CORPORATION INDUSTRY CODING

Allison Clark, Internal Revenue Service

Every tax return in the Corporation Statistics of Income (SOI) program at the Internal Revenue Service (IRS) is coded according to industry for data analysis and tabulations. This four-digit code is determined by the corporation's principal business activity (PBA) and is the basis for all classification. This paper describes IRS' current system for industry coding corporate tax returns and looks at an alternative method to improve accuracy of the codes. Comparisons are then made to industry coding systems used by other agencies.

THE CURRENT SYSTEM

The industry groupings used for corporation income tax returns conform with the Enterprise Standard Industrial Classification (ESIC) codes used by the Department of Commerce, Bureau of the Census. This structure of classification follows closely along the lines of the Standard Industrial Classification (SIC). Some departures from the system, however, are made for the finance industries, reflecting particular provisions of the Internal Revenue Code [1].

The SIC is the statistical classification standard underlying all establishment-based Federal economic statistics classified by industry [2]. It is revised periodically to reflect the economy's changing industrial organization. The last major revision was in 1987. Currently, the codes used by SOI correspond to the prior SIC revision of 1972, and the ESIC of 1974. Most four-digit codes used in SOI data correspond to SIC codes at the three-digit or two-digit levels. Some examples are as follows:

Meat Products.--SOI code 2010 corresponds to SIC industry group Number 201, including the four-digit SIC codes--

- 2011 Meat Packing Plants;
- 2013 Sausages and Other Prepared Meat Products; and
- 2015 Poultry Slaughtering and Processing.

Apparel and Accessory Stores.--SOI 5600 corresponds to SIC major group 56 at the two-digit level, and the following four-digit codes--

- 5611 Men's and Boy's Clothing and Accessory Stores;
- 5621 Women's Clothing Stores;
- 5632 Women's Accessory and Specialty Stores;
- 5641 Children's and Infants' Wear Stores;
- 5651 Family Clothing Stores;
- 5661 Shoe Stores;
- 5699 Miscellaneous Apparel and Accessory Stores [3].

While some SOI minor industry codes are broader than others, the majority correspond to SIC and ESIC three-digit codes.

On the Form 1120, Corporation Income Tax Return, a company is asked to classify itself according to its principal business activity. A list of codes is included with the instructions for the return. The company must also state its business activity and its product or service on the tax return. The answer to these two questions helps to determine the proper SOI industry code. However, corporations do not always answer these questions as precisely as is needed for statistical purposes. Some examples include:

(1) Business Activity- Manufacturing
Product or Service- Stainless Steel Coils
(2) Business Activity- Manufacturing
Product or Service- Metal Products
(3) Business Activity- Manufacturing
Product or Service- Various.

A diversified company may not even classify itself, instead listing its PBA code as 9999 or undeterminable. In SOI, each major group of industries has a miscellaneous or not allocable category, if a precise code cannot be determined. Examples are miscellaneous manufacturing and manufacturing not allocable, other retail stores, and nature of business not allocable. Using the above examples, the first corporation would receive a precise industry code (3370-ferrous metal industries, miscellaneous primary metal products), the second a miscellaneous category within metals (3490-miscellaneous metal products), if a precise product could not be determined, and the final corporation would be classified in miscellaneous manufacturing and manufacturing not allocable (SOI industry 3998), if no further information could be determined. Fortunately, for SOI statistics, these categories are extremely small; every effort is made to classify the corporation into a particular SOI industry. This is done using various reference materials such as Moody's Manuals and Directory of Corporate Affiliations, which have information on large corporations and their business activities.

The principal business activity is a large factor in coding corporate returns. During data abstraction, all records pass through an automatic process known as the Industry Code Library System (ICL) [4], if the company has been included in our statistical sample for any of the three years prior to the current year [5]. If a corporation has a change in its reported PBA from a previous to the current year, the industry code is manually edited. If a record has the same PBA as a previous year, the industry code from the prior year is automatically transferred, making it the SOI industry code of the current year. If there was no prior year PBA match (the corporation was not in our sample in previous years), and the corporation's total assets are less than $2.5 million, the corporation's PBA is used as its SOI industry code. For all other cases (such as a new corporation with total assets greater than

$2.5 million), the industry code is edited manually. In the 1986 SOI sample, of the 85,096 returns in the file, sixty-six percent matched PBA's from the prior year. Thus, the prior year industry code became the current year code. Seven percent did not match PBA's, while twenty-seven percent had no prior year match.

Clearly, there are some problems with the Industry Code Library. What a company reports as its PBA may, in fact, not be where most of its income is derived. Also, a company may continue using a PBA code even when its product mix has changed due to mergers and acquisitions. In the 1986 SOI file, 26,270 records out of 85,096--approximately thirty-one percent--had PBA's which did not match their industry code, which was up slightly from thirty percent in 1985. However, if the returns matched in the ICL, the industry code is not manually examined, it is determined automatically. This is especially a problem with large consolidated returns. Usually a large corporation will have many subsidiaries operating in different types of businesses. From one year to the next, different subsidiaries may generate more income or a company may acquire some new companies and/or sell others. From year to year the percentage of consolidated returns increases. Seventy-five percent of the corporation income tax returns in the 1986 SOI sample were consolidated, steadily increasing from previous years. The increasing number of consolidated returns shows how accurate the statistics must be.

A HYPOTHETICAL CASE STUDY

When filing a consolidated return, the corporation will usually include a schedule of its affiliates and their principal business activities. For example, a hypothetical company, Blue Bank and Consolidated Subsidiaries (ACS), has four subsidiaries and files the schedule shown in Figure 1. If the parent corporation and each subsidiary were coded separately, Blue Bank ACS would represent five industries. Instead of having five industry codes, however, a return is classified according to (with a few exceptions) a "total receipts rule," where the industry code is the code for

Figure 1. -- Blue Bank ACS. Affiliations Schedule

Subsidiary	Business Activity
Blue Leasing	Equipment Leasing
Marine Club	Restaurant
Blue Realty Credit	Real Estate Investments
Aqua International Holdings	Miscellaneous Services

the business activity accounting for the largest percentage of the corporations total receipts [6]. Total receipts is defined as--

(a) gross receipts less returns and allowances;
(b) dividends;
(c) interest in U.S. obligations;
(d) other interest;
(e) rents;
(f) royalties; and
(g) other taxable income, not including gains from the sale or exchange of assets, or capital gain net income.

All this information comes from the income portion (lines 1 through 11) of the Form 1120, U.S. Corporation Income Tax Return.

Suppose Blue Bank ACS files a consolidated income tax return with the income schedule summary shown in (Figure 2). Looking at the income summary and using the total receipts rule, one can compute that Blue Leasing accounts for the most receipts ($175 million). Therefore, the principal business activity of Blue Leasing would become Blue Bank ACS's industry code. However, if Blue Bank (the parent company) and Marine Club were engaged in the same business activity, combined they would account for the largest amount of total receipts ($328 million) and their business activity would become the consolidation's industry code. Note that although Marine Club has the largest amount of total income ($158 million) this is not identical to the amount of "total receipts." From this example it is easy to see how a diversified corporation can change industry codes from year to year.

Figure 2. -- Income Summary of Blue Bank ACS -- A Hypothetical Corporation

Receipts	Blue Bank ACS	Blue Bank	Blue Leasing	Marine Club	Blue Realty	Aqua International
1. Gross Receipts	453,975,238	123,660,553	174,145,997	147,067,852	9,100,836	
a. Less Returns & Allowances	0	0	0	0	0	
b. Balance	453,975,238	123,660,553	174,145,997	147,067,852	9,100,836	
2. Cost of Goods Sold	78,915,026	29,777,421	40,996,459		8,141,146	
3. Gross Profit	375,060,212	93,883,132	133,149,538	147,067,852	959,690	
4. Dividends	13,887,177	13,826,295		60,882		
5. Interest	40,133,075	30,830,462	4,621	9,283,083		14,909
6. Gross Rents	3,016,677	155,387		225,154	2,520,000	116,136
7. Gross Royalties						
8. Capital Gain (Loss)	(286,015)	(1,665,165)			1,379,150	
9. Ordinary Gain (Loss)	165,134	76,888		2,442	85,804	
10. Other Income	3,731,912	2,315,347	1,055,885	302,625		58,055
11. Total Income	435,708,172	139,422,346	134,212,486	158,404,550	3,479,690	189,100

HOW MERGERS AND ACQUISITIONS AFFECT STATISTICS

In the past few years, there have been many mergers between large diversified companies. Consequently, companies filing consolidated tax returns can have subsidiaries in fifty or more distinct business activities. Unfortunately, we are not able to tabulate each subsidiary separately based on its activity; therefore, by using the total receipts rule, the company is classified according to where the majority of receipts are received [7].

Beginning in 1985, all giant returns in the corporate SOI program which have a different industry code from the prior year are researched to verify the current industry code. In 1985, out of 192 industry code changes of large returns, 108 switched into another industry within the same industrial division [8]. Of the 108, sixty-eight switched within the Finance, Insurance and Real Estate Division (FIRE). Twenty-six switched to another industry within manufacturing. In 1986, out of 274 industry code changes, 174 remained within their original division. Again, the FIRE division accounted for most of the change (sixty-four percent), with manufacturing following at twenty percent. A record is also kept of companies with total assets of $500 million or more which were chosen from the sample in a previous year, but which are missing from the current year sample as a result of a merger. In 1984, twenty-eight companies merged, sixty-seven in 1985, and seventy-nine in 1986 [9]. Most of the corporations merged into the same industry, or the company with which they merged kept the same industry code.

It is apparent that the number of large missing returns due to mergers is increasing, and that more returns are merging with companies with different industry codes. This has an effect on statistics from year to year. Although mergers sometimes do change a corporation's industry code, usually the corporations remain within the same industrial division, so, in the aggregate, the statistics are relatively unchanged.

RECLASSIFYING SOI INDUSTRIES USING ASSETS: A HYPOTHETICAL EXAMPLE

The increasing number of mergers for large diversified companies suggested that there may be a need to reassess whether total receipts is an appropriate criterion for coding industries on consolidated corporation income tax returns. With that in mind, a study was conducted using another measure of a company's economic activity--total assets.

Assets are reported on corporation income tax returns in the same manner as receipts, using the line items 1-14 of the Balance Sheet provided in Schedule L of the Form 1120. Like receipts, assets are also reported by the parent and subsidiary companies. Revenue which represents operation was used in determining a "total assets" rule similar to that for total receipts. The components included are shown in Figure 3 on the next page.

Using the hypothetical Blue Bank ACS as an example, again, the industry code can be refigured based on assets rather than receipts (See Figure 4). Here, the parent corporation, Blue Bank, has the majority of assets at $1.2 billion. The industry code of the consolidated return would be that of the parent company, which is 6090 for banks. Although Blue Leasing has a large amount of assets in other investments ($307 million), these are not used in computing the industry code with the "total assets rule," since the company is involved in leasing rather than finance. Blue Leasing's total assets, as computed by the total assets rule, would be $51.6 million instead of $359 million when adding other investments.

Comparing the "total assets" and "total receipts" rules shows that Blue Bank ACS would receive conflicting industry codes. Blue Leasing had the majority of receipts, giving Blue Bank ACS an industry code of 7389 for equipment leasing, while the largest amount of assets were shown by Blue Bank, the parent company, which would give the consolidation an industry code of 6090 for banks.

To determine how likely this was to occur in real life, examples from five industrial divisions--construction (general building contractors), manufacturing (bakery manufacturing), wholesale trade (wholesale metals and minerals), transportation (water transportation), and finance (security brokers and dealers) were examined and the five largest returns were reclassified using assets instead of receipts. Figure 5 summarizes the results.

Surprisingly, there was very little difference in the data between assets and receipts. In the majority of returns examined, the subsidiary or subsidiaries which accounted for the majority of receipts also accounted for the majority of assets. This was true of the five largest returns examined in the bakery products, water transportation, and security brokers industries.

The most fluctuation was in general building contractors. While the largest receipts were reported from subsidiaries involved in construction activity, the largest assets were reported by subsidiaries involved in mortgage activity, largely in mortgage and real estate loans. In construction, it appears that the finance component generates more assets than receipts. In the metals and minerals industry, only one of the five largest corporations changed industry codes (to a petroleum industry) using assets. This occured due to a large amount of assets in the "other assets" category.

From the limited research done, it does not appear that using assets to code corporation returns would differ significantly from receipts. Only in a few instances in the examples used did a corporation vary in the amount of income generated by receipts or assets.

Although there are the obvious drawbacks of placing a consolidated tax return in only one industry, it appears that for now our "total receipts" rule is an accurate coding measure, since the change using assets appears minor.

SOI INDUSTRY DATA vs. OTHER AGENCIES: A COMPARISON

Many individuals and agencies request SOI industry data. The most common complaint is

Figure 3. -- Asset Components Included in Industry Coding, By Industrial Division

Asset Component	Agriculture	Transportation	Construction	Manufacturing	Wholesale & Retail Trade	Finance, Insurance & Real Estate
1. Cash	X	X	X	X	X	X
2. Trade notes & accounts (net)		X	X	X	X	
3. Inventories						
4. Federal and state obligations						X
5. Other current assets	X	X				X
6. Loans to stockholders						X
7. Mortgage and real estate loans						X
8. Other investments						X
9. Depreciable assets (net)	X	X	X	X	X	X
10. Depletable assets (net)	X	X	X	X	X	X
11. Land						X
12. Intangible assets						
13. Other assets	X					X

Figure 4. -- Consolidated Balance Sheet for Blue Bank ASC -- A Hypothetical Corporation

Assets	Blue Bank ACS	Blue Bank	Blue Leasing	Marine Club	Blue Realty	Aqua International
1. Cash	2,485,761	2,200,000		33,542		200,870
2. Trade notes & accounts	38,996,402		38,795,532			
a. Less allowance	7,001,276		6,903,814			97,462
3. Inventories	32,678,545			32,678,545		
4. Federal and state obligations	306,806,337	306,806,337				
5. Other current assets	547,266,803	474,026,970		52,826,937	20,412,896	
6. Loans to stockholders	20,011,724	20,011,724				
7. Mortgage and real estate loans	494,332,655				494,332,655	
8. Other investments	725,681,002	418,268,559	307,412,443			
9. Depreciable assets	8,224,562		4,112,281	4,112,281		
a. Less accumulated depreciation	2,448,917		2,000,000	488,917		
10. Depletable assets						
a. Less accumulated depletion						
11. Land	627,449				627,449	
12. Intangible assets						
a. Less accumulated amortization						
13. Other assets	349,087,662	301,983,371	17,553,618	22,117,469	7,426,752	6,452
14. Total assets	2,516,708,709	1,523,296,961	358,970,060	111,279,857	523,029,219	132,612

Figure 5. -- Reclassification of Industries Using Assets

Prior-Year Industrial Classification: Industry Code*	Prior-Year Industrial Classification: Number of Returns	Subsequent-Year Industrial Classification, Industry Code*: 1510	2050	4200	5050	6120	2910	6150
		Number of Returns						
1510	5	2						3
2050	5		5					
4200	5			5				
5050	5				4		1	
6120	5					5		

*1510 - General building contractors; 2050 - Bakery products; 2910 - Petroleum refining (including integrated); 4200 - Trucking and warehousing; 5050 - Metals and minerals, except petroleum and scrap; 6120 - Savings and loan associations; and 6150 - Business credit institutions.

that the industries defined in the SOI codes are not as specific as most people would like. Another is that unlike subsidiaries of a common parent are grouped together. This is addressed in the SOI Corporation Source Book which reads,

> ...the SOI classification system...has limitations. A return is classified by industry based on the activity accounting for the largest percentage of total receipts. This means that large corporations with diversified activities are included in only one industry even though many of their business operations are unrelated to the industry in which they are classified.

SOI vs. CENSUS

Different agencies classify industries in various manners. In its 1982 Census of Manufacturers, the Bureau of the Census used an "establishment" basis of reporting, collecting data on employment, payroll, value of shipments, assets, depreciation, etc. For the census, a company operating at more than one location was required to file a separate report for each location [10]. Companies engaged in distinctly different lines of activity at one location were requested to submit separate reports, if the plant records permit such a separation, and if the activities were substantial in size. The 1982 Census of Manufacturers universe included approximately 345,000 establishments. All establishments with more than twenty employees were mailed report forms. An establishment was then classified in a particular industry on the basis of its major activity during a particular year, i.e. production of the products primary to that industry exceeds, in value, production of the products primary to any other single industry.

The Bureau of the Census publishes Concentration Ratios (a measure of how concentrated an industry is, in comparison to the largest companies in that industry) in Manufacturing during various years. These are completed for the four largest and eight largest companies in an industry. For this series, their industry coding system is identical to SIC codes at all digit levels. Using their data at the four-digit level, for various food manufacturing industries, their ranking of concentration ratios can be compared to ranked ratios using SOI data [11], as shown in Figure 6.

Suprisingly, given the diversification of SOI industries, there is a high correlation between SOI and Census data. If industry 2060 was deleted, industries 2081, 2040, 2030, 2088 and 2089 would correspond exactly.

SOI vs. FEDERAL TRADE COMMISSION

In 1972, the Federal Trade Commission introduced a "Line of Business Data" which separated large companies according to their lines of business. In 1974, the Commission began collecting 1973 line of business data for large diversified manufacturers. A "line of business" observation was defined as a single company reporting its activities to a given FTC industry category, and the business had to have $10

Figure 6. -- Concentration Ratios,Ranked by Census and SOI Data, from Highest to Lowest

Census	SOI	SIC Codes
2060	2081	2010: Meat Products
2081	2040	2020: Dairy Products
2040	2030	2030: Fruits & Vegetables
2030	2088	2040: Grain Mill Products
2088	2060	2050: Bakery Products
2089	2089	2060: Sugar & Confectionery
2050	2020	2081: Malt Liquors & Malt
2010	2050	2088: Alcoholic Beverages
2020	2010	2089: Bottled Soft Drinks & Flavorings

million or more in net operating revenues. The industry coding system used by the FTC is based on SIC and U.S. Bureau of the Census codes, and most are comparable to SOI codes at the four-digit level. From the data collected, the FTC computed many types of business ratios, among them advertising to sales and assets to sales. Figure 7 shows results when using their statistics from the food manufacturing industries and computing the statistics from SOI data for the comparable year [12].

Figure 7. -- Ranking of SOI to FTC Data

Advertising to Sales		Assets to Sales	
SOI	FTC	SOI	FTC
2089	2088	2088	2088
2060	2089	2030	2081
2040	2040	2060	2089
2050	2081	2089	2060
2030	2030	2050	2030
2096	2096	2081	2040
2081	2050	2040	2096
2088	2060	2020	2050
2020	2020	2096	2020

From the determination of advertising to sales ratios using FTC and SOI data, one can view the similarities and differences. SOI industries 2020, 2030, and 2040 all correspond closely to the FTC data. However, there are also large fluctuations, most notably in SOI industries 2060 and 2088. It is interesting to note that when comparing concentration ratios, industry 2060 varied greatly between the two data sets as well. There is also much fluctuation in the asset to sales ratios, although for both sets of data, SOI industry 2088 was the largest. Taking into account that company data reported to the FTC varies from data reported to the IRS due to different accounting methods, standards and tax years, the data are suprisingly (with a few exceptions) close, as most industries are generally within the same range. In advertising to sales, of the top five companies, three are the same in both sets of data. In the assets to sales, four of the top five companies correspond [13].

CONCLUSION

There is ongoing research in our industry coding process. Currently our industry codes are being updated to reflect new SIC industries, mostly in the services and computer equipment areas [14]. We are also looking at ways to improve our automated industry coding processes, especially with many new mergers. Regular consultations are made with other agencies in order to further streamline our data. These and other future initiatives will allow us to continue producing accurate industry statistics.

ACKNOWLEDGMENTS

Thanks are due to the following people in the Internal Revenue Service who reviewed this paper and provided helpful comments: Fritz Scheuren, Ken Szeflinski, Karen Cys, Keith Gilmour and Glenn Hentz. Additional thanks are also due to Beth Kilss and Wendy Alvey for both review and preparation of the final manuscript.

NOTES AND REFERENCES

[1] Examples include Regulated Investment Companies, I.R. Code Section 851, and Real Estate Investment Trusts, I.R. Code Section 856.

[2] For statistical purposes, an establishment is different from a company. An establishment is "an economic unit, generally at a single physical location, where business is conducted or where services or industrial operations are performed." 1987 Standard Industrial Classification Manual, Executive Office of the President, Office of Management and Budget, p. 12. A company may consist of one or more establishments.

[3] For a comparison of SOI codes to SIC and ESIC codes, see "A General Description of the Corporation Source Book," Statistics of Income, Internal Revenue Service, Publication 647, pp. 9-17.

[4] During data abstraction an individual or "editor" will survey a return and abstract certain data fields by hand while other data fields are automatically printed by computer. For instance if an industry code was not printed by the computer the editor must manually "edit" an industry code based on the corporations principal business activity using the guidelines described in this paper.

[5] The corporation sample is approximately 90,000 returns, although this varies from year to year since the returns are selected through a computer process. Consequently, there are some new returns each year, and others which were in a previous year may not be included in the current year. One exception is "giant" returns, (total assets of $250 million or more) which are included in the sample each year. For further explanation of sampling activity, see 1985 Statistics of Income-1985, Corporation Income Tax Returns, Publication 16.

[6] Some exceptions to the total receipts rule are as follows:

- Mutual savings banks are coded 6030, bank holding companies are coded 6060.
- Regulated investment companies are coded 6742, real estate investment trusts are coded 6743 and small business investment companies are coded 6744.
- Combinations of manufacturing and construction are usually coded for manufacturing, combinations of manufacturing and trade are coded for manufacturing when most of the receipts are from whatever is manufactured by the related co.
- Returns of newpapers or periodicals or other publishings which show advertising as the largest item of receipts are not coded as advertising but are coded as newspapers, periodicals or miscellaneous publishing.

[7] Eisenbies, Robert A., William F. Long and David J. Ravenscraft, "The Financial Performance of Leveraged Buyouts," a paper given at the Workship on Tax Microsimulation Modelling for Businesses, Internal Revenue Service, U.S. Department of the Treasury, also discusses this topic.

[8] Corporation income tax statistics are divided into nine industrial divisions:

1. Agr., Forestry, & Fishing;	6. Whlesale & Ret. Trade;
2. Mining;	7. FIRE;
3. Construction;	8. Services; and
4. Manufacturing;	9. Nature of Bus. Not Allocable.
5. Trans. and Pub. Utilities;	

[9] This number is based on known mergers, where the parent was readily identified through research. This does not include corporations from which we have information that they have merged but have no data on their current parent corporation.

[10] Concentration Ratios in Manufacturing, 1982 Census of Manufactures, U.S. Department of Commerce, Bureau of the Census.

[11] For comparison purposes, Census four digit codes were totaled and averaged to correspond to SOI four digit codes. Only the giant (total assets of $250 million or more) SOI returns were used, to compare to Census' four largest or eight largest companies. While value of shipments are used to determine concentration ratios by the Bureau of the Census, concentration ratios using SOI data were calculated by business receipts as reported in the Corporation Source Book, 1982, Publication 1053, Internal Revenue Service.

[12] Statistical Report: Annual Line of Business Report, 1977, Federal Trade Commission, Report of the Bureau of Economics.

[13] As with the Census data, some FTC four-digit codes were summed and averaged for closer approximation. See Statistical Report: Annual Line of Business Report, 1977, op. cit., pp. 26-27. Again, using SOI data, business receipts were used to approximate sales.

[14] See for example Hanczaryk, Paul S. and W. Joel Richardson, "Industrial Activity Coding of Non-employer Businesses in the 1987 Economic Censuses," U.S. Department of Commerce.