

Vito Natrella, Consultant

The Small Business Administration is required by Title II of the Small Business Act to conduct research in the field of small business. As discussed in this session, a number of approaches have been or are being developed in order to provide the data bases needed for this analysis. One of these is a microdata set representative of both small and large businesses.

Using simulation microdata models, analysis could be developed to evaluate the impacts of different policy choices affecting small business such as economic regulation, environmental and health rules, and taxation. A microdata set could also be used to study the determinants of new business ventures and the characteristics of firms going out of business. A file to meet these needs must contain a variety of information. This would include financial data, geographical information and information on employee characteristics. In addition, a basic requirement is company classification according to employment size.

A considerable amount of data exists from various sources involving these aspects. However, they need to be integrated and made internally consistent. Since no one series contains all the information needed, it is necessary to augment and merge records. Depending on the circumstances this can be achieved by means of an identical match, by a synthetic or statistical match or by a multiple regression technique. Problems associated with these techniques for inputting missing data include high cost, inconsistent definitions, sparsity of cases, poor quality, induced biases, and issues of confidentiality.

#### A Proposal for a Small Business Microdata Base

Initial development of a small business microdata base could be in the direction of a system of three separate but related files based essentially on the IRS Statistics of Income samples: corporate, partnership, and sole proprietorship. The SOI samples have the advantage that they are stratified random samples and contain almost all the needed data. The system would encompass a Company Financial File, an Establishment File and an Employer-Employee File.

The Company Financial File would consist of the SOI sample files augmented by employment data. The SOI files contain complete income account and tax computation items for all forms of business and, in addition, balance sheet items for corporations and partnerships. Geographic designation is included but refers to the central office for large corporations filing consolidated tax returns. Since employment is to be used as the main size

classifier, this item must be added from another source. This might be done from the Standard Statistical Establishment List (SSEL) of Census or from the 941 file of IRS. Each of these has problems which will be discussed later. The Company Financial File could be made more useful by updating some industries on the basis of the Federal Trade Commission Quarterly Financial Report.

In order to permit geographical analysis an Establishment File should be set up. This can be done by obtaining data for the establishments of the firms included in SOI samples in the Company Financial File, from the SSEL of Census. Data would cover industry, employment, receipts, geographic location and employment size of the owning enterprise for each establishment.

The third file in the system of microdata files would contain information on the employees of the firms and would be designated as an Employer-Employee File. This would be developed from the Continuous Work History Sample of Social Security Administration associated with the Establishment File. The file would be arranged by employer and contain data for all its employees. This would permit analysis of the employee mix, characteristics and wages for small business.

#### Access to Microdata

There are really two parts to the problem of access to microdata. One is that, in order to match data from a number of data files, identification is essential if we want to achieve the greatest accuracy. The other involves obtaining access to microdata without identification so that researchers in small business would have the opportunity and flexibility to work directly with the data.

##### 1. Internal Revenue Service

According to Section 6103 of the Internal Revenue Code authorizations are limited, very specific as to conditions, and restricted to particular stated purposes. Paragraph (j) provides for disclosure of tax returns and return information for statistical use. Only four agencies are included in this section to receive authority for specific statistical purposes. They are the Bureau of the Census for all returns and the Bureau of Economic Analysis for corporate returns, to the extent necessary in the structuring of censuses and national economic accounts. The Federal Trade Commission has access to corporate tax returns in order to select its survey sample for the Quarterly Financial Report. The Treasury has access in order to prepare required forecasts, projections and analyses.

These data may be disclosed to others

only in such form which cannot identify, directly or indirectly, a particular taxpayer. The IRS chief counsel has in past years refused to make the business tax models, without direct identification, available since it was felt taxpayers could be indirectly identified by a knowledgeable person.

## 2. Bureau of the Census

Release of microdata in identified form to government agencies is even more restricted for the Census Bureau than it is for the IRS. No agencies currently have access although legislation has been framed to permit statistical agencies (as defined in the law) to receive names and addresses of businesses together with some data for classification purposes from the Standard Statistical Establishment List. The basic purpose of the SSEL is to provide a frame of businesses from which a stratified sample can be drawn. While the SSEL is on an establishment basis, employment is aggregated according to owning company.

## 3. Federal Trade Commission

Access to microdata from the Quarterly Financial Report of the Federal Trade Commission is severely restricted by law similar to the case of the Census Bureau. There are issues involved limiting the usefulness of FTC data. The main deficiency is limitation to only four major industries: manufacturing, mining, wholesale and retail trade. Also, the sample is quite small--15,000 corporations in all four industries. This makes for weakness in the estimates by size of company and industry. Except for the large corporations there is little overlap with the SOI sample. If it is possible to add employment to the QFR survey it may be more feasible to simply have FTC prepare current period estimates by size and relating to SOI tabulations.<sup>1</sup>

## 4. Social Security Administration

Access to data from the Continuous Work History Sample (CWHS) of Social Security is now in a state of suspense. In the past, microdata had been made available in unidentified form for years prior to 1976. However, IRS has objected to making that data available on the grounds that tax return information could be disclosed indirectly.

### Comparability of Data

An important requirement for setting up a microdata file is that data from different files be merged. In addition to the access issue, the issue of comparability of data presents problems. In some cases, while nothing can be done to put the data on a comparable basis, it is desirable to be cognizant of the differences. Some of these arise because of differences in timing, levels of consolidation, or accounting

methods.

The SOI corporation sample file, considered the basic file for a small business data base, has certain characteristics which differ from those of files to be used for merging and enhancements. The SOI file for any particular year includes data for corporations with fiscal years ending in July through June of the following year. Aggregate financial data are roughly centered on the calendar year, although only 40 percent of all corporations have calendar year accounting periods. However, these account for over 70 percent of total net income. Also, for the most part, corporation returns, Form 1120, are filed on a consolidated basis where consolidation includes subsidiaries owned 80 percent or more. Stockholder annual reports, generally the basis for the Dun and Bradstreet Financial-Economics file, consolidate subsidiaries which are over 50 percent owned.

The most important enhancement to the SOI sample file consists of adding employment to the records so that companies can be classified according to number of employees. Data on total employment appear in the IRS Form 941 filed for the first quarter of the calendar year. Of considerable significance is the fact that businesses file Form 941 on a variety of bases. For the most part, data are for each establishment owned by a particular company under separately issued EINs. They are not associated with the EIN of the parent company. Since Form 1120 is filed for a much higher level of aggregation it is usually not possible to locate employment data for many of the multi-establishment firms.

There are a number of partial solutions to this problem currently in progress. IRS is abstracting EINs for subsidiaries of corporations in its tax year 1979 sample. Census' SSEL, which obtains data on employment and payrolls from the IRS 941 file, associates establishments and subsidiaries with owners and parents. This permits aggregation of employment to a company or enterprise level, closer to that used for Form 1120. While these approaches may not give exact employment figures for large multiestablishment firms, they will provide usable employee size classifications.

Accounting methods provide another area of difference between data from tax returns and data from stockholder reports. Methods used in preparing tax returns are geared to minimizing profits so as to minimize tax. Tax laws also provide special incentives to reducing taxes in order to achieve certain economic or social purposes. These may be handled differently according to standard accounting practice. Examples include depreciation, depletion, tax-exempt interest, and installment sales. The main difficulty with the noncomparability due to accounting methods would arise between the SOI

sample files and the Dun and Bradstreet data. These data would also have the problem of different consolidation rules.

### Conclusion

A number of approaches are possible to meet the problems raised above. Some are already underway, while others require considerable planning. The approach that is the most feasible but also the most costly is for SBA to contract out with agencies which already have part of the microdata base to augment and enhance their system. That agency would maintain the file, perform the necessary processing, and produce tabulations needed. SBA has contracted with IRS to have the SOI sample files enhanced with employment data from the 941 file. The Bureau of the Census could also be a candidate for this sort of arrangement. Census has access to the SOI sample files and can use the SSEL to obtain employment figures for enhancement. The SSEL has an advantage over the raw 941 file in that establishment and subsidiary data and EINs are grouped together according to ownership so that total company employment can be determined.

A recent opinion by the Office of the Chief Counsel for the IRS may make IRS access to the Census data possible. The new opinion in effect says that the confidentiality provisions of Title 13 (Census) take precedence over Title 26 (IRS) as applied to IRS employees temporarily sworn in as Census agents. IRS may, therefore,

be able to enhance its SOI sample files using the superior SSEL employment data for companies.

An approach which might be more satisfactory to SBA is to modify the microdata files so that they could be made available to another government agency. The objection on the part of IRS to making microdata business files available is that even with identity and location removed it may be possible to identify a taxpayer indirectly. After the SOI sample file has been enhanced with an employment classifier IRS would first remove all direct identifiers. It could then use a system of grouping of large corporations within a particular industry. These groups could be as small as three. IRS is currently studying this and other approaches.

As mentioned by Kirchhoff and Hirschberg in a paper included in this volume, legislation is being introduced to provide SBA with access to the Census Standard Statistical Establishment List. Although it may be difficult, the attempt should be made to change Section 6103 of the Internal Revenue Code to permit SBA to have access to IRS data similar to Census. SBA's requirements are fully backed by the Small Business Act and should be accepted if proper safeguards are designed and made part of the law.

NOTE: This paper was developed on the basis of reports prepared by the author for the Small Business Administration.

1. After the SOI sample has been enhanced with employment data.