

Statistical Policy Working Paper 11 - A Review of Industry Coding Systems



Statistical Policy Working Paper 11

MEMBERS OF THE FEDERAL COMMITTEE ON
STATISTICAL METHODOLOGY

(February 1984)

Maria Elena Gonzalez (Chair)

Charles D. Jones

Office of information and

Bureau of the Census

Regulatory Affairs (OMB)

(Commerce)

Barbara A. Bailar

Daniel Kasprzyk

Bureau of the Census

Bureau of the Census

(Commerce)

(Commerce)

Norman D. Beller

William E. Kibler

National Center for Education

Statistical Reporting Service

Statistics (Education)

(Agriculture)

Yvonne M. Bishop

David A. Pierce

Energy Information

Federal Reserve Board

Administration (Energy)

Edwin J. Coleman

Thomas Plewes

Bureau of Economic Analysis

Bureau of Labor Statistics

(Commerce)

(Labor)

John E. Cremeans

Fritz Scheuren

Bureau of industrial Economics

Internal Revenue Service

(Commerce)

(Treasury)

Zahava D. Doering

Monroe G. Sirken

Defense Manpower Data Center

National Center for Health

(Defense)

Statistics (Health and

Human Services)

Maria D. Eldridge

Thomas G. Staples

National Center for Education

Social Security Administration

Statistics (Education)

(Health and Human Services)

Daniel E. Garnick

Robert D. Tortora

Bureau of Economic Analysis

Statistical Reporting Service

(Commerce)

(Agriculture)

OFFICE OF INFORMATION AND REGULATORY AFFAIRS

Christopher DeMuth, Administrator

Thomas D. Hopkins, Deputy Administrator for

Regulatory and Statistical Analysis

Dorothy M. Tella, Chief, Statistical Policy Office

Maria E. Gonzalez, Chairperson

Federal Committee on Statistical Methodology

PREFACE

The Working Group on Industry Coding was initiated by the Administrative Records Subcommittee of the Federal Committee on Statistical Methodology to review the various existing industry coding systems and study their relationships, comparability and accuracy. The report presents information on the principles and procedures used to classify and code business establishments by industry within the framework of the Standard Industrial Classification (SIC) system.

This report is intended primarily for Federal agencies that are responsible for industry coding. However, users of data classified by industry should also find it valuable to know more about the coding

procedures and practices that affect the quality of the data.

The findings and recommendations of this report emphasize the need for increased interagency cooperation to improve the quality and comparability of industry codes and reduce the cost and respondent burden of multi-agency coding efforts. A permanent interagency committee is recommended as the mechanism for coordinating improvements in industry coding systems.

Implementation of the recommendations in this report will be explored by the Statistical Policy Office. The report does not necessarily reflect the views of the Office of Management and Budget.

The working Group was chaired by Carl A. Ronschnik, Bureau of the Census, Department of Commerce; the Administrative Records Subcommittee is chaired by Fritz Scheurent Internal Revenue Service.

MEMBERS OF THE ADMINISTRATIVE RECORDS SUBCOMMITTEE

(December 1983)

Fritz Scheuren (Chairman)	Daniel Kasprzyk
Internal Revenue Service	Bureau of the Census
(Treasury)	(Commerce)
Faye Aziz	Beth Kilss
Social Security Administration	Internal Revenue Service
(Health and Human Services)	(Treasury)
Warren Buckler	Carl A. Ronschnik
Social Security Administration	Bureau of the Census
(Health and Human Services)	(Commerce)
Paul Burke	Bruce Levine
Department of Housing and	Bureau of Economic Analysis
Urban Development	(Commerce)

David W. Cartwright

Brian MacDonald

Bureau of Economic Analysis

Bureau of Labor Statistics

(Commerce)

(Labor)

Charles Cowan

James Millette

Bureau of the Census

Bureau of Labor Statistics

(Commerce)

(Labor)

Maria E. Gonzalez

Douglas Sater

Office of Management and

Bureau of the Census

Budget

(Commerce)

David A. Hirschberg

Michael Searson

Small Business Administration

Bureau of Labor Statistics (Labor)

Susan Hostetter

Linda Bouchard Taylor

Bureau of Labor Statistics

Internal Revenue Service

(Labor)

(Treasury)

Thomas B. Jabine

Alan Zempel

Consultant to Committee on Internal Revenue Service

National Statistics (Treasury)

(National Academy of Sciences)

ii

MEMBERS OF THE

INDUSTRY CODING WORKING GROUP

Carl A. Konschnik, Chairperson

Bureau of the Census

(Commerce)

Linda M. Dill

Social Security Administration

(Health and Human Services)

Susan Hostetter

Bureau of Labor Statistics

(Labor)

Thomas B. Jabine

Consultant to Committee on

National Statistics

(National Academy of Sciences)

Beth Kilss

Internal Revenue Service

(Treasury)

Bruce Levine

Bureau of Economic Analysis

(Commerce)

James Millette

Bureau of Labor Statistics

(Labor)

Linda Bouchard Taylor

Internal Revenue Service

(Treasury)

Alan Zempel

Internal Revenue Service

(Treasury)

iii

ACKNOWLEDGMENTS

The idea for this study grew out of the collective interest of the members of the Administrative Records Subcommittee in looking at industry coding issues.

Data for the 16 major industry coding systems reviewed were first

collected from the agencies on a questionnaire prepared by the Working Group. The questionnaire responses and associated documentation were then used to prepare "system descriptions" following a standard format developed by Thomas B. Jabine. Copies of system descriptions, which are in A supplement to this report entitled Description of Selected Industry Coding Systems, may be obtained from the Statistics of Income Division, Internal Revenue Service, D:R:S, 1111 Constitution Avenue, N.W., Washington, D.C. 20224.

In addition to the members of the Working Group, the following persons contributed to the completion of the questionnaires and system descriptions;

Bureau of the Census: Alfred R. Brand, Patricia A. Clark,

Stanley M. Hyman, C. Harvey Monk, Walter E. Neece, Frank M.

Hartman

Bureau of Economic Analysis: George R. Kruer

Internal Revenue Service: Bertie Brame, John Maiden, Patrick

Piet, Paul J. Rose, Nathan Shaifer, Raymond Wolfe

Social Security Administration: Cheryl Williams

The working paper was reviewed by all members of the Working Group.

The chapters were initially drafted by:.

I. Susan Hostetter James Millette

II. Carl A. Konschnik, Bruce Levine

III. Linda M. Dill, Carl A. Konschnik

IV. Thomas B. Jabine

The entire Working Group provided comments to the initial drafts. The final wording was reviewed by the Working Group. Maria E. Gonzalez met with the Working Group throughout its term. Fritz Scheuren and members of the Administrative Records Subcommittee provided additional assistance and encouragement, as did members of the Federal Committee

In the preparation of this working paper, substantial use was made of the following sources:

1. Farrell, M.G., Jabine, T.B., and Konschnik, C.A.

1982 A review of industry coding systems. Proceedings of the Section on Survey Research Methods, American Statistical Association.

2. Jabine, T.B.

1984 The Comparability and Accuracy of Industry Codes in Different Data Systems (in draft). Committee on National Statistics. Commission on Behavioral and social sciences and Education. Washington, D.C.: National Academy of Sciences.

The second item is scheduled for publication in 1984. Several excerpts from it were used directly or with minor changes in Chapters III, IV and VI of this working paper.

v

A REVIEW OF INDUSTRY CODING SYSTEMS

Table of Contents

Page

Preface.	i
Acknowledgments.	iv
1. Findings and Recommendations.	1

A.	Introduction	1
B.	Code Sharing	1
C.	Standardization of industry Coding Principles	2
D.	Standardization of Coding Procedures	5
E.	Documentation.	7
F.	Matching Studies and Other Research.	8
G.	Interagency Cooperation.10

II. Description of the Industry Coding Working .

Group Project.11	
A.	Introduction.11
B.	Scope of the Review14
C.	Major Uses of industry Coding Information15
D.	Composition and Objectives of the Industry Coding Working Group.17
E.	Development of the Basic Documentation for the Federal Industry Coding Systems.18

III. Industry Coding Systems and Their Relationships21
A. Introduction.21
B. Coverage.21
C. Frequency and Timing of initial Coding and Updating28
D. Classification System Used.28
E. Classification Principles30
F. Information Used as input to Coding35
G. Coding Procedures.42
H. Description of Systems Relationships48
IV. Quantitative Information on Comparability and Accuracy.53
A. Introduction53
B. Inter-system Macro-comparisons53
C. Inter-system Micro-comparisons: General.55
D. Interagency Comparisons Between Systems.56

E.	Intra-agency Comparisons Between Systems63
F.	Data on industry Coding Error in Individual Systems.72
V.	References89
VI.	Selected Source Documents and Instructions95

List of Tables

Table		Page
1	Selected Characteristics of Industry Coding Systems Reviewed.23
2	Coverage of Industry Coding Systems Reviewed by SIC Division27
3	Interagency Transfers of Industry Codes51

4	Results of independent Coding of Establishments by Census and SSA	5-7
5	Summary of Errors as a Result of Reconciling BES and Census Records on Delaware Retail Payroll in 1963.59
6	Results of Comparison between Final Industry Codes and SSA-based Mailing List Codes: 1963 Economic Censuses.62
7	An Analysis of 1947-1949 Code Changes for 500 Single-Unit Establishments in Manufacturing64
8	Indexes of Shift for In Scope and out of Scope of Retail Trade by Kind of Business.66
7	Major Kind-of-Business Cross-Classification of Group 1 Retail Trade Establishment Sales in Census and in Current Survey: United States, 1958.67
10	Differences between IRS Master File and Statistics of income	

	(SOI) Industry Classification by SIC Division and Type of organization.69
11	Percent of IRS Master File Codes Agreeing with SOI Codes, by Type of Organization and Level of Detail71
12	Agreement of IRS Master File Codes with SOI Codes at Major Industry Level for Corporations: Tax Years 1972 and 1973.72
13	IRS Statistics of income Program. Number of Incompletely Classified Returns by Industry Division and Type of organization75
14	Reasons for industry Code Differences between Initial and Recheck Surveys: Retail Trade Surveys, 194877

15	Indexes of inconsistency for Selected Major industries: 1960 and 197079
16	Comparison of SIC Codes Based on Census Questionnaires with those Based on Administrative Records: 1977 Economic Censuses.81
17	Evaluation of Published Statistics for Nonemployers in Contract Construction: 1977 Census82
18	Single-Unit Establishments in the SSEL with Current Year Payroll by SIC Division and Source: 1981.85

CHAPTER I

FINDINGS AND RECOMMENDATIONS

A. Introduction

This section presents the findings and recommendations of the Industry Coding Working Group. The recommendations are based on two goals:

1. To improve the quality and comparability of industry codes for all of the data systems reviewed by the working Group; and
2. To reduce the overall cost and respondent burden associated with initial industry coding and updating of codes for these systems.

Meeting these objectives requires increased interagency cooperation in the areas of standardization and code sharing (the transfer of industry codes, for individual establishments or other economic units from one data system to another). With respect to these two areas, the Working Group found that:

Significant improvements in quality and comparability of industry coding can be achieved by increased standardization of coding principles and procedures; however, a substantial increase in code sharing between agencies is needed to achieve the best results.

B. Code Sharing

Chapter III of this report describes the differences found by the Industry Coding Working Group in coding procedures, source documents, procedures for updating codes, and other features of the systems reviewed. These differences, which result in part from cost and respondent burden limitations, cause differences in the industry codes assigned to individual units. This applies both to statistical

data systems and to systems developed primarily for administrative purposes. Chapter IV presents quantitative evidence, from several studies, of differences resulting from system variations.

At present there are few transfers of industry codes between agencies. The primary transfers are from the Social Security Administration (SSA) and the Internal Revenue service (IRS) to the Census Bureau for use in the latter's economic statistics programs. (See Table 3 on page 51 for details.) The Working Group recommends that:

2

Agencies whose systems have been reviewed should expand industry code sharing to improve the quality of codes and to reduce code differences between systems.

Increased code sharing between agencies should lead to more comparable and accurate industry codes in major Federal and Federal/State cooperative data systems. Initially, there would be a

significant cost to develop a system 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, considerable savings could be realized by cutting back on independent data collection activities for assigning and updating industry codes. Currently various agencies collect similar information from the same respondents for use in determining industry codes. Thus the beneficial impact of code sharing between agencies on both respondent burden and cost should be extensive.

To implement the recommendation for code sharing fully will require changes in the confidentiality laws currently governing the Federal statistical community. Except for a few specific cases, agencies may not, under current law, disclose individually identifiable microdata outside their own agency.

C. Standardization of industry Coding Principles

The Working Group found that the agency coding systems reviewed all based their classification systems on the current version

of the SIC Manual, but that each of the systems departs from it in some respects. The nature of these departures from the SIC Manual is described in Chapter III of this report.

It is not clear that all systems would be in a position to follow the principles of the ST exactly in every respect. Administrative requirements and resource limitations may sometimes preclude this. Nevertheless, the Working Group believes that greater adherence to these principles is feasible in most cases, and recommends that:

All Federal and State agencies cooperating in Federal statistical programs that classify economic units (establishments or reporting units) by industrial activity should, to the greatest extent possible, follow the classification principles contained in the 1972 Standard Industrial Classification (SIC) Manual as amended by the 1977 Supplement.

Agencies using the SIC Manual as the basis for assigning industry codes to establishments or reporting units should adhere to the following recommendations on specific classification principles. The

specific recommendations do not necessarily apply for classifying enterprises or similar units.

3

1. The basic business unit should be the establishment as defined in the SIC Manual.

The establishment is normally an economic unit at a single physical location and engaged in one, or predominantly one, type of economic activity. Special rules apply where two or more distinct and separate activities are carried on at a common physical location.

The SIC manual is intended for assigning codes to establishments.

However, some agencies assign codes to similar but somewhat differently defined units-- reporting units. As a long range goal, these agencies should attempt to redefine their reporting units so that they are consistent with the establishment definition.

2. To the extent possible, all units should be classified by 4-

digit SIC industry, using all of the industries included in the current SIC Manual.

Most of the systems reviewed come close to following the SIC structure in the Manual, but use groupings of SIC industries in a few instances. Some aggregation occurs to avoid disclosure of individual establishment data. Some occurs because experience in some agencies shows that for certain industries adequate reporting records are not available on an industry-wide basis. Since different agencies aggregate for different reasons, varying groupings of industries result. Comparability of data by industry would be improved if participating agencies used all of the 4-digit SIC codes or could agree on and use a standard set of codes for grouped industries.

This recommendation is not intended to preclude the use of additional classifiers for the same units. However, classifiers such as those used for administrative or tax purposes should be clearly distinguished from codes based on the SIC. The assignment of SIC's should not be altered or controlled in any way by the assignment of such additional codes. Some agencies, primarily the Census Bureau,

assign industry codes in greater detail than provided by 4-digit SIC codes. This practice is acceptable as long as the detailed classifications are defined within 4-digit industries.

3. When an establishment or reporting unit has multiple activities, the SIC code should be determined according to the principles outlined in the SIC Manual.

This recommendation implies, among other things, that the treatment of multiple activities be based on the variables recommended in the SIC Manual to measure the relative importance of each activity and that 4-digit SIC codes be assigned to each

activity of the establishment. Also it is necessary to assign a percent of total value for each activity for which a 4-digit SIC was determined and then group activities with the same 4-digit SIC's and sum the percent values. The establishment's classification would then be the 4-digit SIC with the greatest percent of total activity.

4. Information that identifies Central Administrative offices (CAO's) and auxiliary units must be collected and reviewed to ensure accurate determination of 4-digit industry codes. All systems should incorporate this information.

As stated in the SIC Manual, a CAO is an establishment primarily engaged in management and general administrative functions performed centrally for other establishments of the same company. An auxiliary unit is an establishment primarily engaged in performing supporting services for other establishments of the same company rather than for the general public or for other business firms. Both CAO's and auxiliary units should be classified according to the primary 4-digit industry activity of the operating establishments they serve.

Additional classification codes describing the type of function performed also should be standardized. The Working Group recommends that agencies responsible for industry coding adopt a uniform set of auxiliary codes for the classification of CAO or auxiliary activities for use in their systems. The codes would delineate activities such as central administration; research and development; warehousing; data

processing; and repair shops.

5. Agencies should work together to arrive at consistent solutions to two problems generally encountered in classifying government operations-- determining ownership and distinguishing between operating and administrative operations.

Many activities are quasi-government and the distinctions between government and private industry are often unclear. Most agencies have guidelines for determining ownership that follow the SIC Manual concept of "owned and operated". However, very little coordination and sharing of the interpretation of the rules have occurred. Developing a system for sharing and comparing concepts would foster consistency among agencies.

The Public Administration division of the SIC Manual includes "...the legislative, judicial, administrative and regulatory activities of Federal, State, local and international governments." However, the government owned and operated

establishments outside of public administration properly should be classified according to the activities in which they are

engaged. Coordination and cooperation among agencies should enhance systematic identification and reporting according to these standards.

D. Standardization of Coding Procedures

This section presents recommendations to improve and standardize coding procedures used by the systems to implement industry coding principles. Coding procedures considered most important are those that relate to the use of source documents, quality assurance training for coders, and resistance principles.

Chapter III of this report describes source documents used by each of the systems reviewed. These source documents vary both in the level of detail requested and the format and wording of the items

included. This variability has clearly contributed to differences between the systems. Chapter VI contains examples of source documents.

Although it was beyond the scope of this Working Group to develop specific questionnaires or standards for questionnaires, the Working Group recommends that:

1. Agencies that do industry coding should work together to increase the uniformity of product, activity and related questions used in their source documents

The Working Group believes that accurate 4-digit industry coding requires questions specifically tailored to SIC division level and for some intermediate groupings of 4-digit industries. Since some agencies may not have the need or resources to use forms designed for specific industry groups, the Working Group suggests the development of two kinds of model source documents: a set for specific industry groups and an abridged general purpose version. Separate versions for initial coding and updating are also suggested.

The development of standardized source documents should be based on thorough research. The Working Group's recommendations for research on source documents are given in section F of this chapter.

This report provides some information on Quality Assurance in Chapter III. However, most of the agencies reviewed had limited information on specific quality assurance measures used for their systems. The systems reviewed show considerable variation in the scope and intensity of procedures for maintaining and improving the accuracy of industry codes. The Working Group recommends that:

2. Each agency should review the procedures it uses to assure the quality of industry coding and should try to upgrade them where needed.

Because technology (both in industries upon which codes are based and in the processing and procedures used by agencies when assigning codes) is changing rapidly, the Working Group suggests that one or

more interagency workshops be organized to discuss new developments in industry coding and to promote the exchange of information on coding procedures. Workshops should cover computerized coding (coding based on verbal descriptions or on quantitative product and service data), computer-assisted coding from activity descriptions, and computer consistency checks. Methods of reducing agency cost and respondent burden also should be examined.

The Working Group found that agencies doing industry coding did not have formal training programs for coders in some of their systems. SSA provides extensive formal training for new coders in their single-unit employer identification (EI) file system. This is followed up by on-the-job training and close quality review. The Census Bureau provides training for large groups of coding technicians during the economic censuses, and ,the Bureau of Labor Statistics (BLS) provides an ongoing training program for all State coding technicians. However, for some systems more on-the-job training and less of a formal program is used. The Working Group recommends that:

3. Agencies should provide periodic training

based on recommended coding course principles

and procedures for their SIC coders.

Such courses should include solutions, preferably those agreed upon by an interagency group, to coding problems arising from the development of new industries and from changes in existing industries.

Resistance principles generally take prior industry codes, and related data into account in determining a current code. The purpose of using them is to avoid erratic shifts back and forth from one industry to another and, in sample-based systems, to help control sampling variability. Lack of uniformity in the use of resistance principles has been one of many causes of industry classification differences between systems.

The Working Group found that resistance principles, while frequently employed in the systems reviewed, were poorly documented and inconsistent among agencies. Therefore, the Working Group recommends that:

4. Agencies that apply resistance principles in updating industry classifications should collaborate to develop uniform guidelines for application of these principles. The rules used for resistance coding should be documented and made readily available.

E. Documentation

A major accomplishment of the Working Group has been the collection of detailed documentation on the characteristics of industry coding systems and source documents used for SIC coding. System descriptions developed by members of the Working Group with the help of Other agency personnel include information about: the basic coding unit, the industry classification principles followed, the source document used, the coding procedures, the volume and timing of coding, the quality measures associated with the coding, the general characteristics of the file in which the codes reside, the timing and methods for updating codes, planned changes to the coding system, and the uses and users of the industry codes. (A collection of these

systems descriptions is available as a supplement to this report

(Internal Revenue Service, 1984).)

This information serves as an essential tool for understanding the content of each system and the data produced from it. Therefore, the Working Group recommends that:

1. Complete documentation for coding systems included in this study should be updated at least every five years. Additionally, major changes occurring in any agency system should be documented and the information updated promptly.
2. All coding principles used by an agency principle which is either in addition to or contrary to those currently in the SIC Manual should be clearly described in agency publications that provide data by industry.
3. Coding rules embedded in programs for computerized

coding systems should be fully documented in a form that makes them accessible to data users.

4. Results of quality control checks and evaluation studies of manual and computerized coding operations should be systematically documented and made available to users.

The Working Group believes that agencies should adhere to certain standards for internal documentation. For example, cumulative files that contain industry codes should show the date of the most recent review and update for each unit and, where relevant, the source. in some cases it may be desirable to show more than one source code to avoid unnecessary restrictions on access. An agency may have data of its own and from other agencies, with differing restrictions on access. All data

sources should be identified to avoid unnecessary restrictions on

release of codes to other agencies for statistical purposes.

P. Hatching Studies and Other Research

Chapter IV documents several matching studies. Generally, the findings of such studies have led to improved methodology within the matched systems, greater awareness of the need for interagency cooperation, and a better understanding of the impact of differences in economic data used for policy determinations. In addition, matching studies provide information on the feasibility of code sharing and supporting evidence for the importance of code sharing. Most major matching studies were conducted more than 10 years ago.

The Working Group recommends that:

1. interagency microdata matching studies be conducted as a way of investigating the feasibility of code sharing and of quantifying differences between the systems.

Matching studies should compare industry codes, along with selected data items such as employment, geographic location, and

payroll, for units which match between agency files. The Working Group suggests that the studies first establish a sound matching process in areas with a high degree of agreement and comparability. Using matching processes identified as successful, a study should then focus on areas where classification is known to be especially difficult, such as wholesale and retail trade. Once differences are quantified, the agency specific procedures that cause the differences should be identified and improved.

A current interagency group, the Employer Reporting Unit Match Study (ERUMS) Working Group, has done initial planning for a micro-record matching study to compare the statistical characteristics of the Social Security, BLS, and IRS systems. The ERUMS Working Group will examine the effects of the variations between agencies in defining the reporting unit. Currently, expectations are that a sample covering 400 employer identification (EI) numbers from one state will be selected from Unemployment insurance (UI) records. ADP and manual matching techniques will be used to match these units with those in SSA and IRS for the same EI's. A natural by-product of the study will be a comparison of the industry codes for matched units. The ERUMS Working Group expects to gain useful information about the

kinds of problems that must be solved to match records from different economic data systems.

While documenting facets of the various industry coding systems, the Working Group made no attempt to judge the relative merits of any specific form, procedure, unit identification or updating method. All of the source documents and procedures used

9

by these cooperating agencies lend themselves to research studies aimed at identifying benefits and limitations. Chapters III and IV of this paper discuss in some detail specific forms, procedures, levels of industry coding, frequency of updating information used to obtain codes, and other details of each system. Based upon the review of these source documents, the Working Group recommends that:

2. Research studies and tests be conducted with a view toward establishing the most effective source documents for SIC coding as standards.

3. Tests and research be conducted on current and new methods and procedures for industry coding.

Tests and studies with varying sets of questions designed to elicit the nature of business activity should be cooperative ventures among agencies. Results of tests should be used to establish the most effective version as a standard. Since not all agencies can collect detailed information for use in industrial classification, the goal should be to develop standard questionnaires with at least two levels of detail.

A research project testing the verification method of SIC updating has been initiated by BLS (Hostetter, 1983). This method utilizes a form containing a description of the four-digit SIC industry in which a particular employer was most recently classified. The form requests the employer to verify the industry description as an accurate indicator of his primary economic activity. If correct, the employer simply checks the appropriate box, answers some other questions on ownership, auxiliary status and multi-establishment

status and returns the form. This reduces both respondent burden and staff time, since forms checked as correct need not be reviewed to assign an industry code. If the industry description does not correctly describe the economic activity, the employer then is asked to provide a detailed product and activity statement so that the correct classification can be determined. Currently, BLS has contracted with five State employment security agencies to conduct independent but identical quality measurement surveys testing the validity of the verification method of refiling.

The Census Bureau has introduced computer-assisted coding and is currently researching and refining the process. Although computer-assisted coding and updating codes by verification both have potential for enhancing SIC coding, the Working Group does not endorse wide use of either method until testing and results substantiate their effectiveness.

Additional cooperation among agencies on methodological research would allow progress toward standardization of all facets of industry coding. Even where standardization is not

possible, such research could produce detailed documentation of differences in data stemming from specific methods or procedures. This should prove useful to users who combine or compare data from different sources.

G. Interagency Cooperation

Increased interagency cooperation is essential for significant progress toward the goals stated at the beginning of this section: improvements in the quality, comparability and efficiency of industry coding systems.

The OMB Statistical Policy Office's Technical Committee on Industrial Classification is devoting most of its attention to planning for the SIC revision scheduled for 1987, with somewhat less attention to the other important aspects of industry classification and coding. The Working Group recommends that:

The activities relating to industrial classification and coding listed below should be undertaken either by the OMB Technical Committee on Industrial Classification or by another permanent interagency committee established for this purpose:

1. Regular meetings to discuss and resolve coding problems caused by the development of new industries and changes in the structure of existing industries. Interim solutions, pending revision of the SIC, should be agreed on and adopted by all of the participating agencies.
2. Promotion, support and coordination of other relevant activities along the lines recommended elsewhere in this chapter.

Some examples of how this continuing committee might operate include: periodic updating of the industry coding system descriptions prepared by the Industry Coding Working Group; conducting interagency workshops for sharing information about new coding methods and procedures and about materials and methods used to train coders;

promoting greater uniformity in source documents used for SIC coding;
coordinating and facilitating interagency matching studies; developing
standards for partial coding and for grouping 4-digit industries; and
developing standards for resistance coding..

In addition to leadership from the Statistical Policy office of
OMB and any interagency groups established for these purposes,,
progress on these recommendations will require full cooperation from
agencies that produce and use data classified by industry, as well as
those that control administrative record sources. from which industry
codes are developed.

CHAPTER II

DESCRIPTION OF THE INDUSTRY CODING WORKING GROUP PROJECT

A. Introduction

Under the auspices of the Administrative Records Subcommittee of the Federal Committee on Statistical Methodology, the Industry Coding Working Group reviewed industry coding systems used by Federal agencies to classify establishments and other economic units for statistical purposes. The objective of this interagency working group) was to review and document the existing industry coding systems with a view toward ultimately improving the comparability and quality of data classified by industry. This report describes the activity of the Working, Group and presents some findings and recommendations.

By industry coding systems here we mean the methods and procedures for assigning industry codes, rather than the technical aspects of constructing a classification framework and numbering scheme within which economic units will be assigned industry codes. Moreover, the term "industry code" is used in a generic sense; it refers to the codes actually used in each system, which are not always equivalent to the four-digit industry codes in the Standard Industrial Classification (Office of Management and Budget, 1972). The coding systems reviewed generally conform to

the SIC, but all are at variance with it to some degree.

The Working Group's effort was responsive to two recommendations made by a predecessor group, the Subcommittee on Statistical Uses of Administrative Records, which also worked under the auspices of the Federal Committee on Statistical Methodology. In its final report (office of Federal Statistical Policy and Standards, 1980), that Subcommittee recommended that:

The quality of administrative records to be used for statistical purposes should be evaluated systematically to determine the appropriateness of the records for the proposed use.

Consistent procedures should be used in administrative and statistical data collection efforts for defining reporting units, identifying and coding reporting unit characteristics, and developing standards for data tabulation.

These recommendations apply with particular force to industry classification and coding, where the information sources are many and of varying quality.

In order to get some idea of the magnitude of the industry code assignment by the Federal government, consider the following.

Annually, the Internal Revenue Service (IRS) assigns industry codes to nearly 16 million business units as part of its revenue processing of the tax returns. Additionally, more than 200,000 units are coded for the IRS Statistics of Income Program. Similarly, the Social Security Administration (SSA) assigns industry codes to over 900,000 new business units each year, with most of these (an estimated 875,000) coded in the Single-unit Employer Identification (EI) File coding operation.

As part of the Employment Security Program, the Bureau of Labor Statistics (BLS) maintains an industry-coded file of about 4.8 million units. Each year about 500,000 new units are coded, and codes are reviewed annually and updated, where appropriate, for about one-third of the existing units.

At the Census Bureau, as part of the annual Company organization Survey, over 900,000 establishments of multi-unit firms have their codes reviewed, and changed if appropriate, while about 75,000 new multi-unit establishments are industry coded. In addition to this, about 50,000 new business births are coded each year. For the quinquennial economic censuses, the Bureau mails census forms covering about half of the total universe of 6.7 million establishments in scope to the censuses. Responses to items included on the census forms are used to assign current industry codes to these establishments. Also, as part of the censuses, another 200,000 or so unclassified establishments are coded via a classification form mailing.

The figures just cited account for a substantial percentage of the volume of industry coding done by, or under the auspices of, the Federal government. However, this is not the whole picture, as can be seen from Table 1 on page 23, where coding volume figures (from columns (9), (10), and (11)) are given along with other data.

No attempt has been made in this work to quantify the substantial

costs associated with industry code assignment. This would indeed be difficult, since the industry coding is a necessary (and in many instances a relatively small) component of the overall administrative or statistical work which is being done concurrently.

Inconsistent industry classification of identical or overlapping populations of economic units by different agencies has led to problems of comparability for analysts and other users who try to compare and combine data from different agency sources. One example of this is in the area of productivity measurement. A recent report on this subject (National Research Council, 1979) said that "A major problem with the comparability of the basic data has been that different agencies assign the same establishments to different industry classifications, as a consequence,, aggregated data at the industry level are not in fact comparable

from agency to agency" (p. 178). Similar problems occur in connection

with the preparation of the national income and product accounts, in manpower studies, in the development of a data base for small businesses, and in other uses of economic statistics.

Several review groups have examined these problems (for example, the Central Statistical Board, 1939; the Hoover Commission, 1949; the President's Commission on Federal Statistics, 1971; the National Research Council, 1979; and the General Accounting office, 1979).

Without exception, they have recommended creation of a central listing of establishments and other economic units, classified by industry, which would be available to Federal and possibly State agencies for statistical purposes. The Census Bureau's Standard Statistical Establishment List (SSEL) was in fact developed for this purpose, but existing statutory restriction-- , on the release of Census Bureau information have so far made it impossible for other agencies to use the SSEL, except in a very limited sense.

At the technical level, several studies of relationships between reporting unit definitions and industry coding practices in different agency systems were undertaken by interagency working groups, under

the general direction of the Office of Statistical Standards of the Bureau of the Budget, in the early 1950's. Several of these studies which were begun in an attempt to account for observed discrepancies between manufacturing employment totals from the 1947 Census of manufactures and the BLS's Current Employment Statistics, involved matching individual reports for selected companies and establishments. These studies identified numerous problems that often impaired uniform reporting, many of which were solved by the working groups or referred to the Office of Statistical Standards SIC Technical Committee for action. The work during this period showed that significant progress toward comparability could result from carefully conducted studies of the coding principles and procedures used by different agencies and their application to particular units (Bureau of the Budget, 1961).

Since that time, however, there does not seem to have been any comprehensive and detailed technical review of the existing industry coding systems: their coverage, the classification principles followed, the coding procedures, and the uses of the industry codes assigned and of aggregate data classified by these codes.

The findings from the present review, the Working Group believes, will suggest changes in individual systems that can lead to significant improvements in quality and to greater comparability between systems. Also, these findings suggest advantages from new code sharing arrangements where these are permitted by law. Some gains can be realized even if there are no new exchanges of codes between agencies (for exchanges at present, see Table 3 on page 51). For example, the applicability

14

of shared software for computer assisted coding could be evaluated. Should future legislation permit the establishment and general use of a central list for statistical purposes, the Working Group's findings, suitably updated, should assist the implementation process.

B. Scope of the Review

The following 16 coding systems have been included in the

Working Group's review:

1. Bureau of Economic Analysis (BEA) System

-- Direct Investment Statistics

2. Bureau of Labor Statistics (BLS) System

-- Employment and Wages Program (ES-202 Report)

3. Bureau of the Census Systems

-- Agriculture Census

-- Business Births

-- Company Organization Survey

-- County Business Patterns

-- Economic Censuses

4. Federal Trade Commission (FTC) System

-- Quarterly Financial Report 1/

5. Internal Revenue Service (IRS) Statistics of Income (SOT)

Systems

-- Sole Proprietorships

-- Partnerships

-- Corporations

6. Internal Revenue Service (IRS) Administrative Systems

(Revenue Processing)

-- Sole Proprietorships

-- Partnerships

-- Corporations

1/ Responsibility for publishing the Quarterly Financial Report

was transferred to the Census Bureau in late, 1982. However,

throughout this paper all references to the FTC system or

Quarterly Financial Report apply to the time period before the transfer.

15

7. Social Security Administration (SSA) System::

- Single-unit Employer identification (EI) File
- Multi-unit EI File

The systems selected for review include some used only for statistical purposes (e.g., all Census systems) and some that are used for both statistical and non-statistical purposes (e.g., the IRS revenue processing systems). All of the systems assign codes to establishments or other economic units; systems that assign industry codes directly to individual workers were not included. most of the systems reviewed have broad coverage in terms of Standard industrial Classification (SIC) divisions; however, there are some exceptions, such as the Agriculture Census system. All are of a more or less

permanent character, i.e., the universe or a sample of it is coded periodically, or the coding is continuous in support of accretions or changes to a cumulative file. Most systems have a relatively large volume of coding, and together they are believed to account for a substantial proportion of the industry coding of establishments and other business units that is done by the Federal government and by State agencies under Federal-State cooperative programs.

It was necessary to distinguish between an industry coding system and the principal file in which the codes reside. To illustrate this, generally, industry codes assigned to establishments by the Census Bureau are placed in the Standard Statistical Establishment List (SSEL). (Industry codes assigned to agriculture establishments during the agriculture census processing are not placed in the SSEL, while those assigned to agricultural services establishments are.) However, the separate industry coding activities done at various times and based upon different source documents are treated as separate industry coding systems.

C. Major Uses of Industry Coding Information

The statistical uses of administrative records are well Documented in Statistical Policy Working Paper 6 (Office of Federal Statistical Policy and Standards, 1980). These uses range widely from the basic publication of statistics describing economic or demographic phenomena to being used as components in the formulation of complex mathematical models.

In general, industrial classification was developed for classifying an establishment by the activity in which it is primarily engaged. The presence of industry codes can facilitate the collection, tabulation, presentation and analysis of data as well as promote uniformity and comparability of data series.

The Federal Government uses industry codes as a means of aggregating much of the administrative and statistical data it collects for publication. Some examples of the regular publication of descriptive statistics by industry from primary data sources include:

- Quarterly Financial Report _for Manufacturing, _Mining and Trade Corporations by the Federal Trade Commission (FTC).

- Corporation Income Tax Returns, Sole Proprietorship) Returns, and Partnership Returns by the Internal Revenue Service (IRS).

- Census Bureau publications such as County Business Patterns and the results of the economic censuses.

- Employment and Earnings and Employment and Wages by the Bureau of Labor Statistics (BLS).

There are other data series published that have been synthesized from several primary data sources. The Bureau of Economic Analysis (BEA) , for the most part, does not collect information directly from firms or individuals. BEA's estimates of current economic activity are based on data obtained from other agencies. The Gross national

Product, which is presented with industry detail, combines data from many sources including the Census Bureau, IRS, BLS, and FTC. The Input-Output Accounts of the U.S. are composed entirely of industry information collected by others. BEA's estimates of State and local area personal income involve the use of several sets of data aggregated by industry. BEA is thus heavily dependent on the comparability of data from its various sources.

In addition, both published and unpublished sets of industry - based data are useful for the collecting agency's internal programs. For example, various units of the Department of Labor use BLS data for purposes such as:

- studies of financial aspects of the Unemployment Insurance program are conducted to set maximum weekly benefit levels.

- States use industry wage and employment data in preparing forecasts of program workloads that are used in developing annual budgets.

- Local area workforce and unemployment statistics are produced by industry which enables classification of areas eligible for benefits under a number of Federal area assistance programs.

- Employment figures are useful in time-series analysis and in the study of seasonal employment, and are used extensively in industry/area comparisons.

1/ Responsibility for publishing the Quarterly Financial Report was transferred to the Census Bureau in late 1982.

- The data serve as a base for labor market information programs at the county, labor market area, State and national levels.

Industry codes from some administrative or statistical record systems are helpful in the processing and tabulation of raw data in other record systems. The Social Security Administration (SSA) assigns industry codes to new firms applying for an employer identification number. A major use of these codes is for identifying industrial activity for workers included in the Continuous Work History Sample (CWHS). These codes are also released to the Census Bureau for incorporation into their standard Statistical establishment List. Reciprocally, on some past occasions, the Census Bureau has provided SSA with updates of industry codes for employers based on the results of the economic censuses.

Some data producers can use the industry codes from other systems as a tool to edit aggregated tabulations. BEA, for example, receives industry codes from FTC and IRS for individual corporations which help to explain changes in their estimates of components in the National income and Product Accounts.

There are other uses that governmental units make of the industry information that they can obtain from data producing

agencies. The IRS, for instance, releases its industry coded Statistics of Income (SOI) files to the office of Tax Analysis and to the Joint Committee on Taxation for use in "tax models" to evaluate the effects of existing or proposed tax policies.

Nongovernment groups such as businesses and nonprofit organizations use industry information from administrative and statistical sources as well. While confidentiality restrictions prohibit the transfer of individual industry codes outside the government (except to contractors of government agencies), aggregated statistics based on industry can be quite useful. Business firms can conduct research to classify and study the industrial profiles of their customers and suppliers. Sales patterns can be analyzed, market potentials can be estimated and commercial strategies can be evaluated.

The industry dimension of administrative and statistical data is one of their most interesting and useful characteristics. It enables the government to improve and evaluate many of its programs. It enhances the research efforts of both public and private groups and it is very helpful to individuals in gaining understanding of the

economic and demographic characteristics of the nation.

D. Composition and objectives of the Industry Coding Working Group

The Working Group members (see list in preface) were in some

cases members of the parent subcommittee or were designated by

18

the subcommittee representative or their agency. Working Group

members met for the first time in May of 1981 and have conducted

meetings, generally monthly, throughout 1982 and 1983.

From the outset the Working Group felt that a fundamental task was to review and document the major industry coding systems. Once this was accomplished, analysis and comparison followed, leading to the proposals for improvements in the comparability and quality of the industry codes which appear in Chapter I . As a further application of this work, a user or potential user of data classified by industry can be provided with essential information concerning the usability

and relative quality of the data.

E. Development of the Basic Documentation for the Federal Industry

Coding Systems

The Working Group constructed a questionnaire on industry coding which requested basic information needed to compare and assess the systems. This questionnaire covered the following main areas:

- The basic coding unit (the unit to which an industry code is assigned), the source or source document from which the coding is done, and the industry classification system used;
- The volume, timing, coding procedures, resource material used, and quality measures associated with the coding;
- General characteristics of the principal file(s) in which the codes reside;
- Updating of the codes and recent or planned changes to the

coding system;

- The uses and users of the industry codes.

Within each of these areas specific questions were asked. Also, related documentation was requested, principally the forms or source documents from which the coding is done, code lists and instructions concerning classification system variations, and any available data bearing on the quality of the coding.

Members of the Working Group identified industry coding systems within their own agencies which fit into the scope of the review. At the same time, they identified key persons who were most knowledgeable about each coding system. The survey questionnaires were then delivered to these respondents by the Working Group members.

Each completed questionnaire was reviewed by one or more members of the Working Group and a meeting was arranged with the respondent for clarification or further information. As a result

of the meeting, the questionnaire was revised, and frequently additional documentation of the system was obtained.

A summary system description was prepared from each questionnaire and the associated materials. These descriptions are designed to put the collected information in a standardized, concise format for easy reference, comparison, and analysis. These summary descriptions form the basis of this report. Copies of system descriptions may be obtained by contacting the Statistics of Income Division, Internal Revenue Service.

CHAPTER III

INDUSTRY CODING SYSTEMS AND THEIR RELATIONSHIPS

A. Introduction

This chapter provides an analysis of the coding systems reviewed. This analysis should provide a stimulus to the agencies maintaining the systems to make changes aimed at increasing comparability with other systems and at improving the accuracy of codes and reducing the cost of coding in their own systems. In addition, the information developed can make possible a technical evaluation of possible new arrangements for interagency code sharing, subject to legal restrictions on such exchanges. Finally, the results should help users of data from these systems to understand their structure and limitations and the extent to which data from different systems are comparable.

An initial step is to identify the system characteristics or dimensions to be compared. The primary dimensions that have been identified are coverage, frequency and timing of initial coding and updating, classification system used, classification principles, information used as input to coding, coding procedures, and description of systems relationships.

Each of these dimensions is discussed in the following sections.

B. Coverage

Systems coverage has 3 sub-dimensions which can be described by the answers to 3 questions: What kinds of units are coded? Which of these units are included in the target population? And, finally, is coding for all units or for a sample?

1. Kinds of Units Coded

The kinds of units that are classified by industry vary widely. The Standard Industrial Classification (SIC) was developed for classification of establishments by industry. Its offshoot, the EnterPrise Standard Industrial Classification (ESIC), was developed for classification by industry of enterprises or companies, many of which consist of two or more establishments (Office of Management and Budget, 1972, 1974, and Office of Federal Statistical Policy and Standards, 1977b.)

Concerning this first aspect of coverage, basic coding units or simply units, i.e., the units of observation to which industry codes are applied, are often determined by intended uses of the data files. For example, the Census Bureau systems, which are established and maintained solely for statistical purposes, use establishments as the basic unit. However, the Standard Statistical Establishment List (SSEL) which is the

22

basic file in which industry codes produced by the various Census Bureau systems reside, is organized to permit the aggregation of groups of establishments to form other units, such as Employer Identification (EI) number units (all establishments operating under a single EI number) and enterprises, and the assignment of industry codes to these units.

By contrast, the units used in the systems of other agencies (e.g., employers, tax entities, consolidated corporations) are determined largely by administrative requirements. Table I on page 23

provides a comparison of the basic coding units used for each system studied, as well as comparisons of SIC level of, detail used, sample or population coverage, an assessment of the level of input data available for assignment of codes, updating cycles, and the average annual volume of coding.

In practice, business enterprises consisting of a single establishment, as defined for purposes of the SIC, are classified in essentially the same way in all of the systems reviewed by the Working Group. There are, to be sure, some elements of judgment in the SIC definition, especially in those instances where "...distinct and separate economic activities are performed at a single physical location..." (Office of Management and Budget, 1972, p.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 separately for each activity on employment, payrolls, sales or receipts and other establishment type data. These criteria clearly allow some latitude for judgment by the agency collecting the data, and one could expect to find some cases where establishments were defined differently by different agencies.

Nevertheless, the major conceptual differences among 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, administrative, or statistical basis, such as employers, taxpayers, corporations, consolidate, corporations, or "reporting units".

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 operational definitions are not identical for .multi-establishment employers. Basically, the reporting unit in each case is a group of two or more establishments under the same employer (El number) in the same county and four-digit industry. It has been so established for the convenience of employers who would find it difficult or burdensome to file separate administrative returns to SSA and to State Employment Security Agencies for each establishment.

The BLS system is primarily an establishment based

system. However, under certain circumstances a "reporting unit" concept is substituted. The "reporting unit" used by BLS

25

includes two or more establishments under the same employer identification (EI) or Unemployment Insurance (UI) account number in the same county and industry. These exceptions to establishment based reporting are allowed in order to reduce employer quarterly unemployment insurance tax reporting burden. Exceptions to county/industry level reporting are discouraged.

SSA also uses a "reporting unit" concept under their establishment Reporting Plan (ERP) to facilitate the processing of large multi-unit employer wage reports. When an employer firm agrees to participate in the plan, it is asked to identify each of the firm's reporting units (which may be establishments or payroll groupings) by geographic location (county) and industrial activity and assign a four-digit reporting unit number to each on a Form SSA-5019. on subsequent annual wage reports the firm groups its employees by

reporting unit, identifying each with the preassigned unit number.

This arrangement provides a basis for SSA to isolate earning discrepancies and to assign geographic and industrial classification to each unit so that wage reports can be used as a source of statistical data. However, it should be noted that due to the voluntary nature of ERP, every effort is made to set up and maintain a breakdown of reporting units that most closely conforms to the firm's internal business structure in order to minimize the reporting burden on the employer. This may or may not result in the use of establishments as the reporting unit. In summary, operational, procedural, and definitional differences make it difficult to compare the net effect of the use of the "reporting unit" concepts in the BLS and SSA systems.

Finally, it is worthwhile to point out that for all systems the nature of the units which are classified by industry in each system is affected not only by the formal definitions but also by the specific procedures used to implement these definitions.

2. Units Included in the Target Population

The second aspect of coverage is to identify which of the specified units are included in the target population for the system.

The 5 principal criteria are:

a. Geographic location. All systems cover units located in the United States and owned by United States citizens or legal entities. Treatment varies for units located in United States territories and possessions, for units with non-United States ownership physically located in the United States, and United States-owned units located outside of the United States.

b. Legal form of organization. Each of the IRS systems covers only one form of organization: sole proprietorship, partnership or corporation. The FTC Quarterly Financial Report system covers only corporations. Most systems cover all forms of organization. However, coverage of government-operated units differs greatly, as described in d. below.

C. Presence of employees. Sole proprietorships or partnerships with no employees are included in the IRS systems if they are required to file tax returns. These nonemployer establishments are incorporated into the economic censuses from IRS records; they are not independently contacted by the Census Bureau. Also, establishments without payroll are included in the Census of Agriculture. All other coding systems code only units with employees.

d. SIC divisions. Some systems are restricted to specified SIC divisions or parts of divisions. For example, the Census of Agriculture covers only part of Division A (Agriculture, Forestry, and Fishing). The FTC Quarterly Financial Report system covers only corporations whose primary activity is in mining, manufacturing, wholesale trade and retail trade. The inclusion of government units varies. They are not covered at all by IRS systems, but are covered in part by several other systems. The BLS Employment and Wages system covers government employees at all levels, except for members of the armed forces.

e. Size. Industry coding in the economic censuses is limited to employer establishments which exceed payroll cutoffs that vary by industry. These cutoffs are set to exclude the smallest establishments within an industry from getting a census form. The census data, including industry codes for these small Establishments, are taken from administrative records. In the Census of Construction, however, census forms are mailed to a probability sample of establishments below the established cutoffs, and sample estimates for this group are included in the census totals.

Table 2 on page 27 shows the coverage of the systems reviewed with respect to criteria b., c., and d. For this purpose, the six IRS systems were grouped to form two "mega-systems": the Revenue Processing and the Statistics of Income systems.

3. Coding for a Sample or a Population

The third aspect of coverage is whether or not sampling is used. If it is, the particular sample design will affect the frequency with which coding is required and the potential for sharing industry codes with other systems. Examples of sample based systems are the IRS

Statistics of Income systems, the FTC Quarterly Financial Report system, and the Census Bureaus Business Births coding system.

Of all systems reviewed, the IRS systems (condensed in Table 2 from six to two systems) are the most complete, covering all SIC divisions except J, Public Administration, and all forms of organization except "government establishments" in the other SIC divisions.

The most complete coverage of Division J, Public Administration, is by the BLS Employment and Wages System, since most public as well as private employers are covered by the Unemployment Insurance system.

It should be-noted that the 1972 revision of the SIC changed the principles for 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 Division J, Public Administration. One result of this change is that the TRS 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.

For employers, i.e., businesses with one or more paid employees, the BLS Employment and Wages and the SSA single-unit EI systems 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 tax-exempt nonprofit organizations.

C. Frequency and Timing of Initial Coding and Updating

The extremes of this dimension can be represented by the IRS revenue processing coding systems and the SSA single-unit EI System. In the IRS revenue processing systems, industry codes are assigned annually to businesses reported on tax returns, without reference to prior year codes. In the SSA system, each covered employer is assigned an industry code at the time of entry into the system, which

occurs when the employer applies for an EI number. This code is generally retained in the system unless and until updated, primarily by matching against economic censuses codes for the employers in the file. These two approaches can be distinguished by the labels "periodic, independent" for the approach represented by the IRS systems and cumulative" for the approach represented by the SSA single-unit system. As another example, BLS has a tight schedule for new code assignments, along with a three year cycle for updating. Many systems lie somewhere in between the extremes. Where industry coding is done for a sample of units in the target population, the approach used will depend on whether and how much the samples for successive time periods overlap.

D. Classification System Used

All of the systems studied use a classification scheme based on the SIC. Some systems which classify groups of establishments, e.g., the IRS systems for corporations, use systems based on the ESIC, which in turn ties into the SIC.

For the systems reviewed by the Industry Coding Working

Group, the following assertion can be made: While each

29

classification system is based on the 1972 SIC 1/ 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:

- grouping of SIC categories
- subdivision of four-digit SIC categories
- addition of categories not covered by the SIC

For the systems reviewed, grouping of SIC categories is more common than subdivision.

The SIC contains 1,005 four-digit and 421 three-digit codes.

The systems of IRS use a much smaller number of categories than the others, currently in the neighborhood of 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 groups 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 (1) hotels, and (2) motels, motor hotels, and tourist courts.

The BLS system uses most (971 of the 1,005) four-digit industry codes. In the 34 remaining industries, BLS experience is that four-digit sic level coding is often unreliable because of conditions that prevail in these industries, such as frequent fluctuations in employer products or services or generally inadequate employer records.

The SSA system also uses most of the four-digit industry codes. 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 i (public administration) , where only the two-digit detail is provided. The

codes used for these groups are called "foldback" codes. Thus, there are 63 of the 1,005 SIC industry codes which are not used at all. For 115 industries, "foldback codes" are used only if the employer does not furnish enough information to code to the four-digit level; followups for additional information are not attempted by SSA. The use of these foldback codes was especially heavy during a period in the early 1970's when SSA was doing "dual coding" (assigning two codes to each employer, one based on the 1967 SI. and one based on the 1972 SIC) in preparation for conversion of their systems to the 1972 SIC. In summary, it seems fair to say that full SIC detail is lacking in SSA's systems for 178 of the 1005 industries in the 1972 SIC.

1/ AS revised by the 1977 Supplement (Office of Federal Statistical Policy and Standards, 1977b).

The Census Bureau's industry classification system for the 1977 Economic Censuses is described in its 1977 Industry and Product

Classification Manual (Bureau of the Census, 1977b). The latest version of this IPC manual for the 1982 Economic Censuses has recently been released. Census 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.

Industries affected by the latter restriction for 1977 are:

- (1) Mercury, 1092, grouped with 1099
- (2) Typewriters, 3572, grouped with 3579
- (3) Electronic tubes, 3671 to 3673, carried as 3671.

In addition, for economic censuses 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 "sub-industries" are identified by adding two digits to the four digit SIC code. For the 1977 Economic Censuses, 83 four-digit industries in these major groups were subdivided to form 256 six-digit sub-industries. 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 5th digit, and the 6th 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 6th position.

All of the systems have conformed to SIC revisions; in addition, many of them have introduced other changes from time to time, usually in the direction of showing more detail.

E. Classification Principles

Given the general principle of adherence to the SIC, there remain several conceptual issues to be dealt with in order to develop the procedures to classify establishments or other units by industry (Simmons, 1953). These include:

1. Classification of units with multiple activities.

Under some conditions, such units may be split and classified separately. This option is more likely to be used when reports are

filed solely for statistical purposes. When it is not used the first decision needed is what measure of activity to use. Options include gross receipts, value of sales, value of production, value of shipments, and employment or payroll associated with each activity covered by a separate SIC code. A second decision is how to use these measures to determine the principal activity. one option is to simply choose the 4-digit (or 6-digit if using IPC) category with the highest value of the measure chosen. An alternative sometimes used is a hierarchical

procedure: choose first the SIC division which has the highest value, next the major (2-digit) industry within that division with the highest value, and so on until the 4-digit or 6-digit level is reached.

For establishments the main question is what measure of the relative importance of different activities should be user? The 1972 SIC Manual (Office of Management and Budget, 1972) is F clear on this.

It states that "Ideally, the principal product or service should be determined by its relative share of "value added" at the establishment" (p. 12). 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. p. 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, communications,	Value of receipts or

electric, gas, and sanitary services	revenues
Wholesale trade	Value of sales
Retail trade	Value of sales
Finance, insurance, and real estate	Value of receipts
Services (including agricultural services) or revenues	Value of receipts
Public administration	Employment or-payroll

The recommendation is qualified in two ways. First, it is stated that these measures should be used "when available." Second, it is stated that 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 by four-digit industry, the establishment is coded to the industry with the largest share

32

of the total, without regard to the shares of higher-level SIC categories (industry groups, major industries, or divisions).

To what extent are these recommendations followed in the systems reviewed by the Industry Coding Working Group? Following is a summary of the practices of the four major agencies. It

will be seen that none of the agencies follows the SIC Manual in every respect.

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 -- 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, where value of receipts is used in place of value of production and Division D, manufacturing, where 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 codes and/or 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 (Schedule r) for farm sole proprietors

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 to base principal activity on sales or receipts was used.

SSA -- Currently employers applying for an EI number are asked to describe their "nature of principal business activity" without any specific reference to the treatment of multiple activities. Multi-unit employers who provide data for their separate establishments or reporting units are asked to provide percentages corresponding to the principal activities of each one, listed in order of importance, but the instructions do not say on what measures these percentages should be based. The report form also asks for 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 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.

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

2. Time interval and reference period

One year is the standard time interval for most systems. The SSA

systems are an exception; the input document asks for a description of the principal activity carried on, without any reference to a specific time period. Most systems use a calendar year, but in some systems the reports are for tax years or fiscal years, which are not equivalent to calendar years for all units coded.

Another important consideration is the relationship between the reference period for code determination and the period for which data are collected and the code assigned. This leads to the question of updating, i.e., how often should industry codes be revised? There is considerable variation both between and within systems as to the frequency of updating industry codes, or refiling, as it is sometimes called.

When a system is used to produce aggregate data such as employment, payroll, receipts, etc., classified by industry, the reference period on which the industry code is based may not be the same as the period covered by the data. The major industry coding systems reviewed do, in fact, differ considerably in this respect. Following is a broad outline of the differing practices followed by

each of the four major industry coding agencies.

IRS -- Returns are industry coded annually, based either on 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 resorting unit is classified initially when the employer enters the unemployment insurance system. It is BLS policy that codes should be reviewed and updated on a fixed time schedule, as follows:

34

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

The timing of the 3-year cycles varies by SIC division, so that review and updating is done for 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 -- Each employer is classified initially at the time an application for an EI number is filed. The application form asks for information about the nature of the business at the time of the filing; there is no defined reference period. Shortly thereafter, eligible multi-unit employers are asked to submit activity information for each of their reporting units, the situation with respect to reference period being the same as for the original application form. For single-unit employers, the last general update was based on a

comparison with codes assigned in the 1972 Economic Censuses. For multi-unit 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. Resources for such correspondence are limited.

Since both the single and multi-unit employer files carry date codes indicating the most recent update of the 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 -- Reference periods vary by coding systems. For units covered by mail (or interview) in economic censuses, the industry classification has the same reference period as the data. This is also true in some but not all current surveys. Perhaps the best approach is to consider the SSEL, which provides the frame for all censuses and surveys and for the annual County Business Patterns program.¹ For the larger multi-unit companies, industry codes for their establishments are updated annually in the Company Organization Survey. Smaller multi-unit

companies are updated once between five-year economic censuses.

At the other end of the spectrum, industry codes for single-unit

1### - true for all units with employees. IRS is the main source of information for zero-employee units.

35

employers outside the industry 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 in 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 EI numbers.

In summary, most agencies use a one-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. (See Table 1 on page 23, Column B.)

3. other considerations

Some data users are troubled by the effects of sudden and/or erratic changes in industry classification, especially when large units are affected. This has led to the application, in some systems, of resistance principles. After a preliminary code has been determined using data from the current reference period, the preliminary code is compared with codes from one or more previous periods. If the preliminary code differs, from the prior one, it is accepted only if certain threshold conditions are met. Several of the systems studied incorporate resistance principles.

There is also the problem of the classification of certain ancillary or auxiliary activities, such as central administrative offices, manufacturers' sales branches, laboratories, and warehouses. Classification of these units is usually based on the activities of the establishments they serve, as specified by the SIC Manual.

F. Information Used as Input to Coding

Various sources of information are used as input for classification of units by industry within the agency systems covered in this study.

The two principal categories are agency source documents, and

information other than agency source documents. The latter

encompasses prior codes assigned within the same agency and codes from

other agencies. The referencing of codes and other information

available from commercial sources and contact with the company by

phone, correspondence, or in person are also methods of obtaining

additional coding information.

1. Agency Source Documents

The principal resource for assigning industry codes to units within each system is usually the source documents used by the agency.

The reason for this is that the codes from other agencies or

commercial business listings may not be fully compatible with the data

classification requirements of the

receiving system because of differences such as the required level of detail, coding principles, code inaccuracy and whether or not the codes apply to the appropriate reference period. Also, in many situations code transfers are prohibited by law.

A study of the source documents used for the different coding systems shows a variation between agencies and in some cases within agencies. Lack of standards in this area could be one reason, but the variation can, in most cases, be justified by the major differences between each agency program's data requirements for the design of their source documents, and whether industry coding is a primary or supplemental consideration in this program.

Some factors that an agency must consider in designing the form are the type of information needed in order to obtain the desired level of industry detail, the scope of instructions needed to secure this information, and whether or not the form can be specialized to

cover specific industries. It is also necessary to determine whether the forms are to be self coded by the respondent, manually coded by the agency's classifiers or coded by computer. In addition, the burden which completing the form places on the respondent must be evaluated.

A. very important factor that should be noted is that

often the coding source documents are designed primarily for

other purposes. For example, the Form SS-4, which is used as the

main coding source for SSA's single unit E1 coding system, is actually

an IRS form utilized by employers and others in applying for an E1

number. Another case would be the IRS' Statistics of Income coding

Systems where tax schedules, such as the Form 1120, are user for

industry coding. Coding information is often a minor part of such

forms.

In contrast, some other agency source documents are specifically

designed for the collection of industrial data. These forms may vary

from the general purpose type to report forms tailored to a specific

industry. Examples, of these latter types of source documents are the

various report forms used in the economic censuses. These forms are specialized to the industry which has been determined by codes assigned from previous censuses or surveys, the Company Organization Survey (COS) or Social Security Administration (SSA) records. If a code is not available and the kind of business cannot be determined from the trade name or other reliable information, a more generalized form is sent.

In general, the principal difference among the source documents is the nature and detail of coding information available on the various forms used in each agency's system(s). The type of information requested on these forms for determining an industry code ranges from brief descriptions of the principal business activity, or pre-listed industry descriptions and codes for self-selection, to percent distributions of gross sales or

receipts by products or services. Specific examples of these varied

kinds of information are: (1) pre-listed taxpayer-selected codes such as on IRS Form 1120; (2) pre-listed kind of business activity check boxes (with or without industry codes) on report forms used to classify establishments lacking industry codes prior to mailing industry-specific forms in the economic censuses; (3) respondent-furnished descriptions of principal products or activities based on percent of total sales on BLS Forms 3023-A and 3023-B (of which there are different versions for each industry division); (4) principal business activity on BLS Form SS-4 used in SSA's single-unit EI coding system; and (5) sales distribution by industry on BEA's Form BE-12 used in their Benchmark Surveys. In the absence of an adequate description of the unit's activities, some agency systems may use the trade name as a coding source (e.g., Hilda's Beauty Shop, Bob's Cafe or Johnson's Department Store). This "name coding" is used in SSA's coding of the Form SS-4.

The following is a comparative analysis of the level of detail available on source documents. It provides a comparison by level of source information detail based on the chart shown below and gives examples for each category (See Chapter VI for actual source documents and brief description of each).

		Level of source
Category	Coding by:	information detail
A	Respondent	Not applicable
B	Agency	Low
C	Agency	Medium
D	Agency	High

Category A (Selfcoded) -- The only systems which use self-coding (i.e., coding by respondents) almost exclusively are the IRS revenue processing systems for partnerships and corporations. Some forms used in BEA's Direct Investment (DI) Statistics Program also request respondents to enter up to eight 3-digit codes which represent DI industry Classifications under which they have sales. However, final code determinations are made and entered on the forms by BEA coders. Bureau of the Census forms, especially in the retail anti wholesale

trade and service areas, also frequently utilize pre-listed, respondents elected descriptions and codes. In most cases, responses to these items are checked against other data furnished on the form in order to determine what industry code to assign.

The source documents for the above mentioned IRS systems are the appropriate tax return forms for these 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 1, Chapter VI. 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 4nstructions. The code list provides a short description for each of the industries included by IRS along with the appropriate codes. 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 to a greater extent in the Statistics of income industry coding.

An observed feature of self-coding is the potential for a high proportion of incorrect codes immediately following a revision of the Standard Industrial Classification. Some evidence on this score is presented in Chapter IV.

Category B (Agency coded, low detail) -- The example for this category is also taken from IRS. Exhibit 2 of Chapter VI shows the relevant data items and instructions from the 1981 tax return schedule used for non farm 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 11 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 are used for some combinations of activities.

It may be noted in passing 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 crops and livestock production. This is probably sufficient to put these source documents in Category D.

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

Category C (Agency coded, medium detail) -- The main example for this category is the Form SS-4 (Application for Employer

Identification number). The complete Form SS-4 and the relevant section of the instructions for it appear as Exhibit 3 of Chapter II. This is an IRS form Used by SSA to classify all employers for the single-unit employer file. (Codes for

39

establishments or reporting units of multi-unit employers are based on a more detailed form which is sent to eligible employers following receipt of the initial application.) 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. other items which may assist in classification are:

Items 1 and 4 -- Name and 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 were not covered in the Category B example. The Form SS-4 is classified in the medium rather than high detail category primarily because it does not provide any breakdown of multiple activities. 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 will be useful to give more than one example.

The source documents which provide the most information for industry coding are the mail questionnaires used in the quinquennial economic censuses. These questionnaires call for detailed information and are tailored to different groups of SIC industries hence they include the specialized inquiries needed to assign industry codes within those groups. Special procedures are, of course, needed to handle questionnaires returned by establishments which are inappropriate to their activities.

Exhibit 4 of Chapter VI shows one 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 industry and Product Classification 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 Standard Statistical Establishment List (SSEL), will appear on the mailing label. A "self-designated" code will be determined on the basis of the respondent's entry in Item 9, Kind of Business. Normally, the final IPC code will

be computer-

40

assigned, 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 code determination only when the data for the items normally used are incomplete, ambiguous, or contradictory.

Other forms that provide a high level of information for industry coding are BLS Forms 3023-A (Industry Classification Statement) and 3023-B (Industry Verification Form), which are designed for each industry and used for updating all industry codes. They are also used to update area, type of ownership, and auxiliary codes of existing units covered by the Unemployment Insurance Employment and Wages (ES-202) Program on a three-year refiling cycle. Form BLS 3023-A is used sometimes by the state agencies to clarify or obtain additional

information necessary to assign SIC codes to new employer accounts.

For both forms, there are separate versions for each industrial division (including an "all industry" version). Each form also provides for the inclusion of other establishments reported by a multi-unit company.

Exhibit 5 of Chapter VI shows BLS Form 3023-A7 (Rev. Dec. 1982), which is one of the forms used to update industry codes for reporting units currently classified in wholesale trade. Unlike other examples discussed in this section, this form is designed primarily to get the information needed for classification of the report unit. The key items on the form for this purpose are items B, D and E. Item B covers the identification of multiple products or activities of the reporting unit, and the percent of total sales (value of receipts) accounted for by each during the most recent calendar year. Item D identifies Central Administrative offices (CAO'S) and auxiliary units, and item E asks for the principal class of customer, as an aid to determining whether the unit is wholesale or retail.

A final example in this category comes from the Federal Trade

Commission's (FTC) Quarterly Financial Report (QFR) Program. (This program was transferred to the Bureau of the Census in late 1982.)

Exhibit 6 of Chapter VI 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 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 which 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 the 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 used 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.

2. Information Other than Agency Source Documents

As stated earlier, most agencies rely primarily on their own source documents as input to their coding systems. However, in certain situations they may resort to other coding sources such as additional contact with the company, prior codes assigned to the same units within their own agency codes supplied by other agencies, and codes and other pertinent information extracted from commercial sources.

The prior codes assigned by an agency are used for various purposes. Listed below are some of the uses and examples of agency systems to which these situations apply.

--- Report form selection. During the economic censuses the

Census Bureau utilizes prior codes as a selection factor in

determining the appropriate form to be mailed.

- Reference for manual editing. Many of the agency coding systems reference prior codes during updating processes for purposes of reviewing code changes, determining accuracy of current codes and making final code determinations. For example, prior codes for permanent sample units in FTC's Quarterly Financial Report (QFR) are available to the coders for determining code changes for large corporations.

Codes supplied by other agencies are also used for various purposes. Some of these are listed below with examples.

- Report form selection. The Census Bureau uses industry codes from SSA records if no previous Census assigned codes are available to determine the appropriate report form to mail in the economic censuses.
- Coding of nonrespondents, and establishments not included in the mail part of the economic censuses. IRS Principal

Industrial Activity (PIA) and SSA assigned codes are two of the various sources used by the Census Bureau for determining an industry code for these cases in the economic censuses.

42

- Coding of units with incomplete data. The Census Bureau references SSA assigned codes when classifying cases with insufficient information in the business births coding system.

- Updating procedures. The Social Security Administration attempts to update its code files every five years through a coordination with census records based on codes resulting from the economic censuses (especially following a major SIC revision). The last such update was based on the 1972 Economic Censuses.

Other sources of coding information are commercial business

listings (e.g., Dun and Bradstreet, Moody's, Thomas Register). many agencies use these as a source when there is insufficient information to assign a complete industry code to a unit. Some examples of the different agency coding systems which utilize these references are:

- (1) business births coding (Census),
- (2) single-unit EI file (SSA),
- (3) Company organization Survey (Census),
- (4) economic censuses (Census),
- (5) Quarterly Financial Report (FTC), and
- (5) Statistics of Income -- Corporations (IRS).

The final coding source (and indeed the first and preferred source for large establishments and firms) by which an agency may obtain coding information for a unit when there is insufficient information is through additional contact with the company by phone, written correspondence, or in person. This is done for most of the systems and, as a case in point, for the Unemployment Insurance (UI) Employment and Wage Program (Bureau of Labor Statistics). Here the State may send a BLS-3023 form (for new accounts), contact the employer by phone or make a personal visit in order to obtain the needed information.

The wide variation among the coding sources used by the various agencies affects the uniformity of codes assigned to the same units in different systems. Greater standardization of the coding systems in this area would seem feasible at this time, but only for agencies which have similar data requirements and have the resources needed to code at the agreed level of detail.

G. Coding Procedures

The procedures developed for use within the different coding systems encompass a variety of activities. These include:

- The methods by which the industry codes are assigned (i.e., manual, computer-assisted, automated).

- Treatment of missing data.

- Data entry.

- Quality assurance procedures (i.e., manual quality control and computer consistency checks).

The following provides descriptions of procedure types available under each of these functions and examples of how they are used. It shows that wide variations exist between the procedures for the systems studied. The fact that these differences will affect the comparability of codes between agencies is self-evident.

1. Methods of assigning codes.

There are three principal methods by which the initial industry codes are assigned. Of these, manual coding is the most frequently used. The other methods used are "automated coding" and "computer-assisted coding," which is also a form of manual coding. At this time the Census Bureau is the only agency which makes use of "computer assisted coding." Listed below are basic descriptions of the procedures which apply to each of these methods:

-- Manual Coding. Under this method the classifier manually assigns an industry code directly to the source document (or other form used for data entry purposes) based on information supplied by the respondent and other available sources such as commercial references or prior codes.

-- Computer-assisted Coding. This system was developed by the Census Bureau to assist the coder during manual operations by computerizing the basic coding routine. This system is being used in several phases of the 1982 Economic Censuses processing.

Under this method, the coder, who is working at an interactive computer terminal, is first required to select the major SIC division which relates to the activity description and/or trade name supplied on the source form. Then the coder selects a "key word" based on the same information and enters it into the terminal. If possible, the system matches the "key word" to one or more verbal descriptions of SIC industries. These industry descriptions

are then displayed, with their associated code, for the coder to select the description and code which is applicable. If the coder is unable to assign a code at this point, the system will then direct the coder through several routines until a code is derived. If this fails the case is referred to an analyst for review.

In addition to its coding functions, this method was also developed to improve the training of

coders, increase consistency, and provide a flexible mechanism for continuous updating of descriptions and codes in the system and IpC Manual. it is also the first step towards a fully automated system of coding through the development of a comprehensive dictionary of industry descriptions.

-- Computer/Automated Coding. Currently no coding system studied by the Working Group is fully automated; however, two agencies (Census Bureau and IRS) are using largely automated coding procedures. within the Census Bureau systems (e.g., the mail portion of the economic censuses, Census of Agriculture for farms with sales of \$2,500 or more and other periodic surveys such as the Annual Survey of Manufactures) which have implemented this method, this is done by using computerized data on receipts or sales by type of product or service to assign and place in the records for each unit an industry code, according to a programmed set of rules. Starting with tax year 1981, IRS's SOI programs have used largely automated procedures for generating current year SoI codes. Procedures vary by type of return and tax year. For most returns, the automated coding process derives the current year SOI code either from the prior year SOI code or from the current year revenue processing code. Manual coding is used only on an exception basis.

The following lists the agencies covered in the review and

describes the manner in which these methods of coding are applied within the various coding systems.

BEA -- An editor manually assigns the industry codes using the "top down method." The SIC Division is first determined by aggregating the sales distributions which are each assigned a three digit Direct Investment industry classification by the respondent. Then a more detailed industry code is assigned based on the subdivision of the industry division which has the largest percentage of sales. This coding procedure is used in coding source documents in Benchmark Surveys (Forms BE-10 and BE-12) and forms filed for new entities and major code changes (Forms BE-507 and BE-607).

BLS -- An initial industry code is manually assigned to each unit-First entering the Unemployment insurance Employment and Wages (ES-202) Program based on the principal business activity (as defined in the SIC Manual) submitted by the employer on an "employer status determination of liability form." Except for problem cases, which are individually handled by regional offices or at the national BLS office, the industry coding is performed at each of the individual State Employment Security Agencies (SESA's). In addition, on a 3-year

refiling cycle,

45

codes for existing units are updated through the use of either BLS Form 3023-A (Industry Classification Statement) and 3023-B (Industry verification Form) or similar state versions. The verification form is currently being used in several States on a trial basis. After testing, it is expected that it will be used in place of the classification statement for most industries in order to reduce respondent burden and the cost of refiling. For both of the BLS forms, there are separate versions for each industry division (including an "all industry" version). The information supplied by the employer on the Industry Classification Statement is manually coded at the SESA's. Manual coding of the Industry Verification Form occurs only when the employer indicates that the current activity for the unit differs from the form's computer-generated description of the industry to which the unit was previously coded.

Census -- This agency uses a combination of the available methods. Codes and descriptions are pretested on report forms wherever possible and practical. If information and data are entered on the report form without change or addition to the pretested material, then subsequent coding operations are largely within the computer. If it is necessary for the respondent to alter or add to the pretested descriptive material, then verification and review become necessary. If new codes are assigned, this is done manually, utilizing the computer assisted method if possible. Codes assigned manually are then processed and checked in the computer processing in the same manner as pretested codes, with final codes based on predetermined criteria and procedures or on manual override.

FTC -- Based on the primary business activity and percent distribution of gross receipts by source, the industry coder manually enters a Quarterly Financial Report (QFR) industry code at the top of the Nature of Business Report or Corporate Structure Schedule.

IRS -- During Revenue Processing industry codes are manually assigned to sole proprietorships based on the main business activity.

Partnership and corporate returns carry a taxpayer assigned Principal Business Activity (PBA) code which is keyed in directly from the schedule during data entry. Since 1981, SOI industry coding has been largely automated, with manual coding on an exception basis. For sole proprietorships, the current year revenue processing industry 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.

SSA -- Industrial Classification of SSA's single-unit and multi-unit employers is a manual operation. Codes are assigned directly to the source documents (Forms SS-4 and SSA-5019) based on the principal activity designated by the employer.

2. Treatment of Missing Data.

Each of the systems relies primarily on its own source documents to supply the level of information necessary to assign a complete industry code. However, in those cases where the respondent does not provide sufficient data for the desired level of coding or fails to return the form, the agency must resort to other alternatives. one route which many of the agencies take is to obtain additional information on the unit through further contact with the employer. Another is the use of commercial listings. A third available is to reference either prior codes assigned within the same agency or codes obtained from other Federal systems.

When no additional information is available for the assignment of

a complete industry code, the agencies resort to a code that represents the level of information available. The principal methods of code assignment to these types of cases are described below, with examples of the agencies which apply them:

-- Assignment of "Unclassified" or "unknown" Code. This is a code used by an agency when there is insufficient information to determine the industrial activity at any level. Of the agencies studied, all but BEA and FTC use such a code. The assigned code varies between agencies. For example, SSA and Census assign "0000" and BLS uses "9999." IRS uses "9000" for the "not allocable" code.

-- Force Coding. This is a last resort method used to a limited extent by the Census Bureau for the elimination of incomplete codes within some of their systems by "imputing" the industry code. For example, for tabulation purposes under the County Business Patterns Coding System partially coded cases may be "force coded" to 4 digit industry codes using known distributions of fully coded establishments

within that industry division or group. Also, under the COS and Directory Unit coding of multi-unit establishments, codes are imputed" for unclassified units based on those assigned to other establishments of the same firm.

47

-- Partial Coding. Except for BLS, BEA and FT, the coding systems studied rely on partial coding when there is insufficient information to code to full industry detail. A partial code could be any valid 2 or 3 digit SIC filled in with 0's under Census and SSA coding systems.

Another form of partial coding, which is utilized by SSA, is foldback coding. "Foldback codes" are special 4-digit codes which are used to consolidate 2 or more SIC codes in related areas where full detail is neither attainable from the level of information supplied by the employer, nor is it required for SSA statistical purposes. When

there is insufficient information to determine the full industry detail under one of these groupings, the appropriate "foldback code" is assigned to avoid additional correspondence. The main difference between an SSA partial code and "foldback code" is that the latter is not "0" filled (with the exception of special code 0100 which is used for farming activities). For example, if a unit is engaged in landscaping activities, not further described, it is assigned "foldback code" 0784, instead of partial SEC code 0780. The elimination of "0's" in the last 2 positions of the industry code suppresses any future correspondence with the employer.

A third method is the use of "not allocable" codes within IRS' SOI systems. These codes are assigned when the taxpayer provides enough information to determine the industry division, but the level of information is "not allocable" to a specific industry within that division.

3. Data Entry.

This is the procedure where either the final code is keyed into

the files or where source information is entered for computer coding.

As for entering the codes to the files, this may be done following the manual coding and quality review operations or during the automated coding procedures. The coded information may be keyed from edit sheets, computer listings or directly from the source documents.

Information on these procedures in the study was very limited.

4. Quality Assurance procedures.

Most of the coding systems apply either manual quality control or computer consistency procedures or both for reviewing accuracy and validity of assigned industry codes.

Manual quality control procedures are used in many of the industry coding systems studied. During manual coding operations, coded cases are systematically selected for additional verification for purposes of controlling coding errors. In most cases this is a sample verification. However, there are situations where a 100 percent review is conducted, either because of the size of the unit, or because the industry

coder is inexperienced, or because the quality control sampling specifications call for an initial 100 percent review before going to a sample review.

For example, in SSA systems peer review of work complete by experienced classifiers is conducted on a sample basis within the coding branch along with re-review, by the technicians, of errors charged before the blocks are returned to the classifiers for correction. If the error rate is more than 3 percent the coder's block will be reviewed 100 percent. Also, trainees' work is reviewed 100 percent by the technician until the codes reach a required level of accuracy. In addition, a weekly audit of approximately 1,000 Single-Unit Employer's (Form SS-4) and 5 Multi-Unit Employers (Form 5019), which have already been subjected to peer review, is conducted by the Office of Research and Statistics in order to detect outstanding coding errors and problems in the areas of code interpretations and procedures. Another example is the Census

Bureau's business births coding system where the forms are placed in blocks of 100 and subjected to a 10 percent sample verification. When the verifier's code determination differs from the initial code assignment the case is referred to a lead clerk or a supervisor for a final decision. if the coding differences reach more than 2, the block is subjected to reworking.

Computer consistency and validity checks are an automated method of review found in all of the systems studied. It is primarily used to check for invalid codes, inconsistencies in coding or continuity of code changes. For example, after the codes have been entered in SSA's single-unit (SU) and multi-unit (MU) EI files, the industry code for each record is Computer checked against a list of valid industry codes. Records with invalid codes are printed out on an exception listing. These listings are then checked against microfilm of the original source documents for corrections. Another example of consistency review would be that done for the economic censuses where inconsistencies are flagged during computer processing through edit checks programmed into the system.

In a sense, IRS's new partly automated SOI coding procedures could be regarded as incorporating a consistency check. In this case, the computer, comparison is between the current and prior year codes, and a difference indicates the need for manual review and coding of the sample return.

The use of computer checks in the BEA system is somewhat different from the other systems in that the computer actually generates an industry code using the same procedures as the editor. It then compares it to the editor's code selection in order to check for consistency and validity.

H. Description of Systems Relationships

Existing systems relationships are of considerable importance for suggesting further systems utilization. In considering

Possible new code sharing arrangements it is useful to know something

about the linkages that already exist among the industry coding systems that were reviewed by the Working Group. These are of two kinds: intra-agency and interagency. Intra-agency linkages are not inhibited by legal restrictions; technical and operational factors determine their feasibility and desirability.

1. Intra-agency Linkages

Census - Most of the Census Bureau systems studied produce industry codes that feed into the Standard Statistical Establishment List (SSEL). The preferred source is the quinquennial economic censuses; industry codes assigned to establishments responding to census mail inquiries take precedence because they are based on more detailed information about the establishments' activities than is available from any other source.

For multi-establishment (multi-unit) companies, industry codes are assigned to new establishments and to existing establishments with activity changes on the basis of the Company Organization Survey. This is done annually for the larger multiunit companies and once between 5-year censuses for the small ones. Special coding systems

have been established for unclassified or partly classified units that are added to the SSEL from administrative record sources. A special classification form is mailed to these units during economic censuses. In non-census years, an attempt is made, in connection with the annual County Business Patterns program, to classify these units based on name and on listings in commercial business directories. The business births coding system and other current sample surveys are additional sources of industry codes based on more detailed and/or more recent data. Within the Census Bureau, the industry classification information flows in both directions; once in the SSEL, codes are used to determine eligibility for inclusion in a wide variety of current statistical programs.

IRS -- The IRS systems cover three types of business units: sole proprietorships, Partnerships and corporations. For each type, there are separate coding operations for revenue processing (all returns) and the Statistics of Income program (a sample of returns). The linkage consists in the fact that the same source documents are used in both systems and that the codes assigned in revenue processing are used indirectly in the Statistics of income industry coding, as

explained in Chapter III.G.

2. Interagency Linkages

Interagency linkages are subject to legal restrictions. in general, codes residing in files of statistical agencies cannot be transferred to other agencies for nonstatistical purposes. There are also severe restrictions on interagency transfers of industry codes from administrative systems for

statistical purposes. Nevertheless, some transfers are permitted and do occur. The more significant ones are listed in Table 3 on page 51. This table shows that SSA is an important source of both single unit and multiunit industry codes for the Census Bureau's SSEL, the Federal Trade Commission (prior to the transfer of the OFR to the Census Bureau), and some State Employment Security Agencies. For the

economic censuses, IRS provides the Census Bureau with codes from the revenue processing Systems for the non mail units, including all establishments with no paid employees.

Prior to the passage of the Tax Reform Act of 1976, the Social Security Administration released employer lists, with industry codes, to several Federal agencies, in addition to those shown in Table 3, for statistical purposes. These lists were usually for selected industries and, in some cases for samples of employers in these industries. At present such releases can be made only to agencies for which specific provisions have been made in Section 6103 of the Internal Revenue Code.

The transfers from Census to SSA have been allowed under a U.S. Attorney General's opinion of January 5, 1953, known to the agencies involved as the "McGranery Decision," which allows the Census Bureau to update industry codes for other Federal and State government statistical agencies for statistical purposes, but only for those EIN numbers whose identities are already known to the agencies receiving the codes. Although this opinion has not been rescinded, the last

such transfer occurred following the 1972 Economic Censuses, and at the time was used only to update industry codes on the SSA Single-unit EI File.

A technical problem, as far as inter-system linkages are concerned, is the fact that in the BLS system, EI numbers are not available for all States in the central name and Address File. Other technical problems are apparent such as the differences between the use of the establishments as the basic reporting unit (or multi-unit firms by the Census Bureau, versus the BLS and SSA use of a reporting unit which sometimes includes more than one establishment.

CHAPTER IV

QUANTITATIVE INFORMATION ON COMPARABILITY AND ACCURACY

A. Introduction

The discussion of comparability and accuracy of industry coding so far has been largely in qualitative terms. Factors which 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. Others, 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 the reader, on the basis of these, may have already begun to form some intuitive judgments as to the relative accuracy of codes in different systems.

The purpose of this section is to present some quantitative data bearing on the comparability and accuracy of industry coding in different systems. The data presented come from both published and unpublished sources, the latter consisting largely of items supplied to the Industry Coding Working Group by the agencies participating.

Section B. covers inter-system macro-comparisons, i.e., comparisons of aggregate data by industry from different systems. Section C. presents results from inter-system micro-comparisons, i.e., comparisons of industry codes from different systems for identical units. Section D presents information on components of error in individual systems.

B. Inter-system Macro-comparisons

It is fairly routine for agencies to compare aggregate data for items such 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 --overage, concepts and definitions; nevertheless, comparisons are sometimes useful as a means of detecting gross errors in one or both data sets. Such comparisons 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 this has often been found to be so when

54

individual unit comparisons have been made (see Section Ideally, a useful sequence of investigation would be:

1. Review descriptive material on the coverage concepts and definitions of the data sets compared.
2. Compare data sets at a broad level, e.g., national totals by SIC division or major group.
3. Where large differences are observed, 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.

This idealized 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 this general approach is found in a 1961 report from the Bureau of the Budget. The 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 by only 182,000 (about 1 percent) from those of BLS. The author of the report took this result as a demonstration that "the work over the years had not been in vain" (Bureau of the Budget, 1947).

Another comparison which led to a matching study involved payroll statistics from the retail portion of the economic censuses for 1958 and 1963 (Bureau of the Census, 1965b). The Census data 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 totals exceeded those from the Census Bureau by 5.8 percent in 1958 and 7.2 percent in 1963. This led to a matching study for the State of Delaware, which is discussed in Section C.

The Bureau of Economic Analysis 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 (Bureau of Economic Analysis, 1972). These comparisons, which involved data from the Continuous Work History Sample (both the 1 percent and 10 percent versions), population censuses, the County Business Patterns program, and the Unemployment Insurance system, are summarized in another BEA report (1976, Chapter VII, A Comparison of the CWHS with Other Data Sets). The observed differences are the result of several different 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 authors of the report say that

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 (P. 92).

Government establishments are, in fact, excluded from County Business Patterns data, so the main implication is that the SSA and BLS systems, both of which include government establishments, may have assigned different classifications to some of them during the period covered by these comparisons (mainly 1971 and 1973).

Other more recent aggregate or macro-comparisons are available in both published and unpublished form (for examples of published comparisons, see Office of Federal Statistical Policy and Standards, 1977a, P. 29, and 1980; Metropolitan Washington Council of Governments, 1977; and Harris, 1981), but they do not offer any additional enlightenment on comparability and accuracy of industry coding in different systems.

C. Inter-system Micro-comparisons: General

This section and the two following sections cover 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 matching records from two or more systems.

Most micro-comparisons require two steps. The first is a matching operation to identify records for corresponding units in the systems compared. The matching normally Produces a certain proportion of one-to-one or "perfect" matches, i.e., pairs Of records, one from each system, which 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. There will also usually be cases where the relationships between units in the two systems are mote complex, e.g., one unit in system A may correspond to a grouping of two or more units in system

B, etc. In such cases, a clear interpretation of differences in industry codes is not always possible.

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 inter-system micro-comparisons 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. Such analyses are time-consuming and generally cannot be done on a large scale. Nevertheless, they can be useful in two ways: first, to improve inter-system comparability by uniform treatment of large units, and second, to suggest changes in coding principles and procedures in either or both systems in order to improve their accuracy and comparability.

D. Interagency Comparisons Between Systems

A very early example (Bureau of the Budget, 1947) is reported as follows:

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, SSB) 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 Census and non-manufacturing by SSA, or vice versa. it was found impossible in most cases to reach agreement on the proper classification by examining the information on 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 classified independently as manufacturing or nonmanufacturing by both agencies. The results are shown in Table 4. 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.

Table 4. Results of independent Coding of Establishments by
Census and SSA

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 classi-

fication being preliminary

subject to change pending

additional information..... 21

Census-OASI classification

difference..... 33

Another Bureau of the Census (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 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 SSA and Census in different categories. The category most clearly prone to error was wholesale trade, for which the Census estimate (based only on

58

matched cases) was 43 percent below the SSA estimate., and the estimated index of inconsistency (a measure of variability) was 53. 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 and the SSA source documents had inquiries specifically designed to distinguish wholesale and retail trade. However, the Census inquiry assumes that the respondent knows the difference between wholesale and retail trade, as defined in the

SIC, whereas the SSA source document inquiries do 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 Census. Matching in each direction required some grouping of census 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 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 5, taken directly from the Census Bureaus report (1965b), 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 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 understatement by Census was about 1.6 percent. Among the units classified in retail trade by both Census and BES, 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 1950's and 1960's, 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" after 1958) directly from employers from the SS-4 (Application for Employer Identification Number) or a followup inquiry.

Table 6 shows the main results of this comparison. Of the 1,958,000 census 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 a final census code the same 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 with those resulting from incorrect classification. However, on a priori grounds, 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 digit level, 205,000 were in retail trade, and 165,000 of these (over half of the total) were in eating and drinking places.

In a study following the 1967 Economic Censuses (Bureau of the Census, 1969), final economic censuses 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 multi-unit matching problems and because the Census 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 these units.

Results of the comparison were shown for 37 industries and industry groups in retail trade for which a direct comparison of census and PIA codes was possible. For the 37 groups based on census SIC codes, it was found that 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 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 SIC codes for these establishments.

In this instance, it seems reasonable to assume that the census SIC codes were generally more accurate than the PIA codes,

Table 6. Results of Comparison Between Final Industry Codes and

SSA-Based Mailing List Codes: 1963 Economic Censuses

Establishments

Percent of matched

Result of comparisons	Number (000)	Percent of total	classified to 4-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 4-digit level by SSA...	279	13.2	---
Classified to 4-digit level by SSA.....	1,679	79.3	100.0

Same 4-digit code.....	1,393	65.8	83.0
Same 3-digit, differ-			
ent 4-digit.....	67	3.1	4.0
Same 2 digit, differ-			
ent 3-digit.....	70	3.3	4.1
Same SIC division,			
different 2-digit...	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

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 probably represent a misuse by IRS of these categories, which are intended to be used for clearly defined activities which 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, 1982) compared industry codes assigned to a small sample of sole proprietorships reported on Form 1040 Schedules C and P for 1978 with 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:

Exact match (at the finest level of detail possible

considering differences in the coding systems).....

Partial match (matching on at least the first digit, but not An exact match).....	15
No match (different first digits).....	47

Total	149

This was a small stratified probability sample of Schedules C and E, 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.

E. 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. One example of this is provided by a study (Bureau of the Census,

64

1951) in which industry codes for 500 single-unit establishments from the 1949 Annual Survey of Manufactures were compared 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 7.

Table 7. An Analysis Of 1947-1949 Code Changes for 500

Single-Unit Establishments in Manufacturing

Item	Number of	Percent of	Percent of
	estab- lishments	all cases examined	

Total number of schedules

examined.....	500	100.0	xxx
---------------	-----	-------	-----

Total code changes, 1947

to 1949.....	57.a	11.4	xxx
--------------	------	------	-----

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	xxx
--------------------	---	-----	-----

a Does not include possible code changes for establishments

(estimated 7 percent of total) reporting product combinations

affecting their industry classification.

The striking finding is that, less than one-third of the apparent changes turned out to be real. 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 4-digit SIC level occurred for only 995, or 3.3 percent. 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."

Another report from 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., 10.0

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 division and within the retail division. In the first instance, establishments were classified as in retail in the 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 8.

For the most part, these differences involved shifts between retail and wholesale trade. However, in the case of milk distributors (part of the category "nonstore retailers") and bakeries, the shifts were largely between retail trade and manufacturing.

Table 9 shows classification differences by major kind of business for establishments classified as retail in both the census and the current survey. (As in Table 8, the large multi-unit 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 groups; 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 more of its sales in meats as a meat market, whereas 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. 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 (those assigned in revenue processing) 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 be more accurate than the master file codes, since the SOI industry codes make fuller use of all relevant information on the returns And resort

Table 8. 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
United States, total.....	0.07	-.02

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

Note: These 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

67

68

to commercial directories in some cases. For partnerships and

corporations, the master file codes are usually those entered by

taxpayers.

Table 10 shows results, at the SIC division level, from two studies that compared SOI an master file codes. The measures shown are base] 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 8 are based on unweighted tabulations of SOI sample cases; hence, the larger units are underrepresented.

Based on Table 10, it can be observed that:

-- There are large Differences between the..two systems, and the large indexes of net shifts for some SIC divisions show that these differences do not always tend to balance out. It is difficult to agree with the statement in one of the IRS reports that "on a broad basis, the two coding systems yielded fairly comparable results" (internal Revenue Service, 1973). Considering that both systems used the same source documents, the differences might be considered

surprisingly large.

-- The master file codes for partnerships were largely those supplied by the taxpayers,, whereas 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. It anything, the data point to the opposite conclusion.

- 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.

-- Partnerships classified in transportation and public utilities in the master file and in services in the SOI coding.

69

70

The results shown in Table 10 were based only on cases for which a return was classified in different SIC divisions in the two systems. Table 11 shows, by SIC division, the percent of cases classified differently in the two systems at the division, major industry (two-digit), and industry group (three-digit) levels. Unlike Table 10, this table includes those SOI sample returns for which there was no valid industry code in the master file. As a result, the division level, percents for sole proprietorships and partnerships in Table 11 are lower than those in Table 10.

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

-- For partnerships in agriculture, forestry and fishing, agreement drops off from 86.9 percent at the major industry level to 61.9 percent at the industry group level. This 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 sole proprietorships in finance, insurance, and real estate, agreement drops off from 67.1 percent at the major industry level to 40.2 percent at the industry group level. This resulted primarily from a group of returns classified in real estate in both systems, but classified differently to the seven industry groups used within the major industry.

Table 12 shows data on the extent of agreement at the major industry 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. 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 4-digit industry codes in the Business, Master File were invalid (Internal Revenue Service, 1975b).

Table 12. Agreement of IRS Master File Codes with SOI Codes at

Major Industry Level for Corporation: Tax Years 1972 and 1973

SIC Division	Percent agreement	
	with SOI Codes	
	1972	1973
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

Source: 1972 data--Powell and Stubbs, 1981

1973 data--Internal Revenue Service, 1975a

F. Data on Industry Coding Error in Individual Systems

Director 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. Available data are presented in this section in the following sequence: errors of

nonresponse leading to incomplete classification; response errors, i.e., those occurring in connection with manual coding or data entry; and general information not restricted to specific components of error.

1. Errors of Nonresponse

There are various methods of dealing with incomplete data for industry classification. The evidence at hand on the results of these effort 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 (Internal Revenue Service, 1984). An agency-by-agency presentation of available data follows:

Census -- The most significant nonresponse problem for the Census Bureau is that connected with new or re-activated establishments (births). For single-unit enterprises, information about new units is received primarily for 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, 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 types of 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; however these 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 multi-unit 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 (Bureau of the Census, 1981) report shows 60,613 or 1.4 percent of all establishments as completely unclassified; however, these 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.

BEA - According to the description of the classification system used for the agency'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.

BLS - 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 systems description prepared for the Industry Coding Working Group, the State Employment Security Agencies, which are responsible for the industry coding, are expected to deal with incompleted 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 the

74

employer contacted by telephone to obtain the needed information. In the interim, the establishment is put in an unclassified 9999 group.

Change to a specific code is made as soon as possible, usually by the next quarter."

FTC - According to the systems description for Quarterly Financial Report (QFR) industry coding, there is no incomplete classification. Over 99 percent of the units are classified by reference to the source documents or commercial 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.

Parenthetically, it can be observed that industry classification errors by IRS could have resulted in coverage errors for the QFR program, since the sampling frame provided by IRS included only corporations classified in the 4 SIC divisions within the scope of the OPR 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.

IRS - The extent of incomplete classification in the SOI (sample-based) files can be determined from publications. Table 13 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 it is taken into account that the figures for corporations are an overstatement, as explained in the footnote to Table 13.

The 1979 data for partnerships, 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 OE businesses included in the ERS individual and business master files. However, in all likelihood the proportions unclassified and partially classified are considerably higher than in the SOI files. It is known, as stated earlier in this parts, that for tax year 1969 there were no valid industry codes in the master file for 20.1 percent of the sole proprietors, 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. modes for "not allocable"

within SIC, division are not used in industry coding for the master files.

Table 13. 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)

Partially classified 1/.....	1.7
Manufacturing, miscellaneous and not allocable 1/.....	0.5
Wholesale, miscellaneous and not allocable 1/.....	1.1

Wholesale and retail,

not allocable..... 0.1

Unclassified..... 0.15

SOLE PROPRIETORSHIPS (1977)

Partially classified..... 3.3

Farms, not allocable..... 1.3

Construction, not allocable..... 0.5

Manufacturing, not allocable..... *

Wholesale, not allocable..... 0.4

Retail, not allocable..... 0.3

Wholesale and retail,

not allocable..... 0.8

Unclassified..... 0.3

* Less than 0.05 percent.

1/ The figures for these categories are overstated, since they

include some fully classified returns in SIC major groups 39
(miscellaneous manufacturing industries) and industry groups 509
(miscellaneous durable goods) and 519. (miscellaneous nondurable
goods).

76

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 one-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 to 1975) it averaged 14.6, with a high of 21.3
percent in 1975. Levine explains this increase as follows:

"...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.

SSA -- 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 or reporting units in the multiunit employer identification file.

Data from a matching operation following the 1963 Economic Censuses presented earlier in this part (Table 6) showed that 279,000 out 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 one-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 and/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.

2. 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.

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 an interval of 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 our principal merchandise lines, whereas the initial interview 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 recheck data without reference to the original questionnaires and codes.

Code differences were observed for 98 establishments, 18 percent of the total included in the recheck. Results of an analysis of the reasons for difference are shown in Table 14. About two-thirds resulted from differences in the information in the original and recheck questionnaires, presumably resulting from the more detailed data requirements in the latter. It was further stated that commodity

breakdowns with percentages were "helpful or necessary for proper ... coding" in 22 of the 98 cases with differences.

Table 14. Reasons for industry Code Differences Between Initial and Recheck Surveys: Retail Trade Surveys, 1948

Reason attributed for difference	No. of cases	Percent of total differences
1. Informational differences.....	67	68
2. Coding differences (same information)	25	26
3. Miscellaneous problems.....	6	6
Total.....	98	100

source: Bureau of the Census, 1951.

The evaluation of industry classification in the Employer Record

Check of the 1960 Population Census (described earlier in this section) 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 Census of Population, 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 their establishment's principal activities, products and services; and industry codes based on this information were

78

compared with those assigned to the same persons from information reported by or for them in the Population Census.

Table 15 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 1970 report:

"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." (Bureau of the Census, 1977a, p.4)

Table 15 also shows that the indexes of inconsistency by industry were lower in 1970 than in 1960. Possible reasons for this change are not discussed directly in. the Census Bureau's report, except for a brief statement in the "Highlights' section as follows:

"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; however, significant changes could have occurred in the training of coders; the quality and coverage of reference materials, such as company name lists, available to coders; the effectiveness of quality control procedures; and 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 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.

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.

(3) Those within scope and having employment, but with employment below designated cutoffs that varied by industry. Only a sample of these establishments was included in the mail portion of the census.

The technique used in each of these studies was to mail economic census questionnaires to a sample of units in the group. The returned questionnaires were used to evaluate the accuracy of census information, including industry codes, that was normally being derived from administrative record sources. Indirectly, therefore, these studies provide information on the quality of industry codes in the

IRS and SSA systems; however the emphasis in the reports of the studies is on the accuracy of economic census results, regardless of their source.

A recent report (Hanczaryk and Sullivan, 1980) studies active establishments with employees included in the SSEL but refined as being out of scope of the economic censuses. The study universe comprised about 558,000 establishments. Of these 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 possessions. 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 classified 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 were 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 16 shows the results of comparing SIC classifications based on census questionnaires with those based on administrative records for the same establishments in these 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. Administrative codes for

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

Table 16. Comparison of SIC Codes Based on Census Questionnaires with those Based on Administrative Records: 1977 SIC Censuses

Type of establishment and SIC division 1/	Percent agreement at 2/			
	Division level	2-digit level	3-digit level	4-digit level
Nonemployers				
Retail trade.....	69.8	58.0	46.7	NA
Service.....	79.1	70.0	NA	NA
Employers below cutoff				
Retail trade.....	95.8	89.6	85.0	81.3
Service.....	97.4	96.1	94.1	94.1

NA - Not available

1/ - Division per administrative record code.

2/ - Weighted to reflect varying sampling rates used.

Source: King and Ricketts, 1980.

The third study reported by King and Ricketts (1980) was a study of nonemployers administratively classified in construction. 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 17, are presented somewhat

82

83

differently; they show the net effects of classification changes on the totals by major industry. Overall, there was a net reduction of 12 percent in the number of nonemployer establishments in

construction. About half of this resulted from the removal, of duplicate listings from the census 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 mail questionnaires. Data were collected for a sample of this group by telephone. 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 one percent, and the relative net changes by major industry, as might be expected, were considerably smaller than those for nonemployers.

3. 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 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	Range for 10 service centers
Nonfarm	0.9%	0.1 to 2.5%
Farm	0.9%	0.0 to 4.9%

Systems descriptions for SSA's single and multi-unit industry coding both stated that "audits" (based on sample verification) conducted by SSA's Office of Research and Statistics "...show approximately a 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 multi unit system), it seems Likely that the overall outgoing quality is somewhat lower than 97 percent.

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 there have been no results published.

Turning once again to the coding of industry for persons, there was a carefully designed study of "coder effects"

84

in the 1960 Census of Population (Bureau of the Census, 1972). This study,, which was based on a comparison of codes entered on the same set of census questionnaires (or copies thereof) by the original census coders and by other coders, 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. Data are presented primarily at the SIC division level. Here may be seen 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.

4. Data on Sources of Codes.

It seems reasonable to suppose that when the industry codes in a file come from several 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 18 (industry codes for multi-unit establishments virtually all come from the economic censuses or from current surveys of the Census Bureau). The first 7 SIC divisions listed in the table are those which are fully or partly included in the economic censuses. The out-of scope division includes two groups: first, about 482,000 establishments in SIC divisions B (mining) through I (services) in industries not included in the economic

censuses, and second, 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 birth listings are provided,monthly to the Census Bureau. industry codes came from this source for 26.4 percent of the in scope and 35.8 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 other characteristics such as employment, payroll and receipts, are distributed by industry source code. No direct data are published, but it can

be observed 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. On the other hand, the division with the lowest proportion of Census-based codes -- construction with 54.4 percent -- averages only 6.8 employees per establishment (Census Bureau, 1982a). Furthermore, virtually all of the industry codes for establishments in multi-unit 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 data codes are not available.

CHAPTER V

REFERENCES

Bailar, B.A., and Kallek, S.

1980 Evaluation of the 1977 Economic Censuses. Pp. 17 in Proceedings
of the Section on Survey Research Methods, 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, American Statistical Association.

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. Washington, D.C.: Bureau of the Budget.

Bureau of the Budget

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. Office of Statistical Standards, unpublished report. Washington, D.C.: Bureau of the Budget.

Bureau of the Census

1951 Effect of different documents, coders and time periods on industry classification of establishments. Memorandum from Morris E. Hansen to Thomas J. Mills, August 7. Attached report

with same title by H.T. Goldstein, July 15. Washington, D.C.:

Department of Commerce.

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.: Department of Commerce.

Bureau of the Census

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.: Department of

Commerce.

Bureau of the Census

1965b Final results of BES-Census retail payroll reconciliation

for the State of Delaware. Memorandum from Peter Harvey

Kailin and William Hurwitz, July 22. Washington, D.C.:

Department of Commerce.

Bureau of the Census

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. Washington, D.C.: Department of Commerce.

Bureau of the Census

1969 Comparison of Census SIC codes with Internal Revenue PIA codes

for same establishments. Memorandum from George Minton to Joseph

F. Daly, November 12. Washington, D.C.: Department of Commerce.

Bureau of the Census

1972 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.: Department of Commerce.

Bureau of the Census

1977a Evaluation and Research Program of the 1970 Census of
Population and Housing: The Employer Record Check. PHC(E)-

12. Washington, D.C.: Department of Commerce.

Bureau of the Census

1977b 1977 Industry and Product Classification manual.

Washington, D.C.: Department of Commerce.

Bureau of the Census

1979 The Standard Statistical Establishment List Program. Technical

Paper No. 44. Washington, D.C.: Department of Commerce.

Bureau of the Census

Washington, D.C.: Department o Commerce.

Bureau of the Census

1982a Special tabulation: number of single-unit and multiunit establishments, and number of employees, by trade area, data from 1979 SSEL. Washington, D.C.: Department of Commerce.

Bureau of the Census

1982b Computer printout: 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. Washington, D.C.: Department of Commerce.

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, Department of Labor.

Washington, D.C.: Department of commerce.

Bureau of Economic Analysis

1976 Regional Work Force Characteristics and Migration Data: A

Handbook on the social security Continuous work History Sample

and Its Application. Washington, D.C.: Department of Commerce.

Central Statistical Board

1939 Annual Report for Fiscal Year 1939. Washington, D.C.

Commission on Organization of the Executive Branch of the

Government (Hoover Commission)

1949 Task Force Report on Statistical Agencies (Appendix D). January.

Washington, D.C.

Farrell, M.G., Jabine, T.B., and Konschnik, C.A.

1982A review of industry coding systems. Pp. 278-283 in
Proceedings of the Section on Survey Research Methods,
American Statistical Association.

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.

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, 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, Susan

1983 The Verification Method as a Solution to an Industry Coding

Problem, pp. 499-504 in Proceedings of the Section on Survey
Research Methods, American Statistical Association.

Internal Revenue Service

1973 A comparison of industry coding of tax returns during revenue

processing and statistical processing. Unpublished report by
Linda Bouchard, Statistics Division. January 15. Washington,

D.C.: Department of the Treasury.

Internal Revenue service

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. Washington, D.C.: Department of the

Treasury.

Internal Revenue Service

1975a Consistency of BMF industry code to SOI industry codes at

major industry level, 1973. Unpublished tabulation,

Statistics Division. Washington, D.C.: Department of the

Treasury.

Internal Revenue Service

1975b Possible and impossible BMF industry codes, 1973.

Unpublished tabulation, Statistics Division. Washington,

D.C.: Department of the Treasury.

90

Internal Revenue Service

1982 A pilot study of the extent and accuracy of EIN and business

activity reporting on tax returns by sole proprietors.

Unpublished report, Statistics Division. Washington, D.C.:

Department of the Treasury.

Internal Revenue Service

1984 Descriptions of Selected Industry Coding Systems, Supplement to A

Review of Industry Coding Systems, Statistics of Income Division

Internal Revenue Service.

Jabine, T.B.

1984 The Comparability and Accuracy of Industry Codes in Different

Data Systems (in draft). Committee on National Statistics.

commission on Behavioral and Social Sciences and Education.

Washington, D.C.: National Academy of Sciences.

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, American Statistical Association.

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, 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.

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.: Department of Commerce.

Office of Federal Statistical Policy and Standards

1977b Standard Industrial Classification Manual, 1977 Supplement.

Washington, Department of Commerce.

Office of Federal Statistical Policy and Standards

1980 Report on Statistical Uses of Administrative Records. Statistical

Policy Working Paper 6. Washington, D.C.:

Department of Commerce.

Office of Management and Budget

1972 Standard Industrial Classification Manual. Statistical Policy

Division. Washington, Office of Management and Budget.

Office of Management and Budget

1974 Enterprise Standard industrial Classification Manual.

Statistical Policy Division. Washington, D.C.: Office of

Management and Budget.

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, American Statistical Association.

President's Commission on Federal Statistics

1971 Federal Statistics. Report of the President's Commission.

Washington, D.C.: Government Printing Office.

Simmons, W.R.

1953 The elements of an industrial classification policy. Journal of

the American Statistical Association (48):429-439.

CHAPTER VI

SELECTED SOURCE DOCUMENTS AND INSTRUCTIONS

A. Introduction

The source documents and instructions in this Chapter 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 purposively to illustrate different levels of detail, as well as the difference between a Document designed for self-coding by the respondent (TRS Form 1065) and documents designed for coding by the 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 the Level of detail available on these forms appears in Chapter III. Seven forms and the corresponding instructions, if any, appear in this chapter as separate exhibits. In the sections which follow, each exhibit (1 through 7) is briefly described: the form and the coding system or Systems for which it is used are identified, and a few explanatory remarks about the items used for coding and the coding system are provided.

B. Exhibit 1 (page 99)

1. Source document

Form 1065, U.S. Partnership Return of income, Tax Year 1981

2. Industry coding systems

The source document shown is used in the following industry

coding systems of the internal Revenue Service:

- a. Revenue processing of partnership-returns;
- b. Statistics of income (SOI) for partnerships (for a sample of

returns)

3. Remarks

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 (Business Code number), 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 (Principal Business Activity), B (Principal Product or Service), 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.

C. Exhibit 2 (page 101)

1. Source document

Schedule C (Form 1040), Profit or (loss) From Business or
Profession (Sole Proprietorship), Tax Year 1981

2. Industry coding systems

The source document shown is used in the following industry
coding systems of the Internal Revenue Service.

- a. Revenue processing of sole proprietorship returns
- b. Statistics of Income for sole proprietorships (for a sample
of returns)

3. Remarks

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. The present coding system for SOI sole proprietorships is partially automated, making use of revenue processing industry codes when available.

D. Exhibit 3 (page 103)

1. Source document

Form SS-4, Application for Employer Identification Number

(Revised 9-82)

2. Industry coding system

The source document shown is used in the Single Unit Employer Identification (EI) Number System of the Social Security Administration.

3. Remarks

Shown are the full form and the instruction 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.

E. Exhibit 4 (page 106)

1. Source document

Form V-B-5502, 1982 Census of Retail Trade: Tires, Batteries,
Parts, Accessories

2. industry coding system

This form is used by the Census Bureau as a source document for coding industry in their economic censuses.

3 Remarks

This is one of a large number of specialized forms that was used in the mail portion of the 1982 Economic Censuses. As explained in the text, many of the items in the questionnaire are used in the largely automated industry coding process. The key item is Item 11-Merchandise Lines.

P. Exhibit 5 (page 108)

1. Source document

Form BLS 3023-A7 (Revised December 1982), Industry Classification

Statement: Wholesale Trade

2. Industry coding system

This source document is used by the Bureau of Labor Statistics (BLS) for their "Employment and Wages (ES-202) System

3. Remarks

The complete form is shown. This is one of several versions tailored to particular ST% divisions; this one is for wholesale trade. The form is used for updating classification information for employers already in the system. This is now being done every three years for most employers. In addition, it may be used, on new employer accounts.

G. Exhibit 6 (page 110)

1. Source document

FTC Form 59-103 (revised 09-79) Nature of Business Report

2. Industry coding system

The source document shown is user for the Quarterly Financial report (QFR) Program. Responsibility for the OFR was transferred to the Census Bureau in late 1982.

3. Remarks

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 two years. This version is used for corporations in manufacturing; a second version is used for the other SIC divisions included in the program.

H. Exhibit 7 (page 112)

1. Source document

BEA Form BE-12, Benchmark Survey of Foreign direct Investment in

the U.S., 1980.

2. industry coding system

The source document shown is used by the Bureau of Economic Analysis (BEA) for their Foreign Direct Investment System.

3. Remarks

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.

Working Paper Series

1. Report on Statistics for Allocation of Funds; GPO Stock Number
003-005-00178-6, price \$2.40

2. Report on Statistical Disclosure and Disclosure-Avoidance
Techniques; GPO Stock Number 003-005-00177-St price \$2.50

3. An Error Profile: Employment as Measured by the Current,
Population Survey; GPO Stock Number 003-005-00182-4, price \$2.75

4. Glossary of Nonsampling Error Terms: An Illustration of a
Semantic Problem in Statistics (A limited number of copies are
available from OMB)

5. Report on Exact and Statistical Matching Techniques; Gpo Stock
Number 003-005-00186-7, price \$3.50

6. Report on Statistical Uses of Administrative 'records; GPO Stock

Number-003-005-00185-9, price \$5.00

7. An Interagency Review of Time-Series Revision Policies (A limited number of copies are available from OMB)

8. Statistical Interagency Agreements (A limited number of copies are available from OMB)

9. Contracting for Surveys (Available through NTIS Document Sales, PB-83-233-148)

10. Approaches to Developing Questionnaires (Available through NTIS Document Sales, PB-84-105-055); Paperback \$16.00; microfiche \$4.50

11. A Review of industry Coding Systems (Available through NTIS Document Sales, PB-84-135-276)

Copies of these working papers, as indicated, may be ordered from the Superintendent of Documents, U.S. Government Printing office,

Washington, D.C.0402 (202-783-3238) or from NTIS Document

Sales, 5285 Port Royal Road, Springfield, VA 22161

(703-487-4650).