the same agency, or they may involve systems in more than one agency. Some comparisons occur as a relatively low-cost by-product of routine processing operations; others require special arrangements for the matching of records from two or more systems.

Most microcomparisons require two steps. The first step is a matching operation to identify records for corresponding units in the systems being compared. The matching normally produces a certain

proportion of one-to-one or "perfect" matches, i.e., pairs of records, one from each system, that clearly are for the same establishment or other unit. For these units, the second step is a straightforward comparison of classifiers, including SIC codes, and data items. However, the second step will also usually require comparisons of cases for which the relationships between units in the two systems are more complex, e.g., one unit in system A may correspond to two or more units in system B, etc. In such cases, a clear interpretation of differences in industry codes is not always possible. In most matching studies, there are also cases for which, as a result of errors or differences in reporting and recording corporate names, addresses, or identification numbers, it is difficult to determine with a high degree of confidence whether or not a good match exists. If a common identifier, such as the EIN, is used in both systems, this is not likely to be a serious problem. If there is a high proportion of doubtful cases, some kind of follow-up to resolve them may be needed.

The comparison of industry codes must, of course, take into consideration the inherent differences in the industry coding principles and procedures used in the systems being compared. In particular, if SIC industries are grouped or subdivided in one or both systems, comparable groupings for the two systems must be established.

What can be learned from intersystem microcomparisons of industry codes? Strictly speaking, the fact that two systems have assigned different industry codes for the same establishment indicates only that at least one of the codes is incorrect. Conclusions as to the accuracy of either system or their relative accuracy require either examination of the reasons for differences or an a priori judgment that one system assigns codes more accurately. Such a priori judgments are sometimes justified. For example, industry codes assigned by IRS in its Statistics of Income Program should, on the average, be more accurate than those assigned in IRS's revenue processing operations, because the SOI coders make fuller use of all information available for classifying each unit.

When individual differences are examined it is often possible to determine why they occurred and what the correct code is, but such analyses are time-consuming and cannot be done on a large scale. However, they can be useful in two ways: first, to improve intersystem comparability by uniform treatment of large units; second, to suggest changes in coding principles and procedures in either or both systems in order to improve their accuracy and comparability.

Interagency Comparisons Between Systems

A very early example of interagency comparisons of industry codes is reported by the Bureau of the Budget (1947):

> In 1939 the Central Statistical Board made an experimental study of 103 largest enterprises (10,000 and more employees), in which the industrial classification of each

agency (SEC, BIR [Bureau of Internal Revenue], SSB [Social Security Board]) was translated into the Standard Industrial Classification and examined for agreement. Result of examination of the list of 103 enterprises: 76 were listed by 3 agencies, 26 by 2, and 1 by 1 agency. Out of 76 listings by 3 agencies, 70 cases were in complete agreement and 6 cases in disagreement. Of the 26 listings by 2 agencies, 20 cases agreed and 6 disagreed.

The Bureau of the Census (1951) describes a special study carried out in connection with the reconciliation of codes assigned in the 1947 Census of Manufactures with those in the SSA (then known as OASI) system. This study covered a sample of 600 establishments classified as manufacturing by the Census Bureau and nonmanufacturing by SSA, or vice versa. It was found impossible in most cases to reach agreement on the proper classification by examining the information in the two agencies' source documents. Therefore, new forms were sent to each establishment to obtain current data. When the forms were returned, each establishment was independently classified as manufacturing or nonmanufacturing by both agencies. The results are shown in Table 6.

TABLE 6 Results of Independent Coding of Establishments by the Bureau of the Census and the Social Security Administration

Outcome	Number of Establishments
Total in sample	600
Out of business since 1947	91
Insufficient information	51
Balance	458
Identical OASI-Census classification	404
Different Census-OASI classification, the Census or OASI classification being preliminary subject to change pending additional information	21
Census-OASI classification difference	33

Source: Bureau of the Census (1951:5).

Considering that the sample cases were generally on the borderline between manufacturing and nonmanufacturing, there was relatively good agreement. The report takes these results as evidence that differences in source documents can often lead to assignment of different codes.

Another Census Bureau (1965a) study provides a comparison of industry codes assigned to a sample of about 2,000 employed persons, based on information reported by or for them in the 1960 population census, with industry codes assigned to their employers by the SSA. Matching was based on employer names and addresses reported in the population census. Results are reported for 14 industry categories corresponding, for the most part, to SIC divisions. Of the matched cases with industry codes, about 15.1 percent (weighted estimate) were classified by the SSA and Census Bureau in different categories. The category most clearly prone to error was wholesale trade, for which the Census Bureau estimate (based only on matched cases) was 43 percent below the SSA estimate, and the estimated index of inconsistency was 53.[2] It is doubtful that the results of this employer record check by themselves could be used to reach any firm conclusions about which system contained more accurate classifications. The SSA's industry codes come from several different sources; it would have been of some interest to tabulate the observed differences and rates separately for each major source. Both the Census Bureau and the SSA source documents had inquiries specifically designed to distinguish wholesale and retail trade: however, the Census Bureau inquiry assumes that the respondent knows the difference between wholesale and retail trade, as defined in the SIC, while the SSA source document inquiry does not.

Still another Census Bureau (1965b) study was undertaken because of differences in aggregate payroll figures for retail trade from the 1958 economic censuses and the current statistics from the Bureau of Employment Security (BES). Individual records for the state of Delaware from the two systems were matched. A sample of about 100 retail establishments from the 1963 retail census was matched against the full BES file, and about 200 sample cases from the BES retail file were matched against the retail census. Matching in each direction required some grouping of establishments from the same company in order to conform to the BES reporting format. All matched cases with differences in SIC classification were reviewed jointly by Census Bureau and BES personnel, using source documents. If information from the two sources was contradictory, telephone calls were made to establish the correct SIC classification.

Table 7, taken directly from the Census Bureau's report (1965b:Table VI), shows the reasons for those cases in which it was determined that an establishment or reporting unit was incorrectly included in or excluded from the Delaware retail universe by one of the two agencies. The table shows that all of the BES errors and nearly two-thirds of the Census Bureau errors (in terms of payroll) resulted from classifying a unit in the wrong SIC division. The estimated net overstatement of retail payroll resulting from incorrect classification by BES was about 7.6 percent, and the net

TABLE 7 Summary of Errors as a Result of Reconciling Bureau of Employment Security and Census Bureau Records on Delaware Retail Payroll, 1963

Nature of Error	Erroneous Exclusion from Delaware Retail Universe	Erroneous Inclusion in Delaware Retail Universe	Net Overstatement of Delaware Retail Universe
A. BES Error			
1. Wholesale unit of retail multiunit included	--	$1,154	
2. Coded wholesale--should be retail	$1,759		
3. Coded retail--should be service		205	
4. Coded retail--should be wholesale		6,336	
5. Coded retail--should be manufacturing		1,033	
Total	1,759	8,728	6,969
B. Census Error			
1. Coded retail--should be service		372	
2. Coded retail--should be wholesale		867	
3. Coded service--should be retail	297		
4. Coded wholesale--should be retail	1,203		
5. Coded manufactures--should be retail	387		
6. Coded out-of-scope by SSA--should be retail	647		
7. No EI number found in SSA file	105		
8. In Census mailout--not in tabulation	272		
9. Combined in the reports for other states	1,820		
Total	4,731	1,239	3,492

Source: Bureau of the Census (1965:Table VI).

understatement by the Census Bureau was about 1.6 percent. Among the units classified in retail trade by both agencies, about 2 percent of payroll was accounted for by units classified in different major groups within retail trade. The results pointed clearly to SIC classification differences as an important factor leading to differences in aggregate data from the two sources.

As the Census Bureau started to make greater use of administrative records in the economic censuses during the 1950s and 1960s, various studies were carried out to evaluate the quality of the administrative record data. One such study (Bureau of the Census, 1968) compared final industry codes for single-unit establishments in the 1963 economic censuses with mailing list codes obtained from SSA. The latter codes had been derived by SSA in part from the 1958 economic censuses and in part (primarily for births, i.e., new or reactivated establishments, after 1958) directly from employers from the SS-4 (application for EIN) or a follow-up inquiry.

Table 8 shows the main results of this comparison. Of the 1,958,000 Census Bureau mail cases matched to the SSA single-unit employer file, 279,000, or about 14 percent, had not been classified to the 4-digit SIC level by SSA. Of the remainder, 83.0 percent were given the same final Census Bureau code as that in the SSA file. Another 11.5 percent were assigned to the same division; for the remaining 5.5 percent, there was not agreement at any level of detail.

Other results showed that SSA-based mailing list codes were changed at almost the same rate whether they were based on the 1958 economic censuses (15 percent) or on information obtained by SSA directly from employers (18 percent). The implications of this finding are not clear, because changes resulting from real activity shifts are confounded wth those resulting from incorrect classification. However, a priori, one would expect fewer differences resulting from real activity shifts in the latter group. Of the 279,000 employers not classified by SSA to the 4-digit level, 205,000 were in retail trade, and of those in retail trade, 165,000 (over half the total) were eating and drinking places.

In a study following the 1967 economic censuses (Bureau of the Census, 1969), final economic census SIC codes were compared with codes assigned by IRS in revenue processing. This study was based on a sample of 22,443 retail, single-unit, sole proprietorships with employees and for which the IRS principal industrial activity (PIA) codes were available. Presumably this group was selected to avoid multiunit matching problems and because the Census Bureau and PIA codes for sole proprietors are more directly comparable than they are for some other SIC divisions. Also, the smaller units are of greatest interest because there is a greater potential for relying entirely on tax returns to obtain economic census data for those units.

For 37 industries and industry groups in retail trade, a direct comparison of Census Bureau and PIA codes was possible. For the 37 groups (based on Census Bureau SIC codes), only 6 groups had the same PIA code for more than 80 percent of the establishments; there were 16 groups that had different codes for more than half of the establishments. Distributions of the number of establishments and

TABLE 8 Results of Comparisons Between Final Industry Codes and SSA-Based Mailing List Codes: 1963 Economic Censuses

Result of Comparisons	Establishments		
	Number (000)	Percent of Total	Percent of Matched Classified to Four-Digit
Total single-unit establishments in censuses	2,117	100.0	--
Not matched to SSA	159	7.5	--
Matched to SSA	1,958	92.5	--
Not classified to four-digit level by SSA	279	13.2	--
Classified to four-digit level by SSA	1,679	79.3	100.0
Same four-digit code	1,393	65.8	83.0
Same three-digit, different four-digit code	67	3.1	4.0
Same two-digit, different three-digit code	70	3.3	4.1
Same SIC division, different two-digit code	57	2.7	3.4
In scope of economic censuses, different division	78	3.7	4.6
Out of scope	15	0.7	0.9

Source: Bureau of the Census (1968:3).

value of sales by industry group showed that there would have been substantial differences in data by industry had the PIA codes been used in place of the Census Bureau SIC codes for these establishments.

In this instance, it seems reasonable to assume that the Census Bureau SIC codes were generally more accurate than IRS's PIA codes, since the former were based on considerably more detailed information about each establishment's sales by merchandise line. This assumption is supported by the fact that PIA codes were more common in some of the more general and "catch-all" categories, such as hardware stores; grocery stores; miscellaneous food stores; and miscellaneous retail stores, not elsewhere classified. The last two categories probably represent a misuse by IRS of these categories, which are intended to be used for clearly defined activities that do not fit into any homogeneous grouping within the SIC major group.

Recently, the Statistics of Income Division of IRS and the Office of Research and Statistics of SSA have been undertaking joint studies with a view toward possible reduction of the overall volume of their coding operations through code sharing. One of these studies (Internal Revenue Service, 1982a) compared industry codes assigned to a small sample of sole proprietorships reported on Form 1040 Schedules C and F for 1978 with the SSA codes for those that could be matched in the SSA single-unit employer file. The assignment of codes to these cases by IRS was done using standard statistics of income procedures, i.e., making use of all relevant information on the Schedule C or F. For 149 cases for which the IRS and SSA industry codes could be compared, the results were as follows:

Category	Number
Exact match (at the finest level of detail possible considering differences in the coding systems)	87
Partial match (matching on at least the first digit, but not an exact match)	15
No match (different first digits)	47
Total	149

This study was based on a small stratified probability sample of Schedules C and F, and the results were not weighted to reflect the different sampling fractions used. Even so, it is probably safe to conclude that there is at present only limited comparability between the codes for sole proprietorships in the IRS and SSA systems. One can only speculate about the relative accuracy of classification in

these systems. In general, the SSA codes are based on greater detail, but the information used by the IRS for coding is more recent.

Intra-Agency Comparisons Between Systems

Prior to the development of the SSEL, industry classification of establishments by the Census Bureau in economic censuses and current surveys was less fully coordinated than it is now. An early study (Bureau of the Census, 1951) compared industry codes for 500 single-unit establishments from the 1949 Annual Survey of Manufactures with codes assigned to the same units in the 1947 Census of Manufactures. For the 57 cases (11.4 percent) with code differences, the census and survey schedules were analyzed to discover the reasons for the differences. The results are shown in Table 9. The striking finding is that less than one-third of the difference turned out to be actual changes. Most of the others could be accounted for by the use of different source documents and product categories and by coding errors.

A more comprehensive analysis of the 30,000 "large" establishments in the 1949 Annual Survey of Manufactures sample showed that real changes in primary activity at the four-digit SIC level occurred for

TABLE 9 An Analysis of 1947-1949 Code Changes for 500 Single-Unit Establishments in Manufacturing

	Number of Establishments	Percent of all Cases Examined	Percent of Code Changes
Total number of schedules examined	500	100.0	
Total code changes, 1947 to 1949	57[a]	11.4	
Classified cases	52	10.4	100.0
"Response" differences	33	6.6	63.5
Coding differences	4	0.8	7.7
Activity changes 1947-1949	14	2.8	26.9
Death-birth	1	0.2	1.8
Unclassified cases	5	1.0	

[a]Does not include possible code changes for establishments (estimated 7 percent of total) reporting product combinations affecting their industry classification.

Source: Bureau of the Census (1951:2).

only 995, or 3.3 percent of the establishments. However, there were an estimated 2,000 to 3,000 additional cases for which ". . . it was found that what appeared to be reported changes in primary activity were actually response differences relating to the same primary activity in both 1947 and 1949" (Bureau of the Census, 1951:3).

Another report from the Census Bureau (1963) describes an intensive analysis of differences between the 1958 census of retail trade and the monthly retail trade sample survey covering the same period. Total retail sales from the two sources showed a net difference of less than 0.5 percent; however, differences for some kinds of business were considerably greater (e.g., 10 percent for gasoline service stations), and the analysis showed that there were significant compensating differences with respect to coverage, classification, and reported sales.

Classification differences were of two types: between SIC divisions and within the retail division. In the first instance, establishments were classified as in retail in the retail trade census and not in retail in the current survey, or vice versa. Data on the size of these differences, for the kinds of business most affected, are shown in Table 10. For the most part, the differences involved shifts between retail and wholesale trade. However, in the case of milk distributors (part of the category "nonstore retailers") and retail bakeries, the shifts were largely between retail trade and manufacturing.

Table 11 shows classification differences by major kind of business for establishments classified as retail in both the retail trade census and the current survey. (As in Table 10, the large multiunit retail firms were excluded.) The largest relative net shift was for nonstore retailers; this category was used to a much larger extent in the census than in the current survey. The second largest relative net shift was for general merchandise stores.

Examination of similar data for 30 detailed kinds of business classes showed indexes of gross shift of 0.30 or more for the following: hardware stores, general merchandise group; variety stores; meat markets; tire, battery, and accessory stores; family clothing stores; household appliance stores; drinking places; and nonstore retailers. A shift between meat markets and grocery stores occurred because of a difference in definition: the census classified any store having 50 percent or more of its sales in meats as a meat market; the cutoff for the current survey was set at 80 percent. In the case of drinking places, the shift was primarily between eating places and drinking places. As noted in Chapter 2, the BLS and SSA systems combine these two categories because of the difficulty in distinguishing between them.

The Statistics of Income Division (formerly Statistics Division) of the Internal Revenue Service has made several studies comparing industry codes contained in the IRS master files for all business returns with those assigned in the Statistics of Income Program to businesses included in the SOI samples for sole proprietorships, partnerships, and corporations (Internal Revenue Service, 1973, 1974; Powell and Stubbs, 1981). In general, the SOI codes are believed to

TABLE 10 Indexes of Shift for In-Scope and Out-of-Scope of Retail Trade, by Kind of Business

Kind of Business	Index of Shift	
	Gross	Net
Lumber, building, hardware, and farm equipment	.17	-.01
Lumber yards	.12	-.05
Hardware stores	.07	-.07
Retail bakeries	.29	-.17
Tire, battery, and accessory stores	.22	-.13
Gasoline service stations	.07	-.03
Household appliance stores	.23	.10
Other retail stores	.22	-.08
Nonstore retailers	.35	-.03
United States, total	.07	-.02

Note: Indexes are defined as follows:

Index of gross shift $(A_i + B_i) / 1/2 (X_i + Y_i)$

Index of net shift $(A_i - B_i) / 1/2 (X_i + Y_i)$

where X_i = the census total for kind of business "i"

Y_i = the current survey total for kind of business "i"

A_i = sales of establishments in scope of census and out of scope of current survey

B_i = sales of establishments in scope of current survey and out of scope of census

Source: Bureau of the Census (1963:5).

TABLE 11 Major Kind-of-Business Cross-Classification of Group I Retail Trade Establishment Sales in Census and in Current Survey: United States, 1958 (millions of dollars; current survey sales estimates throughout)

Current Survey Major Kind-of-Business Classification	Total	Census Major Kind-of-Business Classification										
		Lumber, Building, Hardware, Farm Equipment Dealers	General Merchandise Stores	Food Stores	Automotive Dealers	Gasoline Service Stations	Apparel, Accessory Stores	Furniture, Home Furnishings, Appliance Stores	Eating and Drinking Places	Drug and Proprietary Stores	Other Retail Stores	Nonstore Retailers
Total	137,544	10,526	8,944	26,304	29,816	12,606	8,851	8,053	12,308	5,607	13,343	1,186
Lumber, building, hardware, farm equipment	10,345	9,779	32	7	75	86	1	75	34	6	139	111
General merchandise stores	8,348	63	7,128	403	1	6	489	125	5	1	51	76
Food stores	26,532	5	744	24,998	6	203	3	3	186	62	168	154
Automotive dealers	30,070	158	13	21	29,538	183	--	25	25	--	105	2
Gasoline service stations	12,874	56	106	237	85	12,009	--	2	186	10	180	3
Apparel, accessory stores	8,942	--	575	9	1	2	8,225	3	3	4	63	57
Furniture, home furnishings, appliances	8,390	223	88	49	43	4	19	7,586	13	6	206	153

Eating and drinking places	12,306	2	52	299	5	52	-28	3	11,632	10	239	40
Drug and proprietary stores	5,524	1	36	16	--	--	20	37	7	5,386	18	3
Other retail stores	13,984	239	170	255	62	61	122	194	216	122	12,174	369
Nonstore retailers	229	--	--	10	--	--	--	--	1	--	--	218

Note: The estimates in this table are subject to sampling error and bias.

Source: Bureau of the Census (1963:21).

TABLE 12 Differences Between IRS Master File Codes and SOI Industry Classification, by SIC Division and Type of Organization

SIC Division	Index of Gross Shift[a]		Index of Net Shift[a]		Percent Master File Agreement With SOI	
	Sole proprietorships 1969[b]	Partnerships 1971[c]	Sole proprietorships 1969[b]	Partnerships 1971[c]	Sole proprietorships 1969[b]	Partnerships 1971[c]
Agriculture, forestry, fishing	0.90	0.25	-0.52	-0.14	74.1	94.1
Mining	0.21	0.22	0.09	-0.12	85.6	94.9
Construction	0.23	0.16	-0.08	-0.07	92.3	95.3
Manufacturing	0.71	0.32	0.19	0.11	59.2	79.5
Transportation, public utilities	0.37	0.44	0.09	0.20	78.2	71.0
Wholesale trade	0.74	0.34	0.53	0.18	49.8	76.2
Retail trade	0.20	0.10	-0.05	[d]	92.8	95.0
Finance, insurance, real estate	0.20	0.09	0.05	0.06	88.0	93.0
Services	0.14	0.20	-0.04	-0.13	94.9	96.1

[a]See definition given in Table 9. Negative value for net shift means master file count in category greater than SOI count.
[b]Data from Internal Revenue Service (1973).
[c]Data from Internal Revenue Service (1974).
[d]Absolute value less than 0.005

be more accurate than the master file codes because the SOI industry coders make fuller use of all relevant information on the returns and even use commercial directories in some cases. For partnerships and corporations, the master file codes are usually those entered by taxpayers.

Table 12 shows results, at the SIC division level, from two studies that compared SOI and master file codes. The measures shown are based only on those cases for which a valid industry code, other than "not allocable by SIC division," was assigned in both systems. There were no valid industry codes in the master file for 20.1 percent of the sole proprietorships and 9.1 percent of the partnerships. The measures shown in Table 12 are based on unweighted tablulations of SOI sample cases; hence, the smaller units are underrepresented.

Based on Table 12, several observations can be made. First, there are large differences between the two systems, and the large indexes of net shift for some SIC divisions show that these differences do not always balance out. It is difficult to agree with the statement in one of the IRS reports: "On a broad basis, the two coding systems yielded fairly comparable results" (Internal Revenue Service, 1973:1). Considering that both systems used the same source documents, the differences might be considered surprisingly large.

Second, the master file codes for partnerships were largely those supplied by taxpayers, while for the sole proprietorships the codes were derived by tax examiners from the activity descriptions on the returns. No firm conclusions about the relative accuracy and reliability of these two coding procedures can be drawn from these data; however, there is certainly no clear evidence that self-coding produces worse results. If anything, the data suggest the opposite conclusion. Third, as noted already in several other studies, the differences associated with wholesale trade are especially large.

Further examination of the detailed results shows that the largest indexes of net shift between SIC divisions were accounted for primarily by: sole proprietorships classified in agriculture in the master file and in wholesale trade or services in the SOI coding; sole proprietorships classified in retail trade in the master file and in wholesale trade in the SOI coding; and partnerships classified in transportation and public utilities in the master file and in services in the SOI coding.

The results shown in Table 12 were based only on cases for which a return was classified in different SIC divisions in the two systems. Table 13 shows, by SIC division, the percentage of cases classified differently in the two systems at the division, major group (two-digit), and industry group (three-digit) levels. Unlike Table 12, Table 13 includes those SOI sample returns for which there was no valid industry code in the master file. As a result, the division level percentages for sole proprietorships and partnerships in Table 13 are lower than those in Table 12.

By definition, the percentage agreement must decrease or remain the same as the level of detail increases from division to major group to industry group. Looking at how much the agreement drops off from one level to the next is a useful way of finding out where special coding

TABLE 13 Percent of IRS Master File Codes Agreeing With SOI Codes, by Type of Organization and Level of Detail

	Percent Agreement With SOI Codes at							
	Division Level			Major Industry Level (two-digit)		Industry Group Level (three-digit)		
	Sole proprietorships 1969[a]	Partnerships 1971[b]	Corporations 1972[c]	Sole proprietorships 1969[a]	Partnerships 1971[b]	Corporations 1972[c]	Sole proprietorships 1969[a]	Partnerships 1971[b]
Agriculture, forestry, fishing	50.7	90.8	79.0	36.3	86.9	78.3	35.0	61.9
Mining	40.1	85.5	88.2	39.4	84.1	87.7	n.a.[d]	n.a.[d]
Construction	84.5	89.2	89.2	73.9	82.9	89.2	68.1	79.8
Manufacturing	50.1	69.1	88.2	34.7	37.1	72.8	n.a.[d]	n.a.[d]
Transportation, public utilities	62.9	64.1	75.7	55.0	61.2	70.6	54.7	59.7
Wholesale and retail trade	75.3	83.9	87.7	62.9	73.6	75.4	57.5	71.3
Finance, insurance, real estate	71.5	83.5	84.7	67.1	74.7	75.8	40.2	67.9
Services	81.1	90.3	91.7	72.9	82.4	71.6	68.8	77.8

Note: SOI sample returns with no valid master file codes are included in the base, and are counted as not in agreement.

[a]Data from Internal Revenue Service (1973).
[b]Data from Internal Revenue Service (1974).
[c]Data from Powell and Stubbs (1981).
[d]n.a.--IRS does not classify to three-digit level in these divisions.

problems exist. For example, for partnerships in agriculture and forestry and fishing, agreement drops off from 86.9 percent at the major group level to 61.9 percent at the industry group level. This drop was primarily the result of returns classified as farms in both systems but classified in different farm types (field crop; fruit, tree nut, and vegetable; livestock; animal specialty; and other). For another example, for sole proprietorships in finance, insurance, and real estate, agreement drops off from 67.1 percent at the major group level to 40.2 percent at the industry group level. This drop resulted primarily from a group of returns classified in real estate in both systems, but classified differently to the seven industry groups used within that major industry.

Table 14 shows data on the extent of agreement at the major group level between master file and SOI industry codes for corporations in tax years 1972 and 1973, by SIC division. The percent agreement was lower in 1973 in all divisions except transportation and public utilities. For four divisions--agriculture, forestry, and fishing; construction; wholesale and retail trade; and finance, insurance, and real estate--the percent agreement was substantially lower in 1973.

TABLE 14 Agreement of IRS Master File Codes with SOI Codes at Major Group Level for Corporations: Tax Years 1972 and 1973

	Percent Agreement with SOI codes	
SIC Division	1972[a]	1973[b]
Agriculture, forestry, fishing	78.3	29.5
Mining	87.7	86.2
Construction	89.2	52.1
Manufacturing	72.8	72.3
Transportation, public utilities	70.6	75.7
Wholesale and retail trade	75.4	41.0
Finance, insurance, real estate	75.8	64.7
Services	71.6	70.1

[a]Data from Powell and Stubbs (1981).
[b]Data from Internal Revenue Service (1975a).

The probable explanation for these results is that the 1972 revision of the SIC was first implemented by IRS for tax year 1973. The revision required several changes in the list of activities and codes provided to taxpayers for self-coding on their returns. In all probability, a substantial proportion of taxpayers simply copied their industry codes from their previous year's return without referring to the instructions to see whether the code was still appropriate. This is borne out by a tabulation of the master file codes for 1973 showing that no fewer than 46.3 percent of the four-digit industry codes in the business master file were invalid (Internal Revenue Service, 1975b).

INDUSTRY CODING ERRORS IN INDIVIDUAL SYSTEMS

Direct or indirect evidence about the level of industry coding error in individual systems is available from several sources, such as quality control records, tabulations showing the number of units not classified or only partially classified by industry, and special studies to measure selected components of error. This section presents available data in the following sequence: errors of nonresponse leading to incomplete classification; response errors, i.e., those occurring in the data collection process; processing errors, i.e., those occurring in connection with manual coding or data entry; and general information not restricted to specific components of error.

Nonresponse Errors

Various methods of dealing with incomplete data for industry classification were described in Chapter 4, "Treatment of Incomplete Data." The evidence at hand on the results of these efforts for different systems is not as complete and uniform as might be wished; however, a reasonably good picture can be had from various sources, mostly published. An agency-by-agency presentation of available data follows.

Bureau of the Census

The most significant nonresponse problem for the Census Bureau is that connected with new or reactivated establishments (births). For single-unit enterprises, information about new units is received primarily from the IRS and SSA. Significant proportions of these units are unclassified or only partially classified by four-digit industry. The latter may occur because the source agency system groups some industries (as explained in Chapter 2), because the information on the source document is incomplete, or, especially in the case of IRS, because an invalid code has been assigned.

Before each quinquennial round of economic censuses, special efforts are made to reduce the number of unclassified units in the

SSEL in order to ensure that units within the scope of the economic censuses are included and that those meeting criteria for inclusion in the mail portion of the censuses are sent the appropriate questionnaires. As a result, the number of unclassified units in the SSEL tends to show a cyclical variation, rising to its highest point between each round of economic censuses.

For 1979 (two years after the 1977 economic censuses), approximately 220,000 or 4.2 percent of the active establishments in the SSEL were unclassified, but those establishments accounted for only about 0.6 percent of total employment (Bureau of the Census, 1982a). All of the unclassified establishments were single units. For new establishments in multiunit enterprises, if the information reported in the Company Organization Survey is not enough to assign an industry classification, codes are assigned either by making additional contacts or by imputation based on the pattern of activity for other establishments operated by the same company.

The published 1977 County Business Patterns report (Bureau of the Census, 1981a) shows 60,613 or 1.4 percent of all establishments as completely unclassified; however, those establishments accounted for only about 0.1 percent of total employment. The corresponding published figures for 1979 were 219,736 establishments (4.8 percent of the total), accounting for 0.7 percent of employment.

Bureau of Economic Analysis

According to the description of the classification system used for BEA's direct investment statistics file (prepared for the Industry Coding Working Group), all units are fully classified since they are required by law to report sales distributions.

Bureau of Labor Statistics

No quantitative data were available on the extent of incomplete industry classification in the agency's ES-202 (Unemployment Insurance Employment and Wages Program) report file. According to the system description prepared for the Industry Coding Working Group, state employment security agencies, which are responsible for the industry coding, are expected to deal with incomplete data as follows:

> If there is incomplete information to assign a SIC code, either a BLS-3023 form (for new accounts) is sent to the employer or he is contacted by telephone to obtain the needed information. In the interim, the establishment is put in an unclassified 9999 group. Units with employment greater than 50 should not carry an SIC code of 9999 for longer than one quarter. Smaller units in the 9999 classification remain until the next annual SIC refiling when information necessary to classify them is obtained.

Federal Trade Commission

According to the system description for quarterly financial report (QFR) industry coding, there is no incomplete classification. More than 99 percent of the units are classified by reference to the source documents or commerical lists. The remainder are classified by contacting respondents or, very infrequently, by adopting the industry code on the list provided by IRS for use as a sampling frame.

It can be observed that industry classification errors by IRS could result in coverage errors for the QFR program, since the sampling frame provided by IRS includes only corporations classified in the four SIC divisions within the scope of the QFR program. This coverage problem is likely to be less serious in the future since the QFR program was transferred to the Census Bureau late in 1982, and it will be possible to use the SSEL as a sampling frame.

Internal Revenue Service

The extent of incomplete classification in the SOI (sample-based) files can be determined from publications. Table 15 shows relevant data for corporations (1979) and sole proprietorships (1977). There are very few unclassified returns. Partial classification is more common for sole proprietorships than for corporations, especially when one considers that the figures for corporations are an overstatement (see note a to Table 15). The 1979 data for partnerships (not included in the table), in striking contrast to those for corporations and sole proprietorships, show that the proportion of unclassified and partially classified cases combined is somewhat less than 0.1 percent.

Current data are not available on incomplete classification of businesses included in the IRS individual and business master files. However, the proportions unclassified and partially classified are probably considerably higher than in the SOI files. As stated earlier in this chapter, it is known that for tax year 1969 there were no valid industry codes in the master file for 20.1 percent of the sole proprietorships, and that for tax year 1971 there were no valid industry codes for 9.1 percent of the partnerships. These figures include both returns that were completely unclassified by industry and those that were assigned invalid codes. Codes for "not allocable" within SIC division are not used in industry coding for the master files.

Further evidence on the trend in the proportion of unclassified sole proprietors is found in an article by Levine (1980). The SSA, as part of its Continuous Work History Sample (CWHS) system, maintains a longitudinal 1 percent sample file of self-employed workers with data on their earnings. The percent of workers unclassified by industry in this file averaged 4.9 from 1960 to 1969; however, in the following 6 years (1970-1975) it averaged 14.6, with a high of 21.3 percent in 1975. Levine (1980:475) explains this increase as follows:

TABLE 15 IRS Statistics of Income Program: Number of Incompletely Classified Returns, by Industry Division and Type of Organization

Type of Organization and Industry Classification	Percent of all Returns for This Type of Organization
CORPORATIONS (1979)[a]	
Partially classified[b]	1.7
Manufacturing, miscellaneous and not allocable[b]	0.5
Wholesale, miscellaneous and not allocable[b]	1.1
Wholesale and retail, not allocable	0.1
Unclassified	0.5
SOLE PROPRIETORSHIPS (1977)[c]	
Partially classified	3.3
Farms, not allocable	1.3
Construction, not allocable	0.5
Manufacturing, not allocable	d
Wholesale, not allocable	0.4
Retail, not allocable	0.3
Wholesale and retail, not allocable	0.8
Unclassified	0.3

[a]Data from Internal Revenue Service (1982b:76-78).
[b]The figures for these categories are overstated since they include some fully classified returns in SIC major group 39 (miscellaneous manufacturing industries) and industry group 509 (miscellaneous durable goods) and 519 (miscellaneous nondurable goods).
[c]Data from Internal Revenue Service (1981:14-18).
[d]Less than 0.05 percent.

> . . . before 1968 SSA received the schedule SE's from IRS and assembled the file as a routine part of CWHS processing. Subsequent to 1968, however, IRS began to transmit the SE data on magnetic tape and problem resolution was difficult or impossible.

By taking advantage of the longitudinal nature of the file for imputation, SSA was able to reduce the final percents of unclassified cases considerably.

Social Security Administration

According to the system description prepared for the Industry Coding Working Group, about 7.5 percent of the total records in the single-unit employer identification file as of December 1979 were completely unclassified. No data were given on the proportion of partially classified units, nor was a separate figure available for active employers. There was no corresponding figure available for reporting units in the multiunit employer identification file.

Data from a matching operation following the 1963 economic censuses presented above (Table 8) showed that 279,000 of 1,958,000 establishments (14.2 percent) included in the censuses and matched to SSA records had not been fully classified, i.e., to the four-digit level, by SSA.

Finally, data from the CWHS (Bureau of Economic Analysis, 1976) show that only 1.2 percent of the wage and salary workers in the 1 percent sample were unclassified by industry in the final version of the file for the first quarter of 1972. This suggests that the 7.5 percent of the establishments that were unclassified at the end of 1979 were small or inactive, although some of the difference could be accounted for by a larger proportion of unclassified employers among those added to the system since 1972.

Response Error

There have been a few studies in which industry codes initially assigned have been checked on the basis of additional information obtained from respondents. Reinterview studies of this kind may provide estimates of response bias, response variance, or some combination of these two components of error. All such studies located for use in this report were conducted by the Census Bureau.

Recheck Study

In 1948, the Census Bureau (1951) conducted a "retail trade industry code recheck." A sample of 535 retail trade establishments from the monthly survey were reinterviewed after about two months. Somewhat more detailed information was obtained on each establishment's

sales by merchandise line. In particular, the recheck obtained percent of sales for each of four principal merchandise lines; the initial interview had only called for a listing, in order of importance, of the three principal merchandise lines. Four-digit (and in a few industries more detailed) SIC codes were assigned on the basis of the recheck data without reference to the original questionnaires and codes.

Code differences were found for 98 establishments, 18 percent of the total included in the study. Results of an analysis of the reasons for difference are shown in Table 16. About two-thirds resulted from differences in the information in the original and recheck questionnaires, presumably resulting from the more detailed data requirements of the latter. The study stated that commodity breakdowns with percentages were helpful or necessary for proper coding in 22 of the 98 cases with differences.

Employer Record-Check Studies

The evaluation of industry classification in the employer record check of the 1960 population census (described in the previous section of this chapter) was carried out by comparing industry codes of employed persons based on information reported in the census with industry codes for their employers available in SSA files. A second employer record check was carried out following the 1970 population census, using a different procedure (Bureau of the Census, 1977a). Employers of the sample of 6,245 persons included in the study were asked to provide information about the principal activities, products and services of their establishments; and industry codes based on this

TABLE 16 Reasons for Industry Code Differences Between Initial and Recheck Surveys: Retail Trade Surveys, 1948

Reason Attributed for Difference	No. of Total Cases	Percent of Differences
Informational differences	67	68
Coding differences (same information)	25	26
Miscellaneous problems	6	6
Total	98	100

Source: Bureau of the Census (1951:6).

information were compared with those assigned to the same persons from information reported by or for them in the population census.

The principal statistic used to present data on gross error or response variability from the 1960 and 1970 employer record checks was the index of inconsistency. This index takes on values from 0 to 100, with the higher values indicating greater response variability. It is calculated from the two-by-two table showing the number of units classified in and out of a particular category, e.g., a major industry, in the census and as determined by the record check. The L-fold index of inconsistency is a weighted average of indices for individual categories in a distribution. Further details on calculation and interpretation of the index of inconsistency are given in the reports of the two employer record checks (Bureau of the Census, 1965a, 1977a).

Table 17 shows the indexes of inconsistency by "major industry" (roughly equivalent to SIC division) from the 1960 and 1970 employer record checks. Clearly, wholesale trade was subject to large response error in both censuses. As stated in the report on the 1970 study (Bureau of the Census, 1977a:4):

> This industry has classification problems in two directions. In some cases there is confusion as to whether the case should be manufacturing or wholesale trade. In other cases the confusion is between wholesale and retail trade.

Table 17 also shows that the indexes of inconsistency by industry were lower in 1970 than in 1960. The possible reasons for this change are not discussed directly in the report, except for a brief statement in the "Highlights" section (Bureau of the Census, 1977:4):

> On the whole, the reporting of occupation in the 1970 census was no better nor worse than the reporting in the 1960 census. There did appear to be some improvement in the reporting of industry.

The hypothesis of better "reporting" in 1970 does not seem very tenable, as the industry inquiries in the two censuses were nearly identical, and the collection procedures were similar, although self-enumeration was used somewhat more in 1970.

More likely, the difference resulted from changes in the coding and related processing procedures between 1960 and 1970 or from differences in the procedures used in the record-check studies, or both. Detailed information on differences in processing procedures in the 1960 and 1970 censuses is not available in published form, but significant changes could have occurred: in the training of coders; in the quality and coverage of reference materials, such as company name lists, available to coders; in the effectiveness of quality control procedures; and in the computer edits used to eliminate impossible or unlikely industry codes. The basic difference in the record-check procedures was the collection of the source data

TABLE 17 Indexes of Inconsistency for Selected Major Industries: 1960 and 1970

Employer Classification (ERC)	1960		1970	
	Index of Inconsistency	95-Percent Confidence Interval for Index of Inconsistency	Index of Inconsistency	95-Percent Confidence Interval for Index of Inconsistency
MAJOR INDUSTRY				
L-fold index	19	16.3 to 21.8	14	12.1 to 15.7
Mining	--[a]	--[a]	19	9.7 to 35.8
Construction	20	13.4 to 29.2	9	5.8 to 13.2
Durable goods manufacturing	14	10.8 to 17.4	10	7.7 to 11.9
Nondurable goods manufacturing	17	13.2 to 20.9	14	11.2 to 16.9
Wholesale trade	51	40.3 to 63.4	32	26.0 to 39.9
Retail trade	14	10.7 to 18.1	12	9.7 to 15.0
Business and repair services	32	20.0 to 50.8	18	11.9 to 26.3

[a]Does not meet publication standards.

Source: Bureau of the Census (1977a:6).

for industry classification directly from employers in 1970, as opposed to the use of SSA industry codes in 1960. It is not possible to say with confidence which of these methods provides a better standard for evaluation of industry codes assigned in the census; however, there are at least two points that would appear to favor the direct approach:

(1) As discussed earlier in this report, the updating of SSA's codes to reflect activity changes is incomplete and done with considerable time lag. Thus the direct approach provides more current information for classifying by industry.

(2) The direct approach includes collection of data on each sample person's occupation, which may sometimes be helpful in determining the correct industry.

If, in fact, the 1970 recheck codes were more accurate than those used in 1960, the higher indexes of inconsistency observed in 1960 may have resulted, in part, from errors in the recheck codes.

1977 Economic Census Evaluation Studies

Several evaluation studies conducted in connection with the 1977 economic censuses provide information about the quality of industry codes obtained by the Census Bureau from administrative record sources (Bailar and Kallek, 1980). These studies primarily covered three types of establishments: (1) those classified on the basis of administrative records as being outside the scope of the economic censuses; (2) those within scope, but designated as nonemployers and therefore excluded from the mail portion of the census; for the most part, data for these establishments were obtained from tax returns; and (3) those within scope and having employment, but with employment below designated cutoffs, which varied by industry; only a sample of these establishments was included in the mail portion of the census.

The technique used in the studies of each type of establishment was to mail economic census questionnaires to a sample of units. The returned questionnaires were used to evaluate the accuracy of census information, including industry codes, that was derived from administrative record sources. Indirectly, therefore, these studies provide information on the quality of industry codes in the IRS amd SSA systems; however, the emphasis in the reports of the studies is on the accuracy of economic census results, regardless of their source.

Hanczaryk and Sullivan (1980) report on a study of active establishments with employees included in the SSEL but defined as being out of scope of the economic censuses. The study universe comprised about 558,000 establishments. Of the total, about 77 percent were out of scope because they were classified in SIC industries not included in the economic censuses. Most of the remainder were government organizations, and a few represented units located abroad or in U. S. territories and possssions. A sample was

selected from this population, and copies of the economic censuses General Schedule (NC-X4) were mailed to 5,505 units that were not clearly out of scope.

The returns were classified by industry, and it was then possible to estimate that about 17,200 establishments in the study population were actually in the scope of the economic censuses. This was 3.1 percent of the establishments clasified as out of scope, and they accounted for 0.4 percent of total employees and 0.3 percent of payroll for this group. If these establishments had been included in the censuses, census totals would have been increased by 0.5 percent for number of establishments and 0.2 percent for number of employees and total payroll.

Three other evaluation studies are reported by King and Ricketts (1980). The first two were based on mailings of census questionnaires to samples of nonemployers and "employers below cutoff" classified in the retail trade and service divisions on the basis of administrative record sources. The samples were approximately 10,000 nonemployers and 103,000 employers. Table 18 shows the results of comparing SIC classifications based on census questionnaires with those based on administrative records for the same establishments in the two studies. The percent of agreement was higher for service industries than for retail trade in both studies. Agreement rates for employers below cutoff were considerably better than for nonemployers.

TABLE 18 Comparison of SIC Codes Based on Census Questionnaires with Those Based on Administrative Records: 1977 Economic Censuses

Type of Establishment and SIC Division[b]	Weighted[a] Percent Agreement at Division Level	Two-Digit Level	Three-Digit Level	Four-Digit Level
Nonemployers[c]				
Retail trade	69.8	58.0	46.7	n.a.
Service	79.1	70.0	n.a.	n.a.
Employers below cutoff[d]				
Retail trade	95.8	89.6	85.0	81.3
Service	97.4	96.1	94.1	94.1

n.a.: not available

[a]Weighted to reflect varying sampling rates used.
[b]Division per administrative record code.
[c]Data from King and Ricketts (1980:Table I.1).
[d]Data from King and Ricketts (1980:Table II.1).

Administrative codes for nonemployers are primarily those supplied by IRS; for employers most of the codes come from SSA or from internal Census Bureau programs.

The third study reported by King and Ricketts (1980) was a study of nonemployers administratively classified in construction. Economic census questionnaires were mailed to 2,610 cases selected from this population. The relevant results from this study, some of which are shown in Table 19, are presented somewhat differently; they show the net effects of classification changes on the totals by major group. Overall, there was a net reduction of 12 percent in the number of nonemployer establishments in construction. About half of the changes resulted from the removal of duplicate listings from the census control lists, but the remainder (net) was the result of changes in industry classification.

Finally, King and Ricketts report on a similar study of employers in construction who did not return the Census Bureau mail questionnaires. Data were collected for a sample of this group by telephone, and the results were analyzed in the same way as those from the other construction study. The relative net change in total number of employers, including respondents, was minus 1 percent, and the relative net changes by major group, as might be expected, were considerably smaller than those for nonemployers.

Processing Error

The systems descriptions prepared for the Industry Coding Working Group contained very little quantitative information on errors occurring in manual and automated stages of industry coding. One exception was the IRS Statistics of Income Program industry coding system for sole proprietorships. Records from dependent sample verification of industry coding for tax year 1980 showed the following results (unweighted):

Type of Business	Error Rate (percent)	Range for 10 Service Centers (percent)
Nonfarm	0.9	0.1 to 2.5
Farm	0.9	0.0 to 4.9

System descriptions for SSA's single and multiunit industry coding both stated that "audits" (based on sample verification) conducted by SSA's Office of Research and Statistics show approximately 97 percent accuracy in assignment of codes. Since these audits are conducted on cases that have already been subjected to "peer review," which is also conducted for a sample of cases (10 percent for the multiunit system), it seems likely that the overall outgoing quality is somewhat lower than 97 percent.

TABLE 19 Evaluation of Published Statistics for Nonemployers in Contract Construction: 1977 Census

Category	Number of Establishments (000)			
	SIC 15[a]	SIC 16[b]	SIC 17[c]	Total
Published	708	130	24	554
Changes				
Decreases				
Duplication with employers	42	9	1	32
Reclassified as nonconstruction	86	11	4	71
Reclassified to other construction	80	36	16	27
Increases				
Reclassified from nonconstruction[d]	41	14	1	26
Reclassified from other construction	80	28	2	49
Revised total	621	116	6	499
Net change	-86	-14	-18	-55
Percent change	-12%	-11%	-75%	-10%

Note: Detail may not add to total due to rounding.

[a]SIC 15: General contractors and operative builders.
[b]SIC 16: General contractors other than builders.
[c]SIC 17: Special trade contractors.
[d]Understated because only retail trade and service industries provided cases.

Source: King and Ricketts (1980:178).

No data on processing errors were included in the systems description for the BLS's ES-202 industry coding, which is done by state offices. Boyes and Brown (1974) report on plans for a study of coding reliability based on independent coding of a sample of state product reports, but no results from that study have been published.

Turning once again to the coding of industry for employed persons, there was a carefully designed study of coder effects in the 1960 population census (Bureau of the Census, 1972a). This study, which was based on comparison of codes entered on the original and a duplicate set of census questionnaires by the original census coders and the special coders, respectively, measured both the simple and

correlated components of coder variance. It did not provide estimates of biases common to the original and special coders. The results showed that both simple and correlated coder variances, especially the latter, were quite small in relation to response variances for the same items, measured in other studies that were part of the 1960 Census Evaluation Program. The data were presented primarily at the SIC division level, showed a familiar result: the largest indexes of inconsistency are for wholesale trade, closely followed by business and repair services. The two-way tabulations show relatively large shifts between wholesale trade and manufacturing and between wholesale and retail trade.

Data on Sources of Codes

It seems reasonable to suppose that when the industry codes in a file come from several different sources, their quality may vary by source. Thus the distribution of industry codes in a file by source could be considered an indirect indicator of quality.

Such information is available for single-unit establishments in the SSEL and is shown in Table 20. (Industry codes for multiunit establishments virtually all come from the economic censuses or from current surveys of the Census Bureau.) The first seven SIC divisions listed in the table are those that are fully or partly included in the economic censuses. The out-of-scope division includes two groups: about 482,000 establishments in SIC divisions B (mining) through I (services) in industries not included in the economic censuses and 133,000 establishments in agriculture, government, or located abroad.

The industry codes for establishments in columns (1) and (2) are based on questionnaires from economic censuses and surveys. Codes from census sources account for 68.5 percent of the in-scope establishments and 53.7 percent of the classified out-of-scope establishments. The next largest source is SSA's single-unit file, from which establishment birth listings are provided monthly to the Census Bureau. Industry codes came from this source for 26.4 percent of the in-scope and 38.5 percent of the out-of-scope establishments. Relatively small proportions came from the IRS master files: 3.2 percent of the in-scope and 5.5 percent of the out-of-scope establishments. The remaining cases were classified by industry on the basis of commercial lists or name coding, accounting for 2.0 percent of the in-scope and 2.3 percent of the out-of-scope establishments.

It would be interesting to see how variables such as employment, payroll, and receipts, are distributed by industry source code. Data have not been published, but from other Census Bureau (1982a) data, it can be noted that the division with the highest proportion of codes from Census Bureau sources--manufacturing, with 82.7 percent--has an average of 19.6 employees per single-unit establishment, while the division with the lowest proportion of Census-based codes--construction, with 54.4 percent--averages only 6.8 employees per establishment. Furthermore, virtually all of the industry codes

for establishments in multiunit enterprises, which accounted in 1979 for about 54 percent of total employment, are based on economic censuses or current Census Bureau surveys.

No comparable data are available for other systems. The two SSA files carry source and date codes for each employer's industry classification, but tabulations showing the distribution of currently active employers classified by industry source and date codes are not available.

NOTES

1. For a complete analysis, the units in a particular cell from data set A must be matched against all units in data set B, and vice versa.
2. The index of inconsistency is a measure of response variability, which ranges from 0 to 100. For further discussion, see section "Response Error" below.

TABLE 20 Single-Unit Establishments in the SSEL with Current-Year Payroll, by SIC Division and Source: 1981

SIC Divisions	Source of Industry Code: Economic Censuses (1)	Census Bureau Surveys (2)	IRS (3)	SSA (4)	Dun & Bradstreet (5)	Clerical Name Coding (6)	Computer Name Coding (7)	Not Classified (8)	Total
Part A. Estimated Number of Establishments									
Mining	7,254	7,610	1,552	6,884	458	152	201	--	24,111
Construction	242,666	29,706	15,804	182,387	6,620	2,959	20,429	--	500,571
Manufacturing	115,356	108,409	4,655	37,248	3,119	805	953	--	270,545
Transportation	22,223	24,822	1,951	26,064	550	200	600	--	76,410
Wholesale	199,588	18,275	12,723	58,890	2,130	766	703	--	293,075
Retail	492,657	195,522	33,743	266,645	4,244	3,094	2,318	--	998,223
Services	708,842	157,217	37,994	319,174	2,762	8,983	4,938	--	1,239,910
Out of scope	298,644	32,006	34,175	236,792	8,674	3,655	1,858	--	615,804
Not classified[a]	--	--	--	--	--	--	--	328,526	328,526
Total	2,087,230	573,567	142,597	1,134,084	28,557	20,614	32,000	328,526	4,347,175

Part B. Percent of Industry Codes from Each Source, by SIC Division

Mining	30.1	31.6	6.4	28.6	1.9	0.6	0.8	--	100.0
Construction	48.5	5.9	3.2	36.4	1.3	0.6	4.1	--	100.0
Manufacturing	42.6	40.1	1.7	13.8	1.2	0.3	0.4	--	100.0
Transportation	29.1	32.5	2.6	34.1	0.7	0.3	0.8	--	100.0
Wholesale	68.1	6.2	4.3	20.1	0.7	0.3	0.2	--	100.0
Retail	49.4	19.6	3.4	26.7	0.4	0.3	0.2	--	100.0
Services	57.2	12.7	3.1	25.7	0.2	0.7	0.4	--	100.0
Total in scope	52.6	15.9	3.2	26.4	0.6	0.5	0.9	--	100.0
Out of scope	48.5	5.2	5.5	38.5	1.4	0.6	0.3	--	100.0

[a]May include some cases that are classified out of scope but have no source code.

Source: Bureau of the Census (1982b).

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes the present status of major industry classification and coding activities under federal and federal-state auspices and discusses steps that might be taken to improve the comparability and accuracy of industry codes in different data systems. The current situation is summarized in the first section. The second section covers recommendations for improvements that would be achievable within the existing framework for interagency sharing of industry codes, i.e., without any significant changes in the laws and policies governing code sharing. The third section looks at possibilities for increased interagency sharing of industry codes and some of the associated benefits and problems. The fourth section discusses the need for operational and experimental data to evaluate existing systems and proposed alternative classification principles and coding procedures. The fifth and final section looks at the mechanisms presently available to encourage and implement interagency cooperative efforts and presents recommendations for achieving an integrated, cost-effective system of industrial classification and coding.

SUMMARY OF THE CURRENT SITUATION

As pointed out by Farrell et al. (1982), four agencies--BLS, Census Bureau, IRS, and SSA--are each responsible for a large annual volume of industry coding of establishments or other business enterprises. The average number of units coded annually ranges from roughly 900,000 for SSA (employers applying for EINs) to more than 15,000,000 for IRS (businesses reported on tax returns). The annual volumes are relatively constant for BLS, IRS, and SSA; for the Census Bureau they peak at the time of each quinquennial round of economic censuses.

There are significant differences in the universes of business establishments coded by these four agencies, with respect to both coverage and to the definitions of the basic coding units, but there is also substantial overlap. For most SIC industries, single-unit establishments with employees are coded by all four agencies. The coverage of coding of multiunit employers by BLS and SSA is similar. The Census Bureau's SSEL contains sufficient establishment information

to assign industry codes to most reporting units of the multiunit employers coded by BLS and SSA and to the corporations coded by IRS.

All four agencies use coding systems based on the SIC; however, each of them departs from it in one or more respects. There are substantial differences, both between and within agencies, in the classification principles followed, in the source documents used to obtain information for industry coding, and in the actual coding procedures. Hence it is not surprising that the available evidence (presented in Chapter 5) shows significant differences between systems, both in aggregate data and in the classification of individual units coded in two or more systems.

The different industry coding systems vary widely in the level of effort applied to achieve accurate classifications. That variation is reflected by differences in many system characteristics: the amounts of detail called for in source documents; the use or nonuse of specialized source documents; the techniques used to deal with missing information; the experience, training, and supervision of manual coding clerks; and the kinds of computer consistency checks performed. It seems reasonable to expect a positive correlation between the resources applied and the degree of accuracy achieved, and there is some evidence of such a correlation.

At present, the only significant code sharing among these four agencies results from the periodic transmittal of industry-coded lists from IRS and SSA to the Census Bureau for use in adding new units to the SSEL and for coverage of small employers and nonemployers in the economic censuses. The IRS and SSA industry codes are retained in the SSEL only for those units for which a code based on an economic census or current survey is not yet available. There is no reciprocal transmittal of industry codes from the Census Bureau to other agencies.

What can be done to improve the comparability and accuracy of codes in different systems and to reduce the overall costs of industry classification? The major thesis of this report, as stated in Chapter 1, is that while some modest improvements in comparability and efficiency could be achieved by changes in individual systems, significant overall gains require an increase in the amount of code sharing between agencies.

Legislation to permit greater sharing of individually identifiable records for statistical purposes, under strong safeguards of confidentiality, was drafted early in 1983 by OMB and was circulated informally for review to members of the Working Group on Economic Statistics of the Cabinet Council on Economic Affairs (Office of Management and Budget, 1983). The draft legislation (sometimes referred to as the "enclaves bill") was rather broad in scope, but it included specific provisions to allow agencies other than the Census Bureau to use the SSEL for statistical purposes. In November 1983, however, a decision was made by then-White House counselor Edwin Meese to abandon administration support of the proposed legislation. It was not clear whether the administration would support narrower legislation covering only the sharing of information in the SSEL; however, one cannot be optimistic about the near-term prospects for development of such legislation. Therefore, attention is first given

to improvements that do not require increased code sharing. In the longer run, however, some of the changes suggested might make code sharing more effective if and when laws and policies allow it to take place.

IMPROVEMENTS NOT REQUIRING ADDITIONAL CODE SHARING

The first part of this section covers the industry classification principles followed by the agencies whose industry coding systems were reviewed for this report. The recommendations in this part have as their primary goal greater standardization of classification principles, leading to greater comparability between systems and agencies. The second part of this section covers the coding procedures used, including such aspects as updating schedules, source documents, instruction and reference materials, and processing procedures. Some of the recommendations in this part are also for increased standardization; others are aimed at improvement of quality or reduction of costs. Each recommendation is first presented in general terms; when appropriate, reference is then made to principles and practices of specific agencies and systems.

Other improvements that do not necessarily require code sharing come under the general headings of methodological research and evaluation and interagency coordination. They are of special importance and are treated separately in the last two sections of this chapter.

Classification Principles

Basic Coding Units

Recommendation 1. The definitions of units such as establishments and reporting units that are defined for statistical purposes should be standardized.

The establishment definition in the SIC Manual is somewhat ambiguous; an interagency technical group should explore the possibility of developing specific rules, with illustrations, that would reduce the ambiguity. There is no obvious reason for the differences in the reporting unit definitions used by BLS and SSA (see Chapter 2, "Classification Principles"). The two agencies should try to agree on a single definition.

Use of the SIC and ESIC

Recommendation 2. All federal agencies that classify establishments and other business units by activity for statistical purposes should base their classification on the Standard Industrial Classification (SIC) or the Enterprise Standard Industrial Classification (ESIC), as appropriate.

This recommendation is not intended to preclude the use of additional classifiers for the same units; however, other classifiers used should be clearly distinguishable from the SIC. A clear departure from this policy was identified (see Chapter 2) in connection with data compiled by the Energy Information Administration on marketing of petroleum products: establishments selling primarily to "ultimate consumers" are treated as retail, whether the consumers are in the business or household sector.

Coding to the Four-Digit (SIC Industry) Level

Recommendation 3. Unless clearly contraindicated by cost factors, all systems should code to the full four-digit level. If groupings are made, each four-digit industry should be assigned to a single group, and groupings should adhere strictly to the SIC structure of divisions and two- and three-digit industries. Schemes used for partial coding when input data are incomplete should be reviewed and standardized.

As explained in Chapter 2, "Definition of Basic Coding Units," failure to code uniformly to the same level in all systems can reduce comparability between systems, even for higher-level categories that appear in all systems. In addition, as explained in Chapter 2, "The Effect of SIC Revisions on Comparability," conversion of codes to conform to periodic SIC revisions is made more difficult by the use of grouped industries.

BLS and SSA both use most but not all of the SIC four-digit industries. Together with the Census Bureau, these agencies should try to agree on a standard set of four-digit industries that all would use as their basic categories for the publication of data by industry. Any further splits should be made within the industry categories in that set.

The IRS uses different SIC industry groupings for sole proprietorships, partnerships, and corporations, with each set having about 200 categories, as compared with 1,005 in the 1972 SIC. Switching to full four-digit SIC coding would be a drastic change; the costs and benefits would have to be evaluated carefully. The additional costs might be covered by coding less than annually. If the coding were done in quinquennial economic census years, part of the cost could reasonably be charged to the censuses. The effects of such changes on IRS's nonstatistical uses of industry codes are not obvious, because the details of these uses are administratively classified.

Minor changes in the IRS classification schemes are clearly in order. The kinds of splits and regroupings of SIC four-digit industry codes described in Chapter 2, "Adherence to SIC Categories," should be eliminated, and partial codes should not be combined with full codes for "not elsewhere classified" groups in publications (see Chapter 4, "Partial Coding").

Treatment of Multiple Activities

Recommendation 4. Agencies using the SIC and ESIC systems should agree on and adopt standard classification principles covering the treatment of multiple activities.

Specific departures from standard practice that should be reviewed and eliminated wherever possible (described in Chapter 2, "Treatment of Multiple Activities") include:

(1) Departures from the recommendations in the SIC Manual for the measures of size to be used for activities in each of the SIC divisions in assigning industry codes. The most striking of these departures is IRS's use of assets in assigning industry codes for partnerships. The Census Bureau's SSEL comes closest to using the measures recommended for each division; other agencies tend to use a single measure of size for all or most SIC divisions.

(2) Departures from the basic rule of assigning the code for the four-digit industry accounting for the largest proportion of total activity, without regard to relative shares at higher levels. The SSA assigns an industry code in manufacturing if the associated percentage is 20 percent or more. The Census Bureau, in its Enterprise Statistics Program, uses a top-down approach, which ensures that the detailed code will be assigned in the SIC division with the largest share of total activity.

Resistance Coding

Recommendation 5. All agencies that apply resistance coding rules in updating industry codes should document their rules.

Recommendation 6. Agency practices on resistence coding should be reviewed, with the participation of users, and standards should be developed to promote uniformity in determining when and how resistance principles should be applied.

Documentation

Recommendation 7. All classification principles that are used should be fully documented, including especially those relating to treatment of multiple activities and resistance coding.

Any principles that are either in addition to or contrary to those currently in the SIC Manual should be clearly described in publications presenting data by industry. Classification rules embedded in programs for computerized coding should be documented in a form that makes them accessible to data users.

Coding Procedures

Updating Schedules

Recommendation 8. Agencies should consider whether their updating schedules are optimal or nearly so, taking into account costs and accuracy requirements.

In substantial parts of the SSA employer universe, the industry codes have not been updated for 10 years or more. At the other end of the scale, IRS updates all taxpayer industry codes annually.

Source Documents

Recommendation 9. Activity questions included on source documents as a basis for industry classification should be standardized to the greatest degree possible.

As can be seen from the examples in Appendix B, there is now wide variation in activity questions. The selection of standard items should be based on careful testing of alternatives (see below). Special attention should be given to items designed to distinguish between wholesale and retail trade.

Improving and Controlling the Quality of Manual Coding

Recommendation 10. All manual coding operations should incorporate formal training for new coders, using appropriate instructions and reference materials in addition to the SIC Manual.

Recommendation 11. Formal quality control systems should be adopted if they do not presently exist.

The use of independent, as opposed to dependent, verification procedures should be considered (see Chapter 4, "Factors Affecting the Quality of Manual Coding," for further discussion).

Computer Consistency Checks

Recommendation 12. Each system should incorporate, as a minimum requirement, a computer consistency check to ensure that no invalid industry codes are retained in the system.

Documentation of Coding Procedures and Outputs

Recommendation 13. Imputation procedures, such as "force coding" (discussed in Chapter 4) should be described in publications. Results

of quality control checks and evaluation studies of manual coding operations and of computer consistency checks should be systematically documented and made available to users. Cumulative files that contain industry codes should show the date of the most recent update for each unit and, if relevant, the source.

In some cases it may be desirable to show more than one source code in order to avoid unnecessary restrictions on access. For cumulative files, periodic tabulations of active units (assuming there is some basis for identifying currently active units) should be made to show their distribution by source and date of industry classification and the number and characteristics of incompletely classified units.

IMPROVEMENTS THROUGH ADDITIONAL CODE SHARING

Increased code sharing could lead to greater comparability of industry codes in major federal and federal-state data systems. Depending in part on the priority rules and other reconciliation procedures adopted when different systems are found to have different codes for identical units, it is likely that the accuracy of codes in all the systems could be improved. Initially, there would be significant costs for the development of processing systems to match units in different agency files and to deal with those cases in which the industry codes or the units fail to match; however, once these processing systems were established, savings could be achieved by cutting back on data collection activities now carried out by all the agencies to provide input to largely independent systems for assigning and updating industry codes.

The SSEL is the logical starting point for expanded code-sharing activities. It receives inputs from the Census Bureau's economic censuses and surveys and from the SSA and IRS administrative record files. For each establishment, the industry code is taken from what is judged to be the best available source. The economic censuses are given top priority, followed by other Census Bureau sources (current surveys), then by SSA, and finally by IRS. Based on a comparative analysis of the relevant characteristics of these systems, these priorities seem reasonable.

The one major system that is not currently being tapped by the SSEL is BLS's federal-state employment and wages system. For single-unit employers not covered in the mail part of the economic censuses or in certain of the Census Bureau current economic surveys, the BLS industry codes are probably superior, on average, to those available from SSA or IRS. They are based on more detailed source documents and, unlike the SSA codes, are generally updated every 3 years. There is one technical obstacle to the use of BLS industry codes in the SSEL: not all of the states use the employer identification number (EIN) as a state reporting number for the unemployment insurance (UI) program; and some of the states that do not use the EIN do not include it in the employer records made available to BLS for its UI name and address file. The ideal way to solve this problem would be to require

all states to use the EIN; failing that, it might be possible to establish the necessary linkages by using the IRS Form 940 (Employer's Annual Federal Unemployment Tax Return), which carries both the EIN and UI numbers.

With the addition of BLS source data, the SSEL industry codes could represent a synthesis of the best available information from all major sources. However, the full benefits will not be realized until the exchange between the Census Bureau and other agencies becomes a two-way process, with SSEL information going from the Census Bureau to other federal and state agencies as needed for statistical purposes. This exchange process should include transfer of industry codes to IRS for use in its SOI record systems, but not for inclusion in the individual and business master files, which are used for compliance activities; and there should be an adequate guarantee that the SOI systems would be fully insulated from compliance activities.

The SSEL has the further advantage that its basic unit is the establishment, which is the smallest unit used for most statistical purposes. It also contains the necessary identifiers that generally make it possible (although not always easy) to link data for establishments belonging to larger units, such as employers, reporting units, companies, or corporations, that are the basic units in systems of other agencies.

The establishment of such a multiway exchange network would not immediately produce a statistical utopia. As this report has made clear, there are numerous differences between agencies in coverage and coding unit definitions. Nevertheless, there is substantial overlap, and roughly 80 percent of the employer establishments are single-unit enterprises, defined essentially the same way in all systems. Given reasonable agreement on classification principles for those units, there should be no serious technical obstacles to effective code-sharing arrangements for this part of the universe of establishments.

Once the legal and technical problems are largely solved, the way would be open for something much closer to an ideal system, in which each agency would provide the inputs it is best equipped to provide and the needs of each would be met in the most efficient way possible. Roughly, this might work as follows:

(1) The Census Bureau would provide industry codes for all in-scope establishments canvassed by mail in the quinquennial economic censuses. Between censuses, codes for the larger of these establishments and for new establishments of multi-unit employers would be obtained from the annual Company Organization Survey (COS) and other current surveys. The Census Bureau should consider collecting somewhat more detailed information in the COS for use in updating industry codes of existing establishments.

(2) The Bureau of Labor Statistics would provide industry codes for all other employer establishments. It would probably be sufficient to update these codes only at the time of each round of economic censuses.

(3) The Social Security Administration would continue to identify employer births (new establishments), and its industry codes would be used until replaced by Census Bureau or BLS codes.

(4) The IRS would provide industry codes for nonemployer establishments (and data for these establishments) at the time of each round of economic censuses. Unless codes were clearly needed for compliance purposes, IRS would not code these establishments in noncensus years. In order to meet economic census requirements, the coding in census years should be to the full SIC four-digit industry level.

Once fully established, there is little question that such a system of exchanging industry codes could provide better quality and comparability at a lower overall cost.

METHODOLOGICAL AND EVALUATION RESEARCH

Research Needs

Evaluation and methodological research in connection with statistical programs usually has two main objectives: (1) to inform data users about the accuracy of the data available to them and (2) to enable data producers to develop and evaluate procedures for improving accuracy and efficiency. For research on industry classification and coding, a third objective might be to improve the comparability of data from different systems. Specific research questions can relate either to industry classification principles or to coding procedures.

Classification Principles

Questions about classification principles may arise because an agency follows principles different from those recommended by the SIC Manual or because certain aspects of classification are not fully dealt with in the Manual, e.g., the use of resistance coding (discussed in Chapter 2, "Treatment of Changes Over Time"). Some examples of research questions follow:

- How is the distribution of establishment or enterprise data by industry affected by a "filter" or "top-down" approach to classification, as opposed to assigning the code for the single four-digit industry with the largest share of activity (see Chapter 2, "Treatment of Multiple Activities")?

- How is the distribution of enterprise data by industry affected by the choice between the two principal methods of classification: the aggregate approach and the establishment building-block approach (see Chapter 2, "Treatment of Multiple Activities")?

How is the distribution of data by industry affected by the choice of the variable used to measure the relative importance of different activities within an establishment or other unit? For example, how are industry codes for partnerships reporting to IRS affected by using assets rather than receipts as a measure of size?

How might the uses of time-series data by industry be affected by the use or nonuse of resistance coding principles or by particular formulas used for resistance coding?

It should be possible to answer most questions of this kind by analyses using existing files that contain data on the allocation of receipts and other variables within establishments by SIC industry categories.

Coding Procedures

For any industry coding system it is desirable to have measures of accuracy and information on relative sizes of the different kinds of errors that occur. For the optimum design of industry coding systems, it would also be useful to have answers to such questions as:

What is the relative accuracy of codes supplied directly by respondents (taxpayers, employers, etc.) and those assigned by coders on the basis of responses to various kinds of questions about economic activities?

How is accuracy of coding affected by increasing the amount of detailed information collected about a unit's activities? In particular, how is it affected by the use of specialized as opposed to general-purpose forms?

How is accuracy affected by the specific wording and response format of the relevant items on source documents? Which of the items used to distinguish wholesale and retail trade is most effective?

What is the extent of coverage and the relative accuracy of industry codes in various publicly and commercially available listings used to supplement information collected directly from the units to be coded?

How frequently do real changes occur in industry classification (as opposed to those resulting from response and processing errors)? What are the relative frequencies for various types of units? How do they vary by SIC division and major industry and by size of unit?

- How is the accuracy of updates affected by the use of "shuttle" procedures, i.e., giving respondents information about activities previously reported by them and asking whether these have changed?

In addition to information about how different procedures affect accuracy, information about their relative costs is, of course, necessary in order to make fully informed choices.

System measures of accuracy are of special importance in setting up priority rules in systems such as the SSEL, which have access to codes from more than one source. It is also of considerable importance in such cases to have information about the coverage of each system from which codes are available and about the relationships of the units used in different systems. Information of this kind, obtained through matching of individual units from different systems, serves a dual purpose: it aids in the development of effective code-sharing arrangements, and it may lead to improvements in individual data systems.

Research to Date

Results of past research relevant to industry classification and coding, based on available publications and unpublished reports, were presented in Chapter 5. This section presents an evaluation of the utility and adequacy of that research. A general impression is that the amount of relevant research, especially that involving joint studies by two or more agencies, has declined in recent years. Some of the most interesting and useful studies were conducted in the 1950s and early 1960s.

The majority of methodological studies for which reports could be located were conducted by the Census Bureau. Some additional studies were conducted by IRS. Taken as a whole, the results of the studies reviewed in Chapter 5 provide a reasonable amount of quantitative evidence on the relative accuracy and other characteristics of industry coding in the Census Bureau, IRS, and SSA systems, but very little comparable information for the BLS employment and wages system (except for a study covering only the retail trade SIC division for 1958).

Perhaps the most valuable studies, at least in terms of promoting comparability between agencies, have been those in which industry codes in the different systems were matched and the differences reconciled with the participation of each of the agencies involved. Not only do such studies help to judge the relative accuracy of different systems, they also often lead to the identification of significant problems requiring attention in the next revision of the SIC or interagency agreement on interpretation of the current version.

Studies that simply identify differences between industry codes in two systems, but do not reconcile them, are somewhat less useful, especially if both systems are judged on a priori grounds to have roughly the same level of accuracy. An example would be the

comparisons between revenue processing and SOI industry codes in IRS. In the case of nonfarm sole proprietorships, both coding systems depend on essentially the same limited source information from the Schedule C; the responses of many taxpayers are simply not detailed enough to classify a unit accurately. Nevertheless, one can look at the results of these studies and identify some of the problem areas, such as being able to accurately distinguish between wholesale and retail trade.

The evaluation studies by the Census Bureau following the 1977 economic censuses (see Chapter 5, "Response Error") provided useful indirect information on the accuracy of industry codes in the IRS and SSA systems, but only for smaller units and those in SIC industries not within the scope of the economic censuses.

Little has been reported on experimental research covering such matters as the format and content of source documents, respondent versus agency coding, or independent versus dependent updating procedures. A recent exception is the BLS work on use of a dependent updating procedure aimed at reducing response burden and coding costs (Hofstetter, 1983; see Chapter 3, "Examples of Source Documents").

In summary, there has been relatively little research relevant to industry classification and coding in recent years. There does not appear to be sufficient information available to develop a reasonably complete "error profile" for any of the industry coding systems covering business units. (For an example of an error profile for industry coding of persons, see Jabine and Tepping (1973).) If new opportunities for code sharing arise, there is little in the way of experimental data on which to base important decisions on coding procedures and source documents and on optimum procedures for integrating information from different systems.

Recommendations

Some of the information gaps that have been identified could be filled rather easily; others would require commitment of resources that are already overtaxed. An appropriate strategy is to proceed with the easy things and to evaluate the potential costs and benefits of more costly experiments and evaluation studies to establish priorities for the work that needs to be done.

The first category, i.e., the things that are relatively easy to do, includes some steps to ensure that existing data systems routinely provide data that can be used to evaluate the quality of industry codes and to plan system improvements.

Recommendation 14. Each cumulative system should carry as a miminum the following information for every unit: activity status; source of current industry code; date of assignment of current industry code.

The source and date should be based on the most recent update, even if the code was not changed. In addition, consideration should be given to carrying other information relevant to industry

classification, such as prior codes with sources and dates, and industry codes and measures of size for both primary and secondary activities.

Recommendation 15. Periodic tabulations should be made from each system to permit analyses of distribution of active units and aggregates (e.g., employment, receipts) by industry, by source, and date of most recent industry code update; and frequency and impact of partial or missing industry codes in the system, by industry (for partial codes) and by source.

Recommendation 16. For those systems that match in industry codes from other systems, the matching process should be designed to produce, as a by-product, tabulations showing the proportion of matched units in each system and the extent of agreement of industry codes for the matched units.

In large systems, it may be less expensive and nearly as useful to do some of the recommended tabulations above on a sample basis. A permanent "trace" or "master" sample, with selection based on identification numbers, may be an effective vehicle for keeping track of what is in the system and for doing special analyses of the type recommended.

In the second category, i.e., more formal evaluation studies and experiments, it is recommended that special consideration be given to several kinds of research.

Recommendation 17. Sample matching studies to compare industry codes, along with selected data items such as employment and payroll, in different agency files should be resumed.

Matching studies should be focused on the SIC divisions, such as wholesale and retail trade and, at the enterprise level, on large conglomerates for which classification is known to be especially difficult. The studies will be most useful if differences can be reconciled with participation of all of the agencies involved. A study comparing industry codes in the SSEL with those in the BLS employment and wages (ES-202) system would be of particular interest. Such a study would be facilitated if BLS could persuade all of the state agencies to include EINs in their standard employer and reporting unit records.

Recommendation 18. Methodological experiments should be undertaken, with emphasis on evaluation of alternate format and wording of relevant items on source documents; costs and accuracy associated with alternate procedures for updating industry codes; and the relative accuracy of self-coding versus agency coding, taking into account the problems associated with self-coding immediately following revision of a code structure.

Recommendation 19. A thorough evaluation of the costs and quality effects of using various commercial lists to supply industry codes, either as a primary or secondary source, should be carried out.

Recommendation 20. Studies should be done to evaluate alternative classification principles, e.g., those related to treatment of multiple activities, classification of units above the establishment level, and resistance coding.

Such studies would require the use of data bases with rather detailed information for all of the units involved, such as the SSEL, the economic censuses and, for resistance coding studies, longitudinal files from current surveys such as the Annual Survey of Manufactures.

STEPS TOWARD AN INTEGRATED SYSTEM OF INDUSTRY CLASSIFICATION AND CODING

The lack of comparability of industry codes in different systems is in large part an outcome of the decentralized structure of the U. S. statistical system and of the laws and policies that govern exchange of individually identifiable information among the agencies in the system. Over a period of roughly 25 years, from about 1940 to 1965, with the leadership of a strong central statistical coordination unit, considerable progress was made in refining the Standard Industrial Classification and reaching agreement between agencies on specific interpretations. Major statistical agencies were in close contact and were able to conduct several matching studies in order to compare and reconcile industry codes in different systems.

In the 20 years since then there have been significant setbacks. The central statistical coordination unit has diminished in size and authority, and the level of interagency cooperation depends largely on the willingness of the individual agencies to work together. Legal and policy restrictions on interagency data exchanges have become more stringent, to the extent that industry code sharing, whether for operational or research purposes, is severely limited. Within the Census Bureau, the long-time dream of an integrated business directory has been largely realized, but the SSEL lacks inputs from the Unemployment Insurance system, and is available only for internal agency use. The revision of the SIC scheduled for 1982 has been postponed until 1987, and there is little formal or informal consultation between agencies on questions of interpretation of the current version.

Reversing the Trend: Short-Term Initiatives

Several suggestions for improvements in individual industry coding systems were presented above ("Improvements Not Requiring Additional Code Sharing"). In the present circumstances, it will be largely up to individual agencies to decide on these matters, but it would also

be desirable for the Statistical Policy Office in the Office of Information and Regulatory Policy of OMB to continue and to expand its role. Since 1978, the Federal Committee on Statistical Methodology and the subcommittees and working groups operating under its auspices have made valuable contributions to understanding the issues discussed in this report (see, for example, Office of Federal Statistical Policy and Standards, 1980). To continue the momentum developed by these groups, several steps would be helpful.

Recommendation 21. The scope and intensity of the work of the OMB Technical Committee on Industrial Classification should be expanded.

A major goal for the work of the committee should be to promote more uniform interpretation of the current version of the SIC in the classification of government and auxiliary establishments, and of newly emerging economic activities. For federal government establishments, it should be possible for the committee to reach full agreement on industry codes for specific units since there should be no legal restrictions to discussion of the relevant characteristics of individually identifiable units. For other kinds of establishments, confidentiality requirements of the agencies participating may inhibit discussion of specific cases that illustrate problems in interpretation of the SIC. A second goal should be to plan for an orderly transition to the revised SIC in 1987. Many of the structural changes needed for the revision have already been identified; however, careful planning will be needed to coordinate the transition activities among the agencies that do significant amounts of industry classification, and to ensure that the necessary resources are made available.

Recommendation 22. Agencies should be encouraged to share technical information about their industry coding procedures.

The Technical Committee on Industrial Classification should take responsibility for arranging and disseminating periodic updates of the industry coding systems descriptions prepared by the Industry Coding Working Group. In addition, it would be useful to hold one or more workshops on the use of computers in industry coding, involving persons responsible for coding at the technical level. Topics covered should include coding based on computerized English-language activity descriptions or on quantitative product and service data, computer-assisted coding from activity descriptions, and consistency checks of or involving industry codes.

Recommendation 23. Agencies should be encouraged to undertake joint sample matching studies to clarify and try to resolve current differences between agencies in reporting unit definitions and industry classification.

Members of the Establishment Reporting Working Group of the Administrative Records Subcommittee strongly recommended that a study be undertaken, and a successor group, the Employer Reporting Unit

Match Study Work Group, is developing plans for such a study (Buckler et al., 1983).

Long-Term Requirements and Issues

The more difficult question is how to move toward a fully integrated business directory with uniform industry classification, such as was described in broad outline above ("Improvements Through Additional Code Sharing"). The obstacles are clearly formidable; otherwise, the need so clearly seen over 40 years ago by the Central Statistical Board would have been met long since.

The passage of legislation giving federal and state agencies access to the SSEL for statistical purposes, while a necessary first step, would not immediately bring about a fully integrated system. Answers will be needed to many other technical and policy questions. The technical issues have already been addressed; some of the major policy issues concern resources, organization, and statistical versus general use.

Resources

While a fully integrated system would eventually bring about substantial savings through the elimination of overlapping collection and processing activities, substantial resources would be needed for systems development.

Organization

It is not clear that user agencies will ever be willing to commit themselves to full dependence on a business directory operated by and basically under the control of one of the user agencies, i.e., the Bureau of the Census. It was largely for this reason that the President's Reorganization Project for the Federal Statistical System (1981) recommended that consideration be given to placing the SSEL in a central statistical office; lacking one, it would be desirable to seek ways to give both user agencies and those that provide data for the SSEL a genuine participatory role in management of the system.

Statistical Versus General Use

It is not obvious that all of the information in a business directory established and maintained by the government should be made available only for statistical purposes. Business lists classified by industry are clearly needed for many nonstatistical purposes. Does it make sense to establish completely separate systems to serve these purposes? In many countries, government business directories are published or otherwise released with no restrictions on use (American

Statistical Association, 1980). A possible response to this question would be to suggest going ahead with the statistical business directory and leaving the question of possible nonstatistical uses for later. However, it may be difficult to develop a fully effective system on this basis. It is important that the BLS federal-state employment and wages statistics system be included, but some of the states may be unwilling to accept the restrictions that have so far been included in all versions of proposed legislation covering the use of the SSEL. It should also be noted that any proposal to make some of the information in the SSEL available for unrestricted public use would bring forward the question of competition between the public and private sectors. What would be the impact of such a policy on the companies that maintain and sell industry-coded business lists?

The above discussion has departed considerably from the technical questions relating to comparability and accuracy of industry codes in different data systems, but in the last analysis it is difficult to separate technical and policy issues. The complexity of the task ahead dictates the need for the close cooperation and good will of all of the agencies concerned and for the coordination structure and resources essential to the development of a fully integrated system.

APPENDIX A

REPORTING UNIT DEFINITIONS, BUREAU OF LABOR STATISTICS AND SOCIAL SECURITY ADMINISTRATION

This appendix presents the reporting unit definitions used in two agency coding systems: BLS's employment and wages program (ES-202) system and SSA's multiunit employer identification file system. The differences between the two reporting unit definitions are discussed in Chapter 2, "Definition of Basic Coding Units."

The definitions are quoted directly from the system descriptions prepared by representatives of BLS and SSA for the Industry Coding Working Group. The BLS is developing a revised reporting unit definition that will further limit the conditions under which establishments can be combined to form reporting units; however, the new definition is not yet available.

BLS: EMPLOYMENT AND WAGES PROGRAM (ES-202) SYSTEM (EXCERPT)

Employees who are covered by the Unemployment Insurance (UI) laws of each State. Most employees except some agricultural workers, private household employees, members of the Armed Forces, and the self-employed are covered. Other exceptions have minimal effect on total coverage.

Reporting Units: The economic unit for which data are submitted on an employer's UI contribution report.

Qualifications for defining reporting units:

(1.) Industrial

(a.) A private or public multi-establishment employer who has fewer than 50 workers in all secondary industries combined need not submit an industry breakdown. Therefore, an employer meeting these criteria may report as a single unit.

(b.) A multi-establishment employer who has a total of 50 or more workers in all secondary industries combined is required to submit separate employment and wage data for

the primary and for each of the secondary industries. However, at the option of the State, data relative to any secondary industry at which fewer than six workers are employed need not be reported separately, but may be included under the employer's primary industry.

(2.) County

(a.) A private or public multi-county employer who has a total of fewer than 50 workers in all secondary counties combined need not submit a county breakdown, and all data relative to such an employer may be classified under the primary county. Primary county is defined as the county with the largest employment.

(b.) A multi-county employer who has a total of 50 or more covered workers in all secondary counties combined is required to submit separate employment and wage data for the primary and each of the secondary counties. However, data relative to any secondary county in which fewer than six workers are employed need not be reported separately but may be classified under the employer's primary county.

(c.) Data on workers who in the normal course of their duties perform services in more than one country within the State should be allocated to the county of the head office from which they operate.

SSA: MULTIUNIT EI FILE SYSTEM (EXCERPT)

Units: The Establishment Reporting Plan (ERP) is a voluntary program conducted by the Social Security Administration, under which multiunit employers are requested to group their employees by establishment within their wage reports in order to allow the Administration to collect accurate geographic and industrial data for statistical purposes. The principal source forms used for identifying potential reporters are: (1) the Form SS-4, Application for Employer Identification Number (with 50 or more employees and more than one place of business indicated on the form) and (2) Annual Wage Reports of employers with 50 or more employees. Once the potential employer has been identified he is contacted to see if he is eligible to use the plan.

Under ERP, eligible employers are those who have 50 or more employees and more than one place of business under one EIN and who also report one of the following: (1) six or more employees in a secondary State, (2) 10 percent of employment or at least 50 employees in a secondary county or industry or (3) 2 or more manufacturing establishments in the same geographic area. If the

employer is eligible, he is asked to participate in the plan, by (1) providing SSA with a Form SSA-5019 which lists his establishments with a unit number assigned to each one, (2) grouping his employees under these same unit numbers on the W-3 and or W-2 of his annual wage report, and (3) submitting additional SSA-5019's whenever he needs to correct, supplement or revise his original list of establishments. In listing his reporting units, if the employer is unable to group his employees by establishment, he may use a payroll grouping instead. For example, when an employer operates two or more non-manufacturing establishments in the same county, the employer may combine the employment of such establishments and treat them as one reporting unit for purposes of the plan. Also when employment is reported with no fixed county location, the unit may be identified as nationwide or statewide. Each of the establishments or payroll groupings assigned a reporting unit number under an EIN represents one basic coding unit.

APPENDIX B

SELECTED SOURCE DOCUMENTS AND INSTRUCTIONS

The source documents and instructions in this appendix are included to give an idea of the wide variety in the amount and kinds of information obtained by different agencies, and for different data systems within agencies, to classify units and assign codes. They do not cover all of the systems reviewed by the Industry Coding Working Group; they were selected to illustrate different levels of detail, as well as the difference between a document designed for self-coding by the respondent (IRS Form 1065) and a document designed for coding by an agency. For the longer forms, only those parts directly relevant to industry coding are shown. Similarly, only those parts of respondent instructions relevant to industry coding are included.

A comparative analysis of these forms appears in Chapter 3, "Examples of Source Documents;" the Census Bureau and BEA forms are also discussed in Chapter 4 in connection with coding procedures. Each form and the corresponding instructions, if any, appear as separate exhibits.

EXHIBIT 1: IRS FORM 1065, U. S. PARTNERSHIP RETURN OF INCOME, TAX YEAR 1981

This form is used in two industry coding systems: revenue processing of partnership returns; and statistics of income (SOI) for partnerships (for a sample of returns).

Shown are page 1 of the form and page 12 of the taxpayer instructions. The latter provides the codes to be used by the taxpayers in item C on the form.

For the revenue processing industry coding system, the code entered by the taxpayer in item C is normally accepted. For the statistics of income industry coding system, past practice has been for coders to use items A, B, and C, name of taxpayer, and other relevant items to assign a code which is entered in the margin of the form. A partially automated system, making use of prior year revenue-processing and SOI codes, when available, is now used (see Chapter 4, "Automated Procedures, Coding").

Form **1065**
Department of the Treasury
Internal Revenue Service

U.S. Partnership Return of Income

For calendar year 1981, or fiscal year beginning ..... , 1981, and ending ..... , 19.....

OMB No. 1545-0099

1981

A Principal business activity (see page 12 of Instructions)

B Principal product or service (see page 12 of Instructions)

C Business code number (see page 12 of Instructions)

Use IRS label. Otherwise, please print or type.

Name

Number and street

City or town, State, and ZIP code

D Employer identification no.

E Date business started

F Enter total assets from Schedule L, line 13, column (D). $

G Check method of accounting: (1) ☐ Cash (2) ☐ Accrual (3) ☐ Other (attach explanation)

H Check applicable boxes: (1) ☐ Final return (2) ☐ Change in address.

IMPORTANT—Fill in all applicable lines and schedules. If you need more space, see page 2 of the Instructions. Enter any items specially allocated to the partners on Schedule K, line 17, and not on the numbered lines on this page or in Schedules A through I.

Income	1a	Gross receipts or sales $	1b Minus returns and allowances $ ..... Balance ▶	1c	
	2	Cost of goods sold and/or operations (Schedule A, line 34)		2	
	3	Gross profit (subtract line 2 from line 1c)		3	
	4	Ordinary income (loss) from other partnerships and fiduciaries (attach statement)		4	
	5	Nonqualifying dividends		5	
	6	Nonqualifying interest		6	
	7	Net income (loss) from rents (Schedule H, line 2)		7	
	8	Net income (loss) from royalties (attach schedule)		8	
	9	Net farm profit (loss) (attach Schedule F (Form 1040))		9	
	10	Net gain (loss) (Form 4797, line 11)		10	
	11	Other income (attach schedule)		11	
	12	**TOTAL** income (loss) (combine lines 3 through 11)		12	
Deductions	13a	Salaries and wages (other than to partners) $	13b Minus jobs credit $ ..... Balance ▶	13c	
	14	Guaranteed payments to partners (see page 4 of Instructions)		14	
	15	Rent		15	
	16	Interest (Caution—see page 4 of Instructions)		16	
	17	Taxes		17	
	18	Bad debts (see page 5 of Instructions)		18	
	19	Repairs		19	
	20	Depreciation from Form 4562 (attach Form 4562) $ ....., less depreciation claimed in Schedules A and H and elsewhere on return $ ....., Balance ▶		20	
	21	Amortization (attach schedule)		21	
	22	Depletion (other than oil and gas, attach schedule—see page 5 of Instructions)		22	
	23a	Retirement plans, etc. (see page 5 of Instructions)		23a	
	23b	Employee benefit programs (see page 5 of Instructions)		23b	
	24	Other deductions (attach schedule)		24	
	25	**TOTAL** deductions (add lines 13c through 24)		25	
	26	Ordinary income (loss) (subtract line 25 from line 12)		26	

Schedule A—COST OF GOODS SOLD AND/OR OPERATIONS (See Page 6 of Instructions)

27	Inventory at beginning of year (if different from last year's closing inventory, attach explanation)			27	
28a	Purchases $	28b Minus cost of items withdrawn for personal use $	Balance ▶	28c	
29	Cost of labor			29	
30	Materials and supplies			30	
31	Other costs (attach schedule)			31	
32	Total (add lines 27 through 31)			32	
33	Inventory at end of year			33	
34	Cost of goods sold (subtract line 33 from line 32). Enter here and on line 2, above			34	

Please Sign Here

Under penalties of perjury, I declare that I have examined this return, including accompanying schedules and statements, and to the best of my knowledge and belief it is true, correct, and complete. Declaration of preparer (other than taxpayer) is based on all information of which preparer has any knowledge.

▶ Signature of general partner ▶ Date

Paid Preparer's Use Only

Preparer's signature ▶ | Date | Check if self-employed ▶ ☐ | Preparer's social security no.

Firm's name (or yours, if self-employed) and address ▶ | E.I. No. ▶ | ZIP code ▶

For Paperwork Reduction Act Notice, see page 1 of Form 1065 Instructions

Codes for Principal Business Activity and Principal Product or Service

These industry titles and definitions, for use on Form 1065 partnership returns, are based on the Standard Industrial Classification System authorized by the Statistical Policy Division, Office of Information and Regulatory Affairs, in the Office of Management and Budget, to classify enterprises and establishments by type of activity in which engaged.

Using the list below, enter on page 1, under C, the code for the specific industry group for which the largest percentage of "total assets" is used. "Total assets" means the amount entered on Schedule L, line 13, column (D). On page 1, under A, state the principal business activity and under B, state the principal product or service which accounts for the largest percentage of total assets. For example, if the principal business activity is "Retail food store," the principal product or service may be "dairy products."

Agriculture, Forestry, and Fishing

Code

Farms:

0120 Field crop.
0160 Vegetable and melon farms.
0170 Fruit and nut tree farms.
0180 Horticultural specialty.
0211 Beef cattle feedlots.
0212 Beef cattle, except feedlots.
0213 Hogs, sheep, and goats.
0240 Dairy farms.
0250 Poultry and eggs.
0260 General livestock (except animal specialty).
0270 Animal specialty.

Agricultural services and forestry:

0740 Veterinary services.
0753 Livestock breeding.
0754 Animal services, except livestock breeding and veterinary.
0780 Landscape and horticultural services.
0790 Other agricultural services.
0800 Forestry.

Fishing, hunting, and trapping:

0930 Commercial fishing, hatcheries and preserves.
0970 Hunting, trapping, and game propagation.

Mining

1000 Metal mining.
1150 Coal mining.
1300 Oil and gas extraction.
1400 Nonmetallic minerals except fuel.

Construction

General building contractors and operative builders:

1510 General building contractors.
1531 Operative builders.

Heavy construction contractors:

1611 Highway and street construction.
1620 Heavy construction, except highway.

Special trade contractors:

1711 Plumbing, heating, and air conditioning.
1721 Painting, paperhanging, and decorating.
1731 Electrical work.
1740 Masonry, stonework, and plastering.
1750 Carpentering and flooring.
1761 Roofing and sheet metal work.
1771 Concrete work.
1781 Water well drilling.
1790 Miscellaneous special trade contractors.

Manufacturing

2000 Food and kindred products.
2200 Textile mill products.
2300 Apparel and other textile products.
2400 Lumber and wood products except furniture.
2500 Furniture and fixtures.
2700 Printing, publishing, and allied industries.
2800 Chemicals and allied products.
3100 Leather and leather products.
3200 Stone, clay, and glass products.
3300 Primary metal industries.
3400 Fabricated metal products.
3500 Machinery, except electrical.
3600 Electrical and electronic equipment.
3700 Transportation equipment.
3970 Other manufacturing industries.

Transportation, Communication, Electric, Gas, and Sanitary Services

Local and interurban passenger transit:

4121 Taxicabs.
4189 Other passenger transportation.

Trucking and warehousing:

4210 Trucking, local and long distance.
4289 Public warehousing and trucking terminals.

Other transportation including transportation services:

4400 Water transportation.
4540 Transportation by air.
4722 Passenger transportation arrangement.
4723 Freight transportation arrangement.
4799 Other transportation services.
4800 Communication.
4900 Electric and gas services.
4960 Sanitary services.

Wholesale Trade

Durable:

5010 Motor vehicles and automotive equipment.
5030 Lumber and construction materials.
5060 Electrical goods.
5070 Hardware, plumbing, and heating equipment.
5083 Farm machinery, and equipment.
5089 Other machinery, equipment, and supplies.
5095 Other durable goods.

Nondurable:

5129 Drugs, chemicals, and allied products.
5130 Apparel, piece goods, and notions.
5140 Groceries and related products.
5150 Farm-product raw materials.
5180 Alcoholic beverages.
5195 Other nondurable goods.

Retail Trade

Building materials, hardware, garden supply, and mobile home dealers:

5211 Lumber and other building materials dealers.
5231 Paint, glass, and wallpaper stores.
5251 Hardware stores.
5261 Retail nurseries and garden stores.
5271 Mobile home dealers.

General merchandise:

5331 Variety stores.
5398 Other general merchandise stores.

Food stores:

5411 Grocery stores.
5420 Meat and fish markets freezer provisioners.
5431 Fruit stores and vegetable markets.
5441 Candy, nut, and confectionery stores.
5451 Dairy products stores.
5460 Retail bakeries.
5490 Other food stores.

Automotive dealers and service stations:

5511 New car dealers (franchised).
5521 Used car dealers.
5531 Auto and home supply stores.
5541 Gasoline service stations.
5551 Boat dealers.
5561 Recreational vehicle dealers.
5571 Motorcycle dealers.
5599 Aircraft, and other automotive dealers.

Apparel and accessory stores:

5611 Men's and boys' clothing and furnishings.
5621 Women's ready-to-wear stores.
5631 Women's accessory and specialty stores.
5641 Children's and infants' wear stores.
5651 Family clothing stores.
5661 Shoe stores.
5681 Furriers and fur shops.
5699 Other apparel and accessory stores.

Furniture, home furnishings, and equipment stores:

5712 Furniture stores.
5713 Floor covering stores.
5714 Drapery, curtain, and upholstery stores.
5719 Home furnishings, except appliances.
5722 Household appliance stores.
5732 Radio and television stores.
5733 Music stores.

Eating and drinking places:

5812 Eating places.
5813 Drinking places.

Miscellaneous retail stores:

5912 Drug stores and proprietary stores.
5921 Liquor stores.
5931 Used merchandise stores.
5941 Sporting goods stores and bicycle shops.
5942 Book stores.
5943 Stationery stores.
5944 Jewelry stores.
5945 Hobby, toy, and game shops.
5946 Camera and photographic supply stores.
5947 Gift, novelty, and souvenir shops.
5948 Luggage and leather goods stores.
5949 Sewing, needlework, and piece goods stores.
5961 Mail order houses.
5962 Merchandising machine operators.
5963 Direct selling organizations.
5982 Fuel and ice dealers (except fuel oil and bottled gas dealers).
5983 Fuel oil dealers.
5984 Liquefied petroleum gas (bottled gas) dealers.
5992 Florists.
5993 Cigar stores and stands.
5994 News dealers and newsstands.
5996 Other miscellaneous retail stores.

Finance, Insurance, and Real Estate

6000 Banking.
6100 Credit agencies other than banks.

Security and commodity brokers, dealers, exchanges and services:

6212 Security underwriting syndicates.
6218 Security brokers and dealers, except underwriting syndicates.
6299 Commodity contracts brokers and dealers; security and commodity exchanges; and allied services.

Real estate:

6411 Insurance agents, brokers, and services.
6511 Real estate operators (except developers) and lessors of buildings.
6520 Lessors of real property other than buildings.
6531 Real estate agents, brokers and managers.
6541 Title abstract offices.
6552 Subdividers and developers except cemeteries.
6553 Cemetery subdividers and developers.
6611 Combined real estate, insurance, loans, law offices.

Holding and other investment companies:

6746 Investment clubs.
6747 Common trust funds.
6748 Other holding and investment companies.

Services

Hotels and other lodging places

7012 Hotels.
7013 Motels, motor hotels, and tourist courts.
7021 Rooming and boarding houses.
7032 Sporting and recreational camps.
7033 Trailer parks and camp sites.
7041 Organization hotels and lodging houses on a membership basis.

Personal services:

7215 Coin-operated laundries and dry cleaning.
7219 Other laundry, cleaning, and garment services.
7221 Photographic studios, portrait studios.
7231 Beauty shops.
7241 Barber shops.
7251 Shoe repair and hat cleaning shops.
7261 Funeral services and crematories.
7299 Miscellaneous personal services.

Business services:

7310 Advertising.
7340 Services to buildings.
7370 Computer and data processing services.
7392 Management, consulting, and public relations services.
7394 Equipment rental and leasing.
7398 Other business services.

Automotive repair and services:

7510 Automotive rentals and leasing, without drivers.
7520 Automobile parking.
7531 Automobile top and body repair shops.
7538 General automobile repair shops.
7539 Other automotive repair shops.
7540 Automotive services, except repair.

Miscellaneous repair services:

7622 Radio and TV repair shops.
7628 Electrical repair shops, except radio and TV.
7641 Reupholstery and furniture repair.
7680 Other miscellaneous repair shops.

Motion picture:

7812 Motion picture and video tape production, distribution, and services.
7830 Motion picture theaters.

Amusement and recreation services:

7920 Producers, orchestras, and entertainers.
7932 Billiard and pool establishments.
7933 Bowling alleys.
7941 Professional sports clubs and promoters.
7948 Racing, including track operation.
7980 Other amusement and recreation services.

Medical and health services:

8011 Offices of physicians.
8021 Offices of dentists.
8031 Offices of osteopathic physicians.
8041 Offices of chiropractors.
8042 Offices of optometrists.
8048 Registered and practical nurses.
8050 Nursing and personal care facilities.
8060 Hospitals.
8071 Medical laboratories.
8072 Dental laboratories.
8098 Other medical and health services.

Other services:

8111 Legal services.
8200 Educational services.
8911 Engineering and architectural services.
8931 Certified public accountants.
8933 Other accounting, auditing, and bookkeeping services.
8999 Other services, not elsewhere classified.

U.S. GOVERNMENT PRINTING OFFICE 1993-0-301

EXHIBIT 2: FORM 1040, IRS SCHEDULE C, PROFIT (OR LOSS) FROM BUSINESS OR PROFESSION (SOLE PROPRIETORSHIP), TAX YEAR 1981

This form is used in two industry coding systems: revenue processing of sole proprietorship returns; and statistics of income for sole proprietorships (for a sample of returns).

Shown are page 1 of the form and the paragraph covering item A, main business activity and product, from page 27 of the taxpayer instructions.

For the revenue processing industry coding system, a code based primarily on item A is entered on the return by a coder. For returns in the statistics of income sample, past practice has been to enter a separate code on the return, making full use of all relevant information available. As in the case of the SOI partnership system, the present coding system for SOI sole proprietorships is partially automated (see Chapter 4, "Automated Procedures, Coding").

SCHEDULE C (Form 1040)
Department of the Treasury
Internal Revenue Service (0)

Profit or (Loss) From Business or Profession

(Sole Proprietorship)

Partnerships, Joint Ventures, etc., Must File Form 1065.

▶ Attach to Form 1040 or Form 1041. ▶ See Instructions for Schedule C (Form 1040).

OMB No. 1545-0074

1981

08

Name of proprietor | Social security number of proprietor

A Main business activity (see Instructions) ▶ ; product ▶

B Business name ▶ | C Employer identification number

D Business address (number and street) ▶
City, State and ZIP Code ▶

E Accounting method: (1) ☐ Cash (2) ☐ Accrual (3) ☐ Other (specify) ▶

C

F Method(s) used to value closing inventory:
(1) ☐ Cost (2) ☐ Lower of cost or market (3) ☐ Other (if other, attach explanation)

	Yes	No
G Was there any major change in determining quantities, costs, or valuations between opening and closing inventory? If "Yes," attach explanation.		
H Did you deduct expenses for an office in your home?		

Part I Income

1 a Gross receipts or sales	1a		
b Returns and allowances	1b		
c Balance (subtract line 1b from line 1a)		1c	
2 Cost of goods sold and/or operations (Schedule C-1, line 8)		2	
3 Gross profit (subtract line 2 from line 1c)		3	
4 a Windfall Profit Tax Credit or Refund received in 1981 (see Instructions)		4a	
b Other income (attach schedule)		4b	
5 Total income (add lines 3, 4a, and 4b) ▶		5	

Part II Deductions

6 Advertising		29 a Wages	
7 Amortization		b Jobs credit	
8 Bad debts from sales or services		c WIN credit	
9 Bank service charges		d Total credits	
10 Car and truck expenses		e Subtract line 29d from 29a	
11 Commissions		30 Windfall Profit Tax withheld in 1981	
12 Depletion		31 Other expenses (specify):	
13 Depreciation (see Instructions)		a	
14 Dues and publications		b	
15 Employee benefit programs		c	
16 Freight (not included on Schedule C-1)		d	
17 Insurance		e	
18 Interest on business indebtedness		f	
19 Laundry and cleaning		g	
20 Legal and professional services		h	
21 Office supplies and postage		i	
22 Pension and profit-sharing plans		j	
23 Rent on business property		k	
24 Repairs		l	
25 Supplies (not included on Schedule C-1)		m	
26 Taxes (do not include Windfall Profit Tax, see line 30)		n	
27 Travel and entertainment		o	
28 Utilities and telephone		p	

32 Total deductions (add amounts in columns for lines 6 through 31p) ▶	32	
33 Net profit or (loss) (subtract line 32 from line 5). If a profit, enter on Form 1040, line 11, and on Schedule SE, Part II, line 5a (or Form 1041, line 6). If a loss, go on to line 34	33	

34 If you have a loss, do you have amounts for which you are not "at risk" in this business (see Instructions)? ☐ Yes ☐ No
If you checked "No," enter the loss on Form 1040, line 11, and on Schedule SE, Part II, line 5a (or Form 1041, line 6).

For Paperwork Reduction Act Notice, see Form 1040 Instructions.

Instructions for Schedule C

Profit or (Loss) From Business or Profession

Purpose of Form

If you operated a business or practiced a profession as a sole proprietorship, complete Schedule C. If you had more than one business, or if you and your spouse had separate businesses, you must complete a **Schedule C** for each business. Farmers should use **Schedule F.**

For expenses that are part business and part personal, deduct only the business part. For example, if only half of your car usage was for business, deduct only half of the cost of operating the car.

Deduct interest, taxes, and casualty losses not related to your business as itemized deductions on **Schedule A.**

Report sales, exchanges, and involuntary conversions (other then casualty or theft) of trade or business property on **Form 4797,** Supplemental Schedule of Gains and Losses.

Use **Form 4684,** Casualties and Thefts, to report a casualty or theft involving trade or business, or income-producing property.

You must pay social security self-employment tax on income from any trade or business unless you are specifically excepted. **Please see Schedule SE.**

For more details about business income and expenses, get **Publication 334,** Tax Guide for Small Business.

Information Returns

You may have to file information returns for wages paid to employees, certain payments of fees and other non-employee compensation, interest, rents, royalties, annuities, and pensions, or for sales by you of $5,000 or more of consumer products to a parson on a buy-sell, deposit-commission, or other basis for resale. For more information, see instructions for **Form W-3,** Transmittal of Income and Tax Statements, **Form W-3G,** Transmittal of Certain Information Returns, and **Form 1096,** Annual Summary and Transmittal of U.S. Information Returns.

Item A

Main Business Activity and Product

Report the business activity that accounted for the most income included on Schedule C, Part I, line 1a. Give the general field as well as the product or service. For example, "wholesale—groceries" or "retail—hardware."

Item B

Business Name and Address

Use your home address only if you actually conducted the business from your home. You should show a street address instead of a box number.

Item C

Employer Identification Number

You don't need an employer identification number unless you had a Keogh (H.R. 10) plan or were required to file an employment, excise, or alcohol, tobacco, and firearms tax return.

Item D

Valuation Methods

Your inventories can be valued at:

- cost,
- cost or market value, whichever is lower, or
- any other method approved by the Commissioner of Internal Revenue.

Item E

Accounting Methods

You must use the cash method on your return unless you kept account books. If you kept such books, you can use the cash method, accrual method, or in some cases, the completed contract or percentage-of-completion method. The method used must clearly reflect your income.

To change your accounting method (including the treatment of any item such as inventories or bad debts), you must usually first get the permission of the Commissioner of Internal Revenue. File **Form 3115** within the first 180 days of the tax year in which you want to make the change.

If you use the **cash method,** show all items of taxable income actually or constructively received during the year (in cash, property, or services). Also show amounts actually paid during the year for deductible expenses. Income is constructively received when it is credited to your account or set aside for you to use.

If you use the **accrual method,** report income when you earn it and deduct expenses when you incur them, even if you do not pay them during the tax year.

Item G

Business Use of Your Home

Within certain limits, you may deduct business expenses that apply to a part of your home **only** if that part is exclusively used on a regular basis:

a. as your principal place of business for any of your trades or businesses, or

b. as a place of business used by your patients, clients, or customers to meet or deal with you in the normal course of your trade or business, or

c. in connection with your trade or business if it is a separate structure that is not attached to your home.

You may also deduct expenses that apply to space within your home if it is the **only** fixed location of your trade or business. The space must be used on a regular basis to store inventory held for use in your trade or business of selling products at retail or wholesale.

If you use space in your home on a regular basis in your trade or business of providing day care service, you may be able to deduct the business expenses even though you also use the same space for nonbusiness expenses.

Please get **Publication 587,** Business Use of Your Home, for more information.

Part I

Income (Lines 1 through 5)

Line 1a

Gross Receipts or Sales

Enter gross receipts or sales from your business.

Installment Sales. If you use the installment method of reporting sales income, please attach a schedule showing separately for 1983 and the three preceding years: gross sales; cost of goods sold; gross profit; percentage of gross profits to gross sales; amounts collected; and gross profits on amounts collected.

Line 1b

Returns and Allowances

You should enter on line 1b such items as returned sales, rebates, and allowances from the sales price.

Line 4a

Income from Overpaid Windfall Profit Tax

In certain situations, you must report as income on line 4a the amount of any credit or refund of overpaid windfall profit tax you received in 1983 for tax year 1982, based on overwithholding or the net income limitation.

In general, the amount of credit or refund you received is income to the extent you deducted windfall profit tax withheld in 1982 on Schedule C, and received a tax benefit for the deduction on your 1982 tax return.

Line 4b

Other Income

Include finance reserve income, scrap sales, amounts recovered from bad debts, interest, and other kinds of miscellaneous income from the business or profession.

Part II

Deductions (Lines 6 through 30)

Line 7

Bad Debts from Sales or Services

Caution: *Cash method taxpayers are not entitled to a bad debt deduction unless the amount was previously included in income.*

Include debts and partial debts arising from sales or services that were included in income and are definitely known to be worthless. Instead of this, you may deduct a reasonable amount that was added during the tax year to a bad debt reserve.

If you later collect a debt that you deducted as a bad debt, include it as income in the year you collect it unless you use the bad debt reserve method. For more information, please get **Publication 548,** Deduction for Bad Debts.

Line 9

Car and Truck Expenses

You can deduct the actual cost of running your car or truck, or take the fixed mileage rate. You must use actual costs if you use more than one vehicle in your business. If you deduct actual cost, show depreciation on line 12.

The fixed rate is figured at 20.5 cents a mile up to 15,000 miles, and 11 cents for each mile after that. Add to this amount your parking fees and tolls.

For cars and trucks that have been fully depreciated, the rate is 11 cents a mile.

If you use the fixed rate, the vehicle is considered to have a useful life of 60,000 miles of business use at the maximum fixed mileage rate.

EXHIBIT 3: IRS FORM SS-4, APPLICATION FOR EMPLOYER IDENTIFICATION NUMBER

This form is used for industry coding in the single-unit employer identification number (EIN) system.

Shown are the full form and the instructions to applicants. Several items are used for industry coding (see text). Although this is an Internal Revenue Service form, the industry coding is done by the Social Security Administration.

Form **SS–4**
(Rev. 9–82)
Department of the Treasury
Internal Revenue Service

Application for Employer Identification Number

(For use by employers and others as explained in the instructions.
Please read the instructions before completing this form.)
For Paperwork Reduction Act Notice, see page 2.

OMB No. 1545–0003 Expires 9–30–85

1 Name (True name and not trade name. If partnership, see page 4.)

2 Social security no., if sole proprietor

3 Ending month of accounting year

4 Trade name, if any, of business (if different from item 1)

5 General partner's name, if partnership; principal officer's name, if corporation; or grantor's name, if trust

6 Address of principal place of business (Number and street)

7 Mailing address, if different

8 City, State, and ZIP code

9 County of principal business location

10 Type of organization
☐ Individual ☐ Trust ☐ Partnership ☐ Other (specify)
☐ Governmental ☐ Nonprofit organization ☐ Corporation

11 Date you acquired or started this business (Mo., day, year)

12 Reason for applying
☐ Started new business ☐ Purchased going business ☐ Other (specify)

13 First date you paid or will pay wages for this business (Mo., day, year)

14 Nature of principal business activity (See instructions on page 4.)

15 Do you operate more than one place of business? ☐ Yes ☐ No

16 Peak number of employees expected in next 12 months (If none, enter "0") ▶	Nonagricultural	Agricultural	Household

17 If nature of business is manufacturing, state principal product and raw material used.

18 To whom do you sell most of your products or services?
☐ Business establishments (wholesale) ☐ General public (retail) ☐ Other (specify)

19 Have you ever applied for an identification number for this or any other business? ☐ Yes ☐ No
If "Yes," enter name and trade name. Also enter approx. date, city, and State where you applied and previous number if known. ▶

Under penalties of perjury, I declare that I have examined this application, and to the best of my knowledge and belief it is true, correct, and complete.

Signature and Title ▶ Date ▶

Telephone number (include area code)

Please leave blank ▶	Geo.	Ind.	Class	Size	Reas. for appl.	Part I

General Instructions

Paperwork Reduction Act Notice.—We ask for this information to carry out the Internal Revenue laws of the United States. We need it to ensure that you are complying with these laws. You are required to give us this information.

Purpose.—Use this form to apply for an employer identification number (EIN). Return both parts of this form to the Internal Revenue Service. You will receive your EIN in the mail.

Who must file.—You must file this form if you have not obtained an EIN before and:

(a) You pay wages to one or more employees; or

(b) You are required to have an EIN to use on any return, statement, or other document, even if you are not an employer.

Trusts, estates, corporations, partnerships, or nonprofit organizations (churches, clubs, etc.) must use EINs even if they have no employees.

Individuals who file Schedules C or F (Form 1040) must use EINs if they are required to file excise, employment, or alcohol, tobacco, or firearms returns.

File only one Form SS–4, regardless of the number of businesses operated or the number of trade names a business operates under. However, each corporation of an affiliated group must file a separate application.

If you have become the new owner of an existing business, you cannot use the EIN of the old owner. If you already have an EIN, use that number. If you do not have an EIN, apply for one on this form.

If you have incorporated a sole proprietorship or formed a partnership, you must get a new EIN for the corporation or partnership.

If you do not have a number by the time a return is due, write "Applied for" and the date you applied in the space shown for the number. If you do not have a number by the time a tax deposit is due, send your payment to the Internal Revenue Service Center where you file your returns. Make it payable to IRS and show on it your name (as shown on Form SS–4), address, kind of tax, period covered, and date you applied for an EIN.

For more information about EINs, see Publication 583, Information for Business Taxpayers.

When to file.—File early enough to allow time for us to process Form SS–4 and send you an EIN before you need the number for a return or deposit. (If possible, file 4 weeks before you will need the number.) See "Where to file" on page 4.

Specific Instructions

Most lines on this form are self-explanatory. The instructions that follow are for those lines that may not be.

Lines 1, 2, 4, and 5.

Sole proprietors.—On line 1, enter your first name, middle initial, and last name. On line 2, enter your social security number and, if you have a trade name for business purposes, enter it on line 4.

Partnerships.—On line 1, enter the legal name of the partnership as it appears in the partnership agreement. On line 4, enter the trade name, if any, and on line 5, enter the first name, middle initial, and last name of a general partner. A general partner should sign this form.

Corporations.—On line 1, enter the corporate name as set forth in the corporation's charter or other legal document creating it. On line 4, enter the trade name, if any, and on line 5, enter the first name, middle initial, and last name of a principal officer. A principal officer should sign this form.

Trusts.—On line 1, enter the name of the trust. On line 4, enter the name of the trustee and on line 5, enter the first name, middle initial, and last name of the grantor. The trustee should sign this form. (See the instruction for line 11.)

Estates of a decedent, insolvent, etc.—On line 1, enter the name of the estate. On line 4, enter the first name, middle initial, and last name of the administrator or other fiduciary. The administrator or other fiduciary should sign this form. (See the instruction for line 11.)

Line 3.—If you have not yet established an accounting year, write "not established" on line 3 and notify your IRS Service Center when you establish an accounting year. (Be sure to include your employer identification number when you write.)

(Continued on page 4)

Line 10.—Note the following before you check:

Governmental.—This box is for an organization that is a State, county, school district, municipality, etc., or one that is related to such entities, such as a county hospital or city library.

Nonprofit organization (other than governmental).—This box is for religious, charitable, scientific, literary, educational, humane, or fraternal, etc., organizations. Generally, a nonprofit organization must apply to IRS for an exemption from Federal income tax. Details on how to apply are in IRS Publication 557, Tax-Exempt Status for Your Organization.

Line 11.—For trusts, enter the date the trust was legally created.

For estates, enter the date of death of the decedent whose name appears on line 1.

Line 14.—Describe the principal business engaged in. See the examples that follow.

(a) Governmental.—State the type of governmental organization (whether it is a State, county, school district, municipality, etc.) or its relationship to such entities (for example, a county hospital, city library, etc.).

(b) Nonprofit (other than governmental).—State whether it is organized for religious, charitable, scientific, literary, educational, or humane purposes, and state the principal activity (for example, religious organization—hospital; charitable organization—home for the aged; etc.).

(c) Mining and quarrying.—State the process and the principal product (for example, mining bituminous coal, contract drilling for oil, quarrying dimension stone, etc.).

(d) Contract construction.—State whether it is general contracting or special trade contracting, and show the type of work normally performed (for example, general contractor for residential buildings, electrical subcontractor, etc.).

(e) Trade.—State the type of sale and the principal line of goods sold (for example, wholesale dairy products, manufacturer's representative for mining machinery, retail hardware, etc.).

(f) Manufacturing.—State the type of establishment operated (for example, sawmill, vegetable cannery, etc.). On line 17 state the principal product manufactured and the raw material used.

(g) Other activities.—State the exact type of business operated (for example, advertising agency, farm, labor union, real estate agency, steam laundry, rental of coin-operated vending machines, investment club, etc.).

Where to file.—

If your principal business, office or agency, or legal residence in the case of an individual, is located in:	File with the Internal Revenue Service Center at:	
New Jersey, New York City and counties of Nassau, Rockland, Suffolk, and Westchester	Holtsville, NY	00501
New York (all other counties), Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont	Andover, MA	05501
District of Columbia, Delaware, Maryland, Pennsylvania	Philadelphia, PA	19255
Alabama, Florida, Georgia, Mississippi, South Carolina	Atlanta, GA	31101
Michigan, Ohio	Cincinnati, OH	45999
Arkansas, Kansas, Louisiana, New Mexico, Oklahoma, Texas	Austin, TX	73301
Alaska, Arizona, Colorado, Idaho, Minnesota, Montana, Nebraska, Nevada, North Dakota, Oregon, South Dakota, Utah, Washington, Wyoming	Ogden, UT	84201
Illinois, Iowa, Missouri, Wisconsin	Kansas City, MO	64999
California, Hawaii	Fresno, CA	93888
Indiana, Kentucky, North Carolina, Tennessee, Virginia, West Virginia	Memphis, TN	37501

If you have no legal residence, principal place of business, or principal office or agency in any Internal Revenue district, file your return with the Internal Revenue Service Center, Philadelphia, PA 19255.

Please sign and date this application.

☆ U.S. GOVERNMENT PRINTING OFFICE : 1982—O-363-459

EXHIBIT 4: CENSUS BUREAU FORM CB-5502, 1982 CENSUS OF RETAIL TRADE: TIRES, BATTERIES, PARTS, ACCESSORIES

This form is used for industry coding for economic censuses. This is one of a large number of specialized forms used in the mail portion of the 1982 economic censuses. As explained in Chapter 3, many of the items in the questionnaire are used in the largely automated industry coding process. The key item is item 11--Merchandise Lines.

BUREAU OF THE CENSUS

FORM

CB-5502

1982 CENSUS OF RETAIL TRADE

TIRES, BATTERIES, PARTS, ACCESSORIES

O.M.B. APPROVAL NO. 0607-0371 EXPIRES 12/84

NOTICE – Response to this inquiry is required by law (Title 13, U.S. Code). By the same law, your report to the Census Bureau is confidential. It may be seen only by sworn Census employees and may be used only for statistical purposes. The law also provides that copies retained in your files are immune from legal process.

BUREAU OF THE CENSUS
1201 East Tenth Street
Jeffersonville, Indiana 47134

DUE DATE: FEBRUARY 15, 1983

If filing by the due date causes an undue burden, a time extension request should be sent to the above address, please include your 11-digit Census File Number(CFN).

Note — *Please read the accompanying instructions before answering the questions.*

In correspondence pertaining to this report, please refer to this Census File Number (CFN)

Employer Identification (EI) Number

CB-5502

Please correct errors in name, address, and ZIP code. ENTER street and number if not shown.

Item 1 – EMPLOYER IDENTIFICATION NUMBER

Is the Employer Identification (EI) Number shown on the label the SAME as that used for this establishment on its latest 1982 Employer's Quarterly Federal Tax Return, Treasury Form 941?

094 1 YES

2 NO Enter current EI Number → (9 digits) – | | | | | | | | | |

Item 2 – PHYSICAL LOCATION OF ESTABLISHMENT

Answer items a, b, c, and d

NOTE: P.O. boxes or rural routes are not physical locations.

a. Same as shown in mailing label. If different, indicate change.

NUMBER AND STREET

CITY, TOWN, VILLAGE, ETC. | STATE | ZIP CODE

b. Is this establishment physically located inside the legal boundaries of the city, town, village, etc.?

095 1 YES 3 No legal boundaries

2 NO 4 Don't know

c. Type of municipality where physically located

096 1 City, village, or borough 3 Other or don't know

2 Town or township

d. Name of county where physically located

Item 3 – OPERATIONAL STATUS

a. How many months during 1982 did this firm or organization actively operate this establishment? — Number of months — 002

b. Mark (X) the ONE box which best describes this establishment at the end of 1982.

001 1 In operation

2 Temporarily or seasonally inactive

3 Ceased operation – Give date → (Figures only: Month | Day | Year)

4 Sold or leased to another operator – Give date at right → AND enter name, etc., below

NAME OF NEW OWNER OR OPERATOR

NUMBER AND STREET

CITY | STATE | ZIP CODE

Item 4 – ORGANIZATIONAL STATUS

Mark (X) the ONE box which best describes this establishment during 1982.

003 1 Individual proprietorship

2 Partnership

3 Cooperative association (taxable)

4 Cooperative association (tax-exempt)

5 Government – Specify _____

6 Corporation (Do not mark if any form of cooperative association.)

9 Other – Specify _____

Value figures may be reported in dollars or rounded to thousands.

Example: If a figure is $1,125,628, report either → *PREFERRED* / *Acceptable*

	Millions (000)	Thousands (000)	Dollars (000)
PREFERRED	1	126	
Acceptable	1	125	628

Item 5 – DOLLAR VOLUME OF BUSINESS IN 1982

Sales of merchandise and other operating receipts EXCLUDING sales (or other) taxes collected — 010 — Mil | Thou | Dol

Item 6 – PAYROLL AND EMPLOYMENT

a. Payroll in 1982, before deductions

(1) Total ANNUAL payroll — 030 — Mil | Thou | Dol

(2) FIRST QUARTER payroll — 031 — Mil | Thou | Dol

b. Employment in 1982

Number of paid employees for the pay period including March 12, 1982. (Include both full- and part-time employees.) — 032 — Number

Item 7 – CLASS OF CUSTOMER

Report the approximate percentage of this establishment's total 1982 sales to each class of customer.

		Report in whole percents
a. General public (household consumers and individuals)	310	
b. Other, including retailers, wholesalers, institutional, industrial, commercial, professional, and farm users (for use in farm production), and government	312	

Item 8 – Not applicable to this report.

Item 9 – KIND OF BUSINESS

Mark (X) the ONE box which best describes the PRINCIPAL kind of business of this establishment in 1982.

070

Dealer in new tires, batteries, and accessories		553115
Auto supply store		553115
Home and auto supply store		553125
Dealer in used tires, batteries, and accessories		593125
Auto wrecker		593125
Gasoline service station		554115
Truck stop		554125
Tire recapping and retreading shop		7534OO
Franchised passenger car dealer (domestic and/or imported cars)		561105
Nonfranchised passenger car dealer (used cars)		562105
Boat dealer		555105
Recreational vehicle dealer		556105
Motorcycle, motor scooter dealer		557105
Other kind of business – *Describe kind*		777777

Item 10 – METHOD OF SELLING

Mark (X) the ONE box which best describes this establishment's PRINCIPAL method of selling.

300

Selling at this establishment	1
Mail order (catalog selling)	2
House-to-house or telephone (direct selling)	3
Operating merchandise vending machines	6

PENALTY FOR FAILURE TO REPORT

CONTINUE ON PAGE 2 →

Item 11 – MERCHANDISE LINES

Report sales either in dollar figures (see example on page 1), or as a percent (in whole percents) of total sales (see example below).

If figure is 38,796 of total sales	Mil.	Thou.	Dol.	Percent
o Report whole percents →				39
Not acceptable →				38.76

Merchandise lines	Census use	Estimated sales during 1982 Mil. Thou. Dol.	Percent
1. Automotive tires, tubes, batteries, parts, accessories (Report parts installed in repair on line 26b.)	320	321	322
a. New automobile tires and tubes	746		
b. New truck and bus tires (include industrial, off-the-road, and farm tractor tires.)	747		
c. Retreaded automobile tires	748		
d. Retreaded truck and bus tires (include industrial, off-the-road, and farm tractor tires.)	749		
e. Storage batteries	751		
f. Automotive parts (over-the-counter)	752		
g. Automotive accessories and sundry supplies (include polishes, paint, decorative items, etc.)	753		
h. Total – Sum of lines 1a through 1g →	748		
2. Automotive fuels	720		
3. Automotive lubricants (oil, greases, etc.)	730		
4. Automobiles, vans, trucks, other powered transportation vehicles (include motorcycles, motor scooters, motorbikes.)	700		
5. Household fuels (oil, LP gas, wood, coal)	780		
6. Major household appliances (Report parts installed in repair on line 26b.)			
a. Kitchen appliances, parts, accessories (include refrigerators, freezers, dishwashers, microwave ovens, etc.)	301		
b. Laundry appliances, parts, accessories (clothes washers and dryers)	302		
c. Other major household appliances, parts, accessories (include room air conditioners, dehumidifiers, vacuum cleaners, sewing machines, etc.)	303		
d. Total – Sum of lines 6a through 6c →	300		
7. Small electric appliances	310		
8. Televisions, video recording devices, video tapes, etc. (Include parts and accessories. Report parts installed in repair on line 26b.)	320		
9. Audio equipment, musical instruments and supplies (include radios, record players, tape recorders and players, components, parts, accessories, tapes, records, sheet music, etc. Report parts installed in repair on line 26b.)	330		
10. Furniture, sleep equipment	340		
11. Floor coverings	360		
12. Kitchenware and home furnishings	360		
13. Sporting goods and trophies (include bicycles, boats and motors, snowmobiles, go-carts, parts and accessories, etc.)	500		
14. Recreational vehicles	580		
15. Mobile homes	590		
16. Hardware, tools, plumbing and electrical supplies	600		
17. Lawn and garden equipment and supplies, cut flowers, plants, shrubs, fertilizers, etc.	620		
18. Lumber, building materials, paint, and home improvement equipment and supplies	640		
19. Jewelry (include novelty jewelry and watches.)	680		
20. Hobby and craft goods, games (including video), wheel goods (Report bicycles on line 13.)	852		
21. Toys (Report bicycles on line 13.)	851		
22. Men's and boys' wear	200		
23. Women's, girls', infants', and toddlers' wear	220		
24. Footwear (include accessories.)	260		
25. All other merchandise	890		
(If sales of merchandise on line 25 exceed 9% of total, specify principal lines and estimated sales.)			
a.	891		
b.	892		
c.	893		
26. All nonmerchandise receipts from customers, EXCLUDING SALES AND OTHER TAXES			
a. Labor charges	900		
b. Parts installed in repair	912		
c. All other nonmerchandise receipts (include receipts from customers for storage, rental or lease of tools and equipment, etc.)	916		
d. Total – Sum of lines 26a through 26c →	900		
27. TOTAL – Should equal Item 5 if reporting in dollars →	990		100%

Item 12 – SPECIAL INQUIRIES

a. Was 50 percent or more of this establishment's 1982 sales and receipts (item 5) derived from retreading tires at this establishment? 390 1 YES 2 NO

b. Did this establishment have any automotive service bays as of December 31, 1982? 380 1 YES 2 NO
If "Yes," complete c
If "No," SKIP to item 13

c. How many automotive service bays did this establishment have? Number 381

Answer item 13 only if your Census File Number (CFN), shown in the address label of this report form, begins with a zero.

Item 13 – OWNERSHIP, CONTROL, AND LOCATIONS OF OPERATION

a. Is this company owned or controlled by another company? 097 1 YES → 2 NO
ENTER OWNING OR CONTROLLING COMPANY NAME ADDRESS AND ZIP CODE
EI Number (9 digits)

b. Does this company own or control any other company or companies? 098 1 YES → 2 NO
ENTER ONE CONTROLLED COMPANY NAME ADDRESS AND ZIP CODE
EI Number (9 digits)

c. How many establishments were operated under the EI Number shown in the address label (or as corrected in item 1) at the end of 1982? → Number 079

If more than one, provide the physical location address and other information indicated below for each establishment. Continue with same format in item 14 (or attach a separate sheet) if necessary.

	NAME, ADDRESS AND ZIP CODE / KIND OF BUSINESS (DESCRIPTION)	1982	Mil.	Thou.	Dol.
1		Sales 081			
		Annual payroll 082			
		Census use 086			
2		Sales 081			
		Annual payroll 082			
		Census use 086			
3		Sales 081			
		Annual payroll 082			
		Census use 086			
4		Sales 081			
		Annual payroll 082			
		Census use 086			

Item 14 – REMARKS

Please use this space for any explanations that may be essential in understanding your reported data.

Item 15 – CERTIFICATION

This report is substantially accurate and has been prepared in accordance with instructions.

Period covered by this report →	FROM Mo. Year	TO Mo. Year
Name of person to contact regarding this report (Print or type)		
Telephone →	Area code / Number	Extension
Signature of authorized person		Date

FORM CB 82-1

EXHIBIT 5: BLS FORM 3023-A7, INDUSTRIAL CLASSIFICATION STATEMENT: WHOLESALE TRADE

This form is used for industry coding in the employment and wages (ES-202) system.

The complete form is shown. This is one of several versions tailored to particular SIC divisions; the one shown is for wholesale trade. The form is used for updating classification information for employers already in the system, which is now being done every 3 years for most employers.

Bureau of Labor Statistics
Industry Classification Statement
Wholesale Trade

U.S. Department of Labor

The information collected on this form by the Bureau of Labor Statistics and the State agencies cooperating in its statistical programs will be held in confidence and will be used for statistical purposes only.

This report is authorized by law 29 U.S.C. 2. Your voluntary cooperation is needed to make the results of this survey comprehensive, accurate, and timely.

OFFICE USE ONLY			
1972 SIC	Ownership	County Code	Aux.

TO AVOID DUPLICATE MAILINGS, please return as soon as possible in the enclosed envelope, to the address below.

Telephone Number

A. This report will be used to ensure the proper industrial and area classification of the establishment(s) for which you reported employment and payroll data on the **QUARTERLY EMPLOYER'S TAX REPORT** under the unemployment insurance account number shown above.

Please complete all the questions on this form as completely and accurately as possible and return promptly. Descriptions of principal activities usually may be brief, but for some industries require some detail. For example: Distribution of groceries and related products—specify is a general line, or specialized products such as coffee (green or roasted), canned goods, packaged frozen meat; distribution of ophthalmic goods—specify if activity includes grinding prescription lenses for retailers; distribution of petroleum products—specify if with or without bulk station or terminal; wholesaling of tools and hardware, specify type.

B. Principal Products or Activities during most recent calendar year in this establishment(s)	Percent of total sales (value or receipts) this period	For State use
1. Wholesale Trade (specify below)		
2. Other Products or Activities (specify below)		

C. Type of ownership (check one)

Private	Government	
1. ☐ Corporate	3. ☐ Federal	5. ☐ Local
2. ☐ Noncorporate	4. ☐ State	6. ☐ International

D. Is the establishment primarily engaged in performing services for other units of the company? ☐ Yes ☐ No

If yes, indicate nature of activity of this establishment.

1. ☐ Central administrative office
2. ☐ Research, development, or testing
3. ☐ Storage (warehouse)
4. ☐ Other (specify)

E. To whom are most of your products or services sold?

1. ☐ General Public
2. ☐ Retailers
3. ☐ Construction Contractors
4. ☐ Wholesalers
5. ☐ Other (specify)

F. Is this establishment part of a multi-unit company?

1. ☐ Yes 2. ☐ No If no, skip to question H on the back of this form.

If yes, enter name and location of controlling company on the space provided below and complete the questions on the back of this form.

BLS 3023-A7 (Rev. Dec. 1982)

EXHIBIT 6: FTC FORM 59-103, NATURE OF BUSINESS REPORT

This form is used for industry coding in the Quarterly Financial Report (QFR) Program.

Only the first page of the form is shown. (The second page covers the corporate structure and organization--parents, subsidiaries, changes, etc.--of the unit responding.) The form is used both for new corporations entering the sample and for updating the classification of units remaining on the sample for more than 2 years. This version is used for corporations in manufacturing; a second version is used for the other SIC divisions included in the program.

APPROVED BY GAO XXXXXXXXXX)
B-180229 (R0456)XXXXXXXXXX)
EXPIRES 6-30-83

NATURE OF BUSINESS REPORT

IMPORTANT: You are advised that unless this report is filed within the prescribed time limit below, your corporation may be subject to compulsory legal process (15 U.S.C. 46). It will be afforded confidential status. COMPLETE EACH ITEM 1 THROUGH 9. Omissions and inconsistencies will result in correspondence which is both costly and time consuming. WITHIN 7 DAYS, complete and return one copy to:

FEDERAL TRADE COMMISSION-ESQ
DIVISION OF FINANCIAL STATISTICS
WASHINGTON, D.C. 20580

Please read enclosed Rules for Consolidation before completing this report. **CONSOLIDATE every domestic corporation which is owned more than 50 percent by your corporation and its majority-owned corporations, EXCEPT those explicitly excluded.**

YOUR REPORT IS URGENTLY NEEDED

PLEASE SUBMIT IMMEDIATELY

(Please correct if name or address has changed.)

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXX XXXXXXX XXXXXXXX

1' 1

(In all correspondence, refer to number at right of address.)

1. For its latest accounting year ended ______________ (Month, day, year) the reporting company's:
   a. Total assets were $______________
   b. Total gross receipts were $______________

2. Specify the particular type of operation which most clearly describes the primary business activity of this company (for example: book publisher, contractor of women's house dresses, manufacturer of machine tool accessories, etc.): ______________

3. Total gross receipts reported in item 1b above were derived from *(list all sources, using attachments as necessary)*:

Source of Gross Receipts	Estimated Percent of Gross Receipts	Materials Used
List products made, processed, or assembled by this company with its own facilities (include contract work done for others on their materials):		Principal raw materials used in production (indicate form in which purchased):
a. ______	______%	______
b. ______	______%	______
c. ______	______%	______
d. ______	______%	______
e. ______	______%	______
f. ______	______%	______
g. ______	______%	______
List products made, processed, or assembled for this company by others (from materials owned by this company):		Materials from which products were made:
h. ______	______%	______
i. ______	______%	______
j. ______	______%	______
List products bought and resold without further processing or assembling by this company:		
k. ______	______%	XXXXXXXXXXXXXXX
l. ______	______%	XXXXXXXXXXXXXXX
List all other sources of income:		
m. ______	______%	XXXXXXXXXXXXXXX
n. ______	______%	XXXXXXXXXXXXXXX
TOTAL GROSS RECEIPTS	100 %	XXXXXXXXXXXXXXX

4. The primary processes performed by or for this company in making the products listed on lines 3(a) through 3(j) above were:

____

____

NB-1
FTC Form 50-103 (rev. 09-79)

OVER

EXHIBIT 7: BEA FORM BE-12, BENCHMARK SURVEY OF FOREIGN DIRECT INVESTMENT IN THE U. S., 1980, PART I, ITEMS 23-33

This form is used for industry coding in the foreign direct investment system.

Shown is page 3 of a form used in a baseline survey, conducted at approximately 5-year intervals to collect data for U.S. affiliates of foreign persons (firms or governments). This part of the form is used to determine the overall industry classification for the unit responding. Note that respondents are asked to enter an industry code for each 3-digit industry accounting for significant sales or revenues.

Part I – IDENTIFICATION OF U.S. AFFILIATE (Continued)

Industry classification of fully consolidated U.S. affiliate (based on sales or gross operating revenues) – Enter the appropriate 3-digit industry code(s) and the sales (as defined in Item 56) associated with each code. For a full explanation of each code, see the *Direct Investment Industry and Foreign Trade Classifications Booklet*. If you use fewer than eight codes you must account for total sales. For an inactive affiliate, show the industry classification(s) pertinent to the last active period; for "start-ups" with no sales, show the intended activity(ies).

		DI Industry code (1)	Sales (2) Bil. Mil. Thous. Dols.
23. Enter code with largest sales	1164		
24. Enter code with second largest sales	1165		
25. Enter code with third largest sales	1166		
26. Enter code with fourth largest sales	1167		
27. Enter code with fifth largest sales	1168		
28. Enter code with sixth largest sales	1169		
29. Enter code with seventh largest sales	1170		
30. Enter code with eighth largest sales	1171		
31. Sales accounted for (Sum of items 23 through 30)	1172		
32. Sales not accounted for above	1173		
33. Total sales – Sum of items 31 and 32 and must equal item 56. →	1174		
BEA USE ONLY	1175		

SUMMARY OF DIRECT INVESTMENT (DI) INDUSTRY CLASSIFICATIONS

The titles of some DI industry classifications by themselves, may be insufficient to classify certain activities. Consult the more detailed descriptions provided in the *Direct Investment Industry and Foreign Trade Classifications Booklet* to be sure of the correct classification.

AGRICULTURE, FORESTRY, AND FISHING

- 010 Agricultural production – crops
- 020 Agricultural production – livestock, except beef cattle feedlots
- 021 Agricultural production – beef cattle feedlots
- 070 Agricultural services
- 080 Forestry
- 090 Fishing, hunting, and trapping

MINING

- 101 Iron ores
- 102 Copper, lead, zinc, gold, and silver ores
- 103 Bauxite and other aluminum ores
- 109 Other metallic ores and metal mining services
- 120 Coal
- 133 Crude petroleum extraction (no refining) and natural gas
- 138 Oil and gas field services
- 140 Nonmetallic minerals, except fuels

CONSTRUCTION

- 150 Construction

MANUFACTURING

- 201 Meat products
- 202 Dairy products
- 203 Preserved fruits and vegetables
- 204 Grain mill products
- 205 Bakery products
- 208 Beverages
- 209 Other food and kindred products
- 210 Tobacco manufactures
- 220 Textile mill products
- 230 Apparel and other textile products
- 240 Lumber and wood products
- 250 Furniture and fixtures
- 262 Pulp, paper, and board mills
- 264 Miscellaneous converted paper products
- 265 Paperboard containers and boxes
- 270 Printing and publishing
- 281 Industrial chemicals and synthetics
- 283 Drugs
- 284 Soap, cleaners, and toilet goods
- 285 Paints and allied products
- 287 Agricultural chemicals
- 289 Chemical products, nec
- 291 Integrated petroleum refining and extraction
- 292 Petroleum refining without extraction
- 299 Petroleum and coal products, nec
- 305 Rubber products
- 307 Miscellaneous plastics products
- 310 Leather and leather products

MANUFACTURING – Continued

- 321 Glass products
- 329 Stone, clay, concrete, gypsum, and other nonmetallic mineral products
- 331 Primary metal industries, ferrous
- 335 Primary metal industries, non-ferrous
- 341 Metal cans and shipping containers
- 342 Cutlery, hand tools, and hardware
- 343 Metal plumbing fixtures and heating equipment, except electric
- 344 Fabricated structural metal products
- 345 Screw machine products, bolts, etc.
- 346 Metal forgings and stampings
- 349 Fabricated metal products, nec, ordnance, and metal services
- 351 Engines and turbines
- 352 Farm and garden machinery
- 353 Construction, mining, and materials handling machinery
- 354 Metalworking machinery
- 355 Special industry machinery
- 356 General industrial machinery
- 357 Office and computing machines
- 358 Refrigeration and service industry machinery
- 359 Machinery, except electrical, nec
- 363 Household appliances
- 364 Electric lighting and wiring equipment
- 366 Radio, television, and communication equipment
- 367 Electronic components and accessories
- 369 Electrical machinery, nec
- 371 Motor vehicles and equipment
- 379 Other transportation equipment
- 381 Scientific and measuring instruments
- 383 Optical and ophthalmic goods
- 384 Medical instruments and supplies
- 386 Photographic equipment and supplies
- 387 Watches, clocks, and watchcases
- 390 Miscellaneous manufacturing industries, nec

TRANSPORTATION, COMMUNICATION AND PUBLIC UTILITIES

- 401 Railroads
- 441 Petroleum tanker operations
- 449 Other water transportation
- 450 Transportation by air
- 461 Pipe lines, petroleum and natural gas
- 462 Pipe lines, except petroleum and natural gas
- 470 Petroleum storage for hire
- 478 Transportation, nec, warehousing, terminal facilities, travel agents, and related services
- 480 Communication
- 490 Electric, gas, and sanitary services

WHOLESALE TRADE

- 501 Motor vehicles and equipment
- 503 Lumber and construction materials
- 504 Farm and garden machinery, equipment and supplies
- 505 Metals and minerals, except petroleum
- 506 Electrical goods
- 507 Hardware, plumbing and heating equipment and supplies
- 508 Other machinery, equipment, and supplies
- 509 Durable goods, nec
- 511 Paper and paper products
- 512 Drugs, proprietaries, and sundries
- 513 Apparel, piece goods, and notions
- 514 Groceries and related products
- 515 Farm-product raw materials
- 517 Petroleum and petroleum products
- 519 Nondurable goods, nec

RETAIL TRADE

- 540 Food stores and eating and drinking places
- 554 Gasoline service stations
- 590 Retail trade, nec

FINANCE, INSURANCE, AND REAL ESTATE

- 600 Banking
- 610 Finance, except banking
- 630 Insurance
- 640 Lessors of agricultural and forestry real estate
- 650 Real estate, nec
- 671 Holding companies

SERVICES

- 700 Hotels and other lodging places
- 731 Advertising
- 732 Business services, nec
- 780 Motion pictures, including television tape and film
- 891 Engineering, architectural, and surveying services
- 893 Accounting, auditing, and bookkeeping services
- 898 Services, nec, provided on a commercial basis

NONBUSINESS ENTITIES

- 900 Government entities
- 905 Nonbusiness entities, except Government

NOTE nec means not elsewhere classified

REFERENCES

American Statistical Association
1980 Business directories: findings and recommendations of the ASA Committee on Privacy and Confidentiality. The American Statistician 34:8-11.

Bailar, B. A., and Kallek, S.
1980 Evaluation of the 1977 Economic Censuses. Pp. 1-7 in Proceedings of the Section on Survey Research Methods. Washington, D.C.: American Statistical Association.

Boyes, B. A., and Brown, J. R.
1974 Quality measurement and quality of sampling frames. Pp. 162-166 in Proceedings of the Business and Economic Statistics Section. Washington, D.C.: American Statistical Association.

Buckler, W., Cartwright, D., and Levine, B.
1983 An Update on Establishment Reporting Issues: Practical Considerations. Paper presented at the annual meeting of the American Statistical Association, Toronto, Canada.

Bureau of the Budget
1947 History of the Development of Industrial Classification of the United States. Memorandum from V. S. Kolesnikoff to members of the Interagency Committee on Development and Application of Standard Industrial Classification, March 14. Bureau of the Budget, Washington, D.C.
1961 Brief History of the Movement in the Federal Government for a Central Directory and of Related Efforts Aimed at Improving Quality and Comparability of Economic Statistics. Unpublished report. Office of Statistical Standards, Bureau of the Budget, Washington, D.C.

Bureau of the Census
1951 Effect of Different Documents, Coders and Time Periods on Industry Classification of Establishments. Memorandum from Morris H. Hansen to Thomas J. Mills, August 7. Attached report with same title by H. T. Goldstein, July 25. U.S. Department of Commerce, Washington, D.C.

Bureau of the Census

1963 Reconciliation of the 1958 Census of Retail Trade with the Monthly Retail Trade Report. Technical Paper No. 9. Washington, D.C.: U.S. Department of Commerce.

1965a Evaluation and Research Program of the U.S. Censuses of Population and Housing 1960: The Employer Record Check. Series ER 60, No. 6. Washington, D.C.: U.S. Department of Commerce.

1965b Final Results of BES-Census Retail Payroll Reconciliation for the State of Delaware. Memorandum from Peter Ohs and Ralph Woodruff to Harvey Kailin and William Hurwitz, July 22. U.S. Department of Commerce, Washington, D.C.

1968 Analysis of 1963 Economic Censuses Changes of Mailing List Industry Codes--Single-Unit Establishments with Payroll. Memorandum from Harold T. Goldstein to Owen C. Gretton and Harvey Kailin, October 9. U.S. Department of Commerce, Washington, D.C.

1969 Comparison of Census SIC Codes with Internal Revenue PIA Codes for Same Establishments. Memorandum from George Minton to Joseph F. Daly, November 12. U.S. Department of Commerce, Washington, D.C.

1971 The Annual Survey of Manufactures: A Report on Methodology. Technical Paper No. 24. Washington, D.C.: U.S. Department of Commerce.

1972a Evaluation and Research Program of the U.S. Censuses of Population and Housing 1960: Effects of Coders. Series ER 60, No. 9. Washington, D.C.: U.S. Department of Commerce.

1972b From the Old to the New SIC for Establishments. Technical Paper No. 28. Washington, D.C.: U.S. Department of Commerce.

1977a Evaluation and Research Program of the 1970 Census of Population and Housing: The Employer Record Check. PHC(E)-12. Washington, D.C.: U.S. Department of Commerce.

1977b 1977 Industry and Product Classification Manual. Washington, D.C.: U.S. Department of Commerce.

1979 The Standard Statistical Establishment List Program. Technical Paper No. 44. Washington, D.C.: U.S. Department of Commerce.

1981a County Business Patterns 1968-1977: 10-Year History. CBP 77-55. Washington, D.C.: U.S. Department of Commerce.

1981b 1977 Enterprise Statistics: General Report on Industrial Organization. Series ES-77, No. 1. Washington, D.C.: U.S. Department of Commerce.

1982a Special Tabulation: Number of Single-Unit and Multi-Unit Establishments, and Number of Employees, by Trade Area, Data from 1979 SSEL. U.S. Department of Commerce, Washington, D.C.

1982b From the EI Master Sample of the 1981 BMF Update, the Weighted Number of Single Unit Establishments with Current Year Payroll by LFO [legal form of organization] by Source Code for Different Division Codes. Computer printout. U.S. Department of Commerce, Washington, D.C.

Bureau of Economic Analysis
1972 An Evaluation of the Usefulness of the Social Security Administration's Continuous Work History Sample. Report prepared for the Manpower Administration, U.S. Department of Labor. Washington, D.C.: U.S. Department of Commerce.
1976 Regional Work Force Characteristics and Migration Data: A Handbook on the Social Security Continuous Work History Sample and Its Applications. Washington, D.C.: U.S. Department of Commerce.

Bureau of Labor Statistics
1974 Handbook on Standard Industrial Classification 1972 (revised). Washington, D.C.: U.S. Department of Labor.
n.d. Standard Industrial Classification Coding Workbook. Office of Field Operations, Program Related Training. Washington, D.C.: U.S. Department of Labor.

Cys, K., Hinkins, S. and Rehula, V.
1982 Automatic and manual edits for corporation tax returns. Pp. 443-448 in Proceedings of the Section on Survey Research Methods. Washington, D.C.: American Statistical Association.

Dun & Bradstreet
n.d. Assigning SIC Numbers. New York: Dun & Bradstreet.

Energy Information Administration
1983 Part 9--Price. Monthly Energy Review 83(2):81-94.

Farrell, M. G., Jabine, T. B., and Konschnik, C. A.
1982 A review of industry coding systems. Pp. 278-283 in Proceedings of the Section on Survey Research Methods. Washington, D.C.: American Statistical Association.

Farrell, M. G., Richardson, W. J., and Wikoff, J. R.
1983 1982 Economic and Agricultural Censuses: Automation and Other Improvements. Paper presented at the annual meeting of the American Statistical Association, Toronto, Canada.

Federal Statistics Users' Conference
1982 No revision will be made in Standard Industrial Classification in 1982. Newsletter (23)2:10-11.

Hanczaryk, R. S., and Sullivan, J. M.
1980 Evaluation of coverage of the administrative records frame for the 1977 economic censuses--employer segment. Pp. 154-159 in Proceedings of the Section on Survey Research Methods. Washington, D.C.: American Statistical Association.

Harris, C. S.
1981 A Comparison of Employment Data for Several Business Data Sources: County Business Patterns, Unemployment Insurance, and the Brookings' U.S. Establishment and Enterprise Microdata File. Washington, D.C.: The Brookings Institution.

Hostetter, S. C.
1983 The Verification Method as a Solution to an Industry Coding Problem. Paper presented at the annual meeting of the American Statistical Association, Toronto, Canada.

Internal Revenue Service
1973 A Comparision of Industry Coding of Tax Returns During Revenue Processing and Statistical Processing. Unpublished report by Linda Bouchard, Statistics Division, January 15. U.S. Department of the Treasury, Washington, D.C.
1974 Partnerships: A Comparison Study of the Principal Industrial Activity Code with the Statistics of Income Industry Code. Unpublished report, business statistics staff, Statistics Division, December 11. U.S. Department of the Treasury, Washington, D.C.
1975 Possible and Impossible BMF Industry Codes, 1973. Unpublished tabulation, Statistics Division. U.S. Department of the Treasury, Washington, D.C.
1981 Statistics of Income 1977: Sole Proprietorship Returns. Statistics Division. Washington, D.C.: U.S. Department of the Treasury.
1982a A Pilot Study of the Extent and Accuracy of EIN and Business Activity Reporting on Tax Returns by Sole Proprietors. Unpublished report, Statistics Division. U.S. Department of the Treasury, Washington, D.C.
1982b Statistics of Income--1978-79: Corporation Income Tax Returns. Statistics Division. Washington, D.C.: U.S. Department of the Treasury.

Jabine, T. B., and Tepping, B. J.
1973 Controlling the quality of occupation and industry data. Bulletin of the Internatonal Statistical Institute 45(3):360-89.

King, C., and Ricketts, E.
1980 Evaluation of the use of administrative record data in the economic censuses. Pp. 175-180 in Proceedings of the Section on Survey Research Methods. Washington, D.C.: American Statistical Association.

Lakatos, E.
n.d. Automated I & O Coding. Unpublished report. Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

Levine, B.
1980 Improving industry and place of work coding in the Continuous Work History Samples. Pp. 472-477 in Proceedings of the Section on Survey Research Methods. Washington, D.C.: American Statistical Association.

Metropolitan Washington Council of Governments
1977 Comparison of COG's Regional Employment Census (REC) Counts by Jurisdiction with State Employment Security Estimates and Other Sources of County-Level Employment. Memorandum from Stuart Bendelow to Project Advisory Committee, November 11. Washington, D.C.

Morgenstern, O.
1963 On the Accuracy of Economic Observations, 2nd ed. Princeton, N.J.: Princeton University Press.

National Research Council
1979 Measurement and Interpretation of Productivity. Panel to Review Productivity Statistics, Committee on National Statistics, Assembly of Behavioral and Social Sciences. Washington, D.C.: National Academy of Sciences.

Office of Federal Statistical Policy and Standards
1977a Gross National Product Data Improvement Project Report: Report of the Advisory Committee on Gross National Product Data Improvement. Washington, D.C.: U.S. Department of Commerce.
1977b Standard Industrial Classification Manual, 1977 Supplement. Washington, D.C.: U.S. Department of Commerce.
1980 Report on Statistical Uses of Administrative Records. Statistical Policy Working Paper 6. Washington, D.C.: U.S. Department of Commerce.

Office of Management and Budget
1972 Standard Industrial Classification Manual. Statistical Policy Division. Washington, D.C.: Office of Management and Budget.
1974 Enterprise Standard Industrial Classification Manual. Statistical Policy Division. Washington, D.C.: Office of Management and Budget.
1983 Statistical Confidentiality Legislation. Memorandum from Thomas D. Hopkins, Deputy Administrator for Regulatory and Statistical Analysis, to the Working Group on Economic Statistics, Cabinet Council on Economic Affairs, February 10.

O'Reagan, R. T.
1972 Computer-assigned codes from verbal responses. Communications of the ACM (Association for Computing Machinery) 15:455-459.

Powell, W. T., and Stubbs, J. R.
1981 Using Business Master File data for Statistics of Income purposes. Pp. 471-744 in Proceedings of the Section on Survey Research Methods. Washington, D.C.: American Statistical Association.

President's Reorganization Project for the Federal Statistical Program
1981 Improving the federal statistical system: issues and options. Statistical Reporter 81-5:133-221.

Sailer, P., Kilss, B., and Crabbe, P.
1983 Occupation Data from Tax Returns: Another Progress Report. Paper presented at the annual meeting of the American Statistical Association, Toronto, Canada.

Sailer, P., Orcutt, H., and Clark, P.
1980 Coming soon: taxpayer data classified by occupation. Pp. 187-191 in Proceedings of the Section on Survey Reseach Methods. Washington, D.C.: American Statistical Association.

Securities and Exchange Commission
1981 Directory of Companies Required to File Annual Reports with the Securities and Exchange Commission. December 31. Washington, D.C.: Securities and Exchange Commission.

Simmons, W. R.
1953 The elements of an industrial classification policy. Journal of the American Statistical Association 48(263):429-439.

Singelmann, J.
1978 From Agriculture to Service, the Transformation of Industrial Employment. Beverly Hills, Calif.: Sage Publications.

U.S. General Accounting Office
1979 After Six Years, Legal Obstacles Continue to Restrict Government Use of the Standard Statistical Establishment List. Report to the Congress by the Comptroller General of the United States. Washington, D.C.: U.S. Government Printing Office.