CONSTRUCTING AND MAINTAINING FRAMES AND SAMPLES FOR BUSINESS SURVEYS

Carl A. Konschnik, Nash J. Monsour, and Ruth E. Detlefsen
Bureau of the Census

I. Overview

This paper describes the list frame construction and maintenance, sample updating, and survey control procedures for the current business surveys of the United States Bureau of the Census. In this first section an overview is given beginning with a brief discussion of the current business surveys and their sample revisions and continuing on with summary descriptions of the source for the list sample frame, the initial frame construction and sample selection, and the periodic updating of the frame and samples. Following this, the supplementary area sample and the notion of controlling the surveys over time are briefly introduced. In subsequent sections these topics are covered in greater detail.

I.A. The Current Business Surveys

The current business surveys consist of seven distinct surveys for which various economic data are collected and published. Monthly sales are the principal data collected and published for the Monthly Retail Trade Survey (MRTS) and the Monthly Advance Retail Trade Survey (MARTS), while both monthly sales and end-of-month inventories are included in the Monthly Wholesale Trade Survey (MWTS). The Monthly Retail Inventory Survey (MRIS) covers retail inventories. The Annual Retail Trade Survey (ARTS) for retail and the Annual Trade Survey (ATS) for wholesale use the same basic samples as their monthly counterparts, the MRTS and the MWTS, respectively. The MARTS and MRIS samples are subsamples of the MRTS sample. For the annual surveys additional data, most notably those relating to inventory valuation methods, value of purchases, accounts receivables (for retail only), and sales taxes, are requested along with annual sales and end-of-year inventories. For selected services, the Services Annual Survey (SAS) collects annual service receipts data. A monthly survey for selected services industries was discontinued in 1981 due to budget cuts.

The monthly surveys provide monthly estimates of levels and of changes in levels from a month ago and a year ago at the total U.S. level for major Standard Industrial Classification (SIC) codes. The MRTS also provides estimates of retail sales by SIC and geographic area (regions, divisions, selected states, and selected large Standard Metropolitan Statistical Areas (SMSAs) and cities). The annual surveys give annual estimates of levels and year-to-year changes in levels for major SICs at the U.S. level only.

From 1951 to 1968, the U.S. Bureau of the Census published monthly estimates of total retail sales and other related statistics on the basis of an area sample supplemented by a relatively small list of the largest firms. Since 1968, a list sample has primarily been used for the current surveys, taking advantage of administrative records for frame construction and survey control, while retaining a small area sample for coverage purposes. A description of the change from primary reliance on an area sample to primary reliance on a list sample, along with comparisons of statistical, cost, and control considerations is given by Daly et al. (1972).

I.B. Sample Revisions for the Current Business Surveys

Periodically the Business Division of the Bureau of the Census selects new samples of employer business firms for the current business surveys covering the retail and wholesale trades and selected services industries. These sample revisions occur roughly every five years, with the most recent revisions introduced in January of 1982, and work in progress on revisions to take effect in January of 1987. Descriptions of all of the various processes followed in revising the business sample surveys are given by Isaki et al. (1976) and Wolter et al. (1976).

I.C. Source for the List Sample Frame

The list sample frame for the current business surveys is constructed from the Census Bureau's Standard Statistical Establishment List (SSEL), a universe list of employer business firms and establishments constructed and periodically updated with the administrative records of the Federal Government (chiefly those of the Internal Revenue Service (IRS) and the Social Security Administration (SSA)), the results of the quinquennial economic censuses, and current surveys. By establishment we mean an economic unit, generally at a single physical location, where commerce is conducted or where services or industrial operations are performed. By business firm we mean the company, business enterprise, or corporation at the highest level of corporate structure, which owns or operates one or more establishments. For example, a retail firm may own or operate several stores in which it sells its goods. Each store is considered as an establishment of the firm.

I.D. Initial Frame Construction and Sample Selection

The initial sampling frame for the sample revisions is constructed by first extracting the appropriate establishments and their associated data fields from the SSEL. Establishments not in the retail, wholesale, or services SICs of interest are excluded, as well as those which are excluded for other reasons.

Two types of sampling units make up the list frame for business surveys. All multi-establishment (multiunit) firms which exceed

predetermined SIC sales size cutoffs are considered as "alpha" (or sometimes called "company") sampling units. Multiunit firms which do not exceed the cutoffs and single establishment (singleunit) firms are sampled as Employer Identification (EI) number sampling units.

In the initial sample selection, alpha sampling units are selected with certainty (weight = 1). Also, singleunit EI sampling units are selected with certainty if they exceed the same sales size cutoffs mentioned above. The remaining EI sampling units are assigned to strata based on SIC and sales size and subjected to a simple random sampling within strata.

I.E. Updating of the List Frame and Samples

To account for new business employers (birth employers) which come into being after the initial sampling frame is created and the initial sample selected, it is necessary to conduct additional sampling. Similarly, selected sampling units which become inactive must be deleted from the active mailing lists for the surveys. The appropriate updating of the initially selected samples is accomplished first by a large "birth backlog" sampling operation before the new samples are introduced. This is followed by quarterly birth processings which continue for as long as the samples are in use.

An additional list frame, the EI Control File, is constructed for this initial birth backlog and subsequent quarterly birth sampling operations. The EI number is the primary identifier on the EI Control File. This file is constructed and maintained primarily from administrative records data supplied by IRS and SSA, and, once established, is updated quarterly.

I.F. The Area Sample

An area sample is used as a supplementary sample for the Monthly Retail Trade Survey and for the Services Annual Survey. The area sample is a sample of land segments and is used to cover business establishments not on the list frame. These businesses are of two types: nonemployer businesses and new employer businesses (birth employers). Nonemployer businesses never become part of the list sample frame, while birth employers must be represented by the area sample until they are subjected to sampling in the quarterly birth processing.

I.G. Controlling the Surveys Over Time

Survey control ensures that the samples properly represent the target population at any time period for which estimates are made. The need for survey control measures is evident because of the births and deaths of businesses as well as other types of changes, and the need to unduplicate the list and area samples in the estimates. This work encompasses the proper coordination of the computer files and operations used to control the surveys, ensuring proper mailing, processing, and estimation.

II. Composition and Updating of the Standard Statistical Establishment List (SSEL)

II.A. Composition of the SSEL

The SSEL contains approximately 4.2 million singleunit firms with positive payroll in the most recent year, and about one million establishments of multiunit firms with positive payroll in the most recent year.

The primary identifier for singleunits on the SSEL is the Employer Identification (EI) number, which is assigned by IRS and which is the primary taxpayer identifier used by employer business firms. Under the Federal Insurance Contributions Act (FICA), every organization with paid employees must have an EI number. For multiunits, the primary identifier on the SSEL is a Census Bureau assigned "alpha-plant" number. The "alpha" portion uniquely identifies the firm, company, or enterprise to which the establishment belongs; the "plant" portion gives each establishment or plant a number within the firm. Since every employer firm is required to file tax returns under an EI number, the EI number used by the employer to file a return covering each establishment is assigned as a secondary identifier for that multiunit establishment. For clarification, note that an employer firm may have as few as one EI number under which it files tax reports for all its establishments, or as many as one EI for each establishment, at the firm's option. Generally, multiunit firms have only one active EI number which covers all of its establishments.

The SSEL maintains many data fields for each of its establishments. For sampling purposes, the following are among the most important: SIC code, type of operation code, name and address, employment, payroll, sales or receipts, and geographic codes. The use of these and other data fields are discussed more fully as they are encountered in this paper. SIC codes are defined in the Standard Industrial Classification Manual of the Office of Management and Budget (1972) and the Standard Industrial Classification Manual Supplement of the Office of Federal Statistical Policy and Standards (1977). The retail and wholesale trades, selected services industries, and all other industry groups are defined by SIC codes. SIC codes are two, three, or four digits depending on the detail of classification. The Census Bureau further expands the four digit SIC codes to as many as six digits for certain retail, wholesale, and services SICs where more detail is needed. These expanded codes are called Industry and Product Classification (IPC) codes and are defined in the IPC Manual, Bureau of the Census (1982). However, our use of the term SIC code will refer to either the SIC code or the expanded IPC code.

The SSEL contains all employer establishments of all divisions as defined by the SIC Manual with these exceptions: Major Groups 01 (Agricultural Production - Crops) and 02 (Agricultural Production - Livestock) of Division A (Agriculture, Forestry, and Fishing) are not carried on the SSEL; establishments in

Major Group 88 (Private Households) of Division I (Services) are not contained on the SSEL.

II.B. Updating of the SSEL

The SSEL is updated quarterly with two computer files for EIs which are obtained by the Census Bureau from IRS, the 941 payroll file, and the Business Master File (BMF). All firms with paid employees are required to file IRS Form 941, Employer's Quarterly Federal Tax Return, with IRS on a quarterly basis. This form is filed under the appropriate EI number of each establishment of the firm. The report includes the total quarterly 941 payroll for the EI. About 4.4 million EI records are on the 941 file for the quarterly update. This quarterly file is accumulated by the Census Bureau from payroll files which are sent weekly from IRS.

The BMF contains information about EIs and their 941 and other tax return filing requirements. An EI is said to be "BMF active" if it has an active 941 filing requirement, or equivalently, if it is required to file Form 941 for the next quarterly period. If there is an indication that an EI has gone out of business or will no longer have payroll, IRS assigns it an inactive 941 filing requirement code. In this case we say the EI is "BMF inactive." The BMF also provides the mailing address for the EIs. The BMF is obtained from IRS on an annual basis followed by monthly supplements which indicate changes since the annual file. The annual file contains about seven million BMF active EIs. The monthly supplements accumulated to a quarterly file contain about 1 1/2 million records of various types of changes, such as name and address or filing requirement changes, to the BMF annual file. Since only EIs with payroll get added to the singleunit SSEL, the BMF changes affect only those which have been added because of their payroll activity. BMF and other data are retained on separate files for EIs which have no current payroll and thus have not been added to the SSEL.

Another file which is obtained from IRS less frequently contains the receipts data for EIs from the annual tax returns. These data are received by the Census Bureau only for the economic censuses years and the year preceding them. In the economic censuses, the receipts data are used chiefly for EIs which are nonmail units or are nonrespondents. Also, these files contain industry coding information in the form of Primary Industrial Activity (PIA) codes which are useful substitutes for SIC codes if no more reliable classification data are available.

SSA provides the Census Bureau with its chief source of SIC codes for new business EIs. The basis for SSA's industry coding is the IRS Form SS-4 (Application for Employer Identification Number), of which SSA gets a copy. The SS-4 requests information on the principal business activity, number of expected employees, type of organization, class of customer, and other data. SSA assigns an SIC code based on the information provided on the Form SS-4. On a yearly basis, the Census Bureau receives approximately 3/4 million EI numbers for businesses from SSA along with an SIC code and other information. If SSA cannot assign an SIC code due to inadequate information, an "unclassified" code is assigned. About 80 percent of the EIs received each year have an SIC code assigned, though some do not have a full four-digit SIC code assigned. The business birth EI records containing SSA assigned SIC codes are received monthly by the Census Bureau.

Recent economic censuses files are used to update the SSEL. SIC codes, census year payroll, employment, receipts, type of operation codes, and other data are updated for all active business establishments. This updating is done following the completion of the economic censuses which occur at five-year intervals in years ending with 2 or 7.

Multiunit establishments are updated on the SSEL through the Census Bureau's annual Company Organization Survey (COS). This survey canvasses all multiunit firms with 50 or more employees on an annual basis, requesting an updated establishment list and payroll, employment, and SIC information for each establishment. In addition, multiunit firms with fewer than 50 employees are surveyed every third year on a rotating basis for the second, third, and fourth year after the economic censuses. In economic censuses years, only multiunit firms out of scope to the economic censuses are canvassed in the COS, since the inscope multiunits are included in the censuses.

A final source of information for updating the SSEL comes from the current surveys of the Census Bureau. For example, during the monthly canvassing of the MRTS, firms are asked to list any new establishments or indicate establishments which have been closed, sold, or otherwise have become inactive. In the birth sampling operations, refined SIC and additional establishment information is obtained for the selected sample of birth EIs, and these data are used to update the SSEL.

III. Extraction of the Business Surveys Frame from the SSEL and Initial Sample Selection

In this section the extraction of the sampling units from the SSEL and the resultant sampling frame are described. Then the initial sample selection is explained. These operations take place every five years as mentioned above. Some important points are not discussed here. These include: first, the reasons for the choice of the sampling units (both EIs and alphas are used); and secondly, the parameter studies needed to determine the design and allocation of the samples subject to the many variance constraints on the estimates to be produced by the business surveys. These are fully discussed in Isaki et al. (1976).

III.A. Extraction from the SSEL

The extraction of the records from the SSEL used in constructing the initial list sample frame is done in two separate operations because the SSEL is separated into two main files: the singleunit SSEL file consisting of single establishment firms with the EI as the primary identifier; and the multiunit SSEL file which

contains firm level data under the alpha identifier and the associated establishments of that firm under an alpha-plant identifier. In addition to the single establishment EIs, the singleunit SSEL contains the EIs of all establishments on the multiunit SSEL. Since 941 payroll and BMF activity data are available from IRS only on an EI basis, these multiunit EIs (which are referred to as "EI submasters") contain the relevant payroll and BMF activity data of their associated establishments.

As part of the extraction operations, each inscope establishment with payroll is assigned a measure of size for stratification. In this paper, by "inscope" we mean that the SIC is included in the survey, that the type of establishment is included in the survey (most government establishments are excluded, for example, as well as certain nonmerchant wholesalers identified by their type of operation), that the establishments are located in the U.S., and that certain other restrictions apply. Since sampling strata are set up by SIC and by annual sales (or receipts in the case of services) of the most recent census year, the measure of size for stratification is a census year "pseudo sales" measure. For each establishment the pseudo sales are obtained by multiplying the establishment's census sales (for establishments in the most recent census) by the ratio of its most recent year's annual payroll to its census year's annual payroll, then deflating that product to the census year basis by an overall SIC level deflation factor. For an establishment not in the census, its nonzero payroll in the most recent year is converted to a pseudo sales by multiplying by a sales-to-payroll factor at the SIC level, then its pseudo sales are deflated back to the census year.

In the singleunit extraction, the EI submaster records are placed on a separate file for later use since they contain the BMF activity information for their associated establishments. A singleunit EI is placed on the EI frame only if it: 1) is inscope to one of the retail, wholesale, or services trade areas and has the required number of digits in its SIC code to qualify for sampling (generally four digits); 2) is payroll active (i.e., has nonzero payroll in the most recent year); and 3) is BMF active. Singleunit EIs which do not have the required number of digits in their SIC code are subjected to sampling in the birth backlog discussed in section V.

In the extraction, the EI records are edited for correct sales and payroll to ensure that each establishment gets a proper pseudo sales assigned. EIs which fail these edits are reviewed, corrected, reedited, and added to the frame. In the most recent extraction in 1980 using payroll data for 1979 as the most recent year available, about two million singleunit EIs were placed on the sampling frame for all three trade areas, retail, wholesale, and services, out of a total SSEL file of 4.2 million payroll active singleunit EIs.

The multiunit extraction creates a file of all establishments which satisfy 1) and 2) as for the singleunits above. These multiunit plants are also edited, corrected, and placed on a file to be used first for conducting the multiunit certainty determination, then for constructing an EI frame for noncertainty sampling.

As part of the singleunit and multiunit extraction, a "proof of universe" operation is done. This process compares the number of establishments, payroll, and pseudo sales of the extracted files with comparable data from the most recent censuses. Tables are produced at the three-digit SIC summary level. This proof of universe serves to help ensure that there are no deficiencies in the constructed sampling frame.

III.B. The Initial Sample Selection

Prior to sample selection the extracted files of inscope singleunit and multiunit establishments are separated into the retail, wholesale, or services trade area. If a multiunit firm has payroll active establishments in more than one inscope trade area, these establishments will be found in the appropriate trade area's file. A firm is subjected to sampling in any trade area in which it has active establishments. Thus three separate sampling frames are established, and the following sampling operations are repeated for each trade area. It is worth noting that the SIC system assigns a single SIC code to each establishment based on its principal business activity. Thus, an establishment can be in one and only one trade area.

Before determining which firms qualify as alpha sampling units, the pseudo sales of the plants under the firm's alpha are accumulated to each SIC level represented. A major SIC for the alpha is assigned which is the SIC with the largest pseudo sales. The remaining SICs represented by the alpha are referred to as minor SICs.

The first step in sampling is the multiunit certainty determination, that is, the determination of all alphas which qualify for certainty status. The major SIC and all minor SICs of each alpha are compared to SIC certainty cutoffs. If the alpha's total pseudo sales across all SICs exceeds the certainty cutoff for the major SIC or if the pseudo sales for any minor SIC exceeds the certainty cutoff for that minor SIC, the alpha is selected as a certainty alpha sampling unit. In particular, the certainty alpha sampling unit consists of all establishments or plants of the firm without regard to the EI number or numbers assigned to its various establishments. In the retail sampling for the 1982 sample revisions, about 1600 certainty alphas were selected. These alphas include about 190,000 retail establishments. For wholesale and services, the comparable number of certainty alphas selected were about 1000 and 1100 respectively.

A singleunit certainty determination is also done whereby singleunit EIs are selected based on the same SIC by sales size cutoffs. In the 1982 sample revisions there were about 500 retail, 210 wholesale, and 360 services singleunit EIs selected with certainty in the singleunit certainty determination.

Multiunit firms or alphas not selected with certainty are formed into EI sampling units based on the EI associated with each establishment. There may be one or more EI sampling units for each alpha. The EI sampling unit is an EI summary level record with establishment pseudo sales accumulated to an EI total for each SIC found among the establishments. The SIC with the largest pseudo sales is designated as the major SIC of the EI sampling unit and is the SIC in which the EI will be subjected to sampling. The sum of the pseudo sales for all SICs is the pseudo sales to be used as the EI sampling unit's measure of size.

The noncertainty sample selection program is then run using the EI sampling frame consisting of singleunit EIs which did not exceed the certainty cutoffs and multiunit EIs of alphas which did not qualify for certainty selection. The frame is stratified by SIC and sales size, with each EI assigned to a stratum on the basis of its total pseudo sales and its major SIC. Simple random sampling is used within each stratum. In retail the frame is also stratified by geographic area for some SICs.

Three monthly panels of noncertainty units are selected in each trade area. In the most recent sample revisions, about 8700 retail EI sampling units were selected per panel with monthly weights ranging from about 3 to about 950. For wholesale the number of selected EIs was about 1400 per panel with monthly weights ranging from about 3 to about 600. For services there were about 4200 selected EIs per panel with monthly weights ranging from 3 to about 1000. The weights are the inverses of the probabilities of selection.

Table 1 gives summary characteristics of the retail employer universe, frame construction, and selected sample for the Monthly Retail Trade Survey. Of the 7000 singleunit EIs in the sample, about 500 are singleunit certainties. The remaining 6500 singleunit EIs, when added to the 2200 selected multiunit EIs, comprise the 8700 selected noncertainty EIs for the monthly surveys. The numbers in the table have been rounded, but they nonetheless give an accurate picture of the relationships among the universe, the frame, and the sample.

Table 1

Monthly Retail Trade Survey (MRTS)
Initial Sampling (in thousands)

	Universe	Frame	Monthly Sample
MU Firms	44	1.6 Alpha 50 EI	1.6 Alpha 2.2 EI
MU Estabs.	330		190 Alpha 10 EI
SU Estabs.	970	970 EI	7 EI
Total Estabs.	1,300		

IV. Setting Up the Files for Survey Control: EI Control File, EI Universe of Sampled Cases, and Survey Registers

IV.A. The Files for Survey Control

To ensure that the samples properly represent the target population for any time for which estimates are made, the appropriate files must be created and maintained to keep current with business unit changes. The changing nature of the business sampling units in the population is illustrated by the fact that each year newly selected birth EIs consisting of approximately 10 percent of the selected noncertainty sampling units will be added to the survey registers and an equal amount of the noncertainty units will drop out of the samples because they are no longer BMF active. Since births as well as initially selected EIs go BMF inactive, over the five-year period that the samples are in use, about 1/3 of the initially selected noncertainty units become BMF inactive. In all, the noncertainty sampling units account for about 60 percent of the retail estimates, and 75 and 60 percent of services and wholesale estimates respectively. The percentage of unit changes for the larger certainty (weight one) sampling units is considerably less than for the noncertainty units.

The EI Control File is the principal file used for survey control. Updated quarterly with new EIs from payroll, BMF activity and SIC coded files, it serves as the frame for birth sampling as well as the depository of activity change and sampling status information for EIs. The EI Control File is also essential for area sample processing, enabling the determination to be made as to whether an establishment in the area sample should or should not be tabulated for the monthly survey. This file is created following the initial sample selection and is used to determine the birth backlog EIs and all EI changes (births, reactivations, and deaths) on a quarterly basis thereafter.

Other important computer files which are used for survey control are the EI Universe file and the survey registers. The EI Universe contains a record of each EI number which has been subjected to sampling in the business surveys, along with key data associated with the sampling. These key data are the alpha (if the EI is associated with a multiunit firm), sampling weight, stratum in which subjected to sampling, date of sampling, SIC, and the latest BMF activity. A select or nonselect code is present and, if selected, the panel and other information are present. As noted earlier, sampling is done independently for each of the retail, wholesale, and services areas in the initial sample selection. Thus, an EI which has establishments in retail and service, for example, would be sampled in each area, and the results of each sampling would be reflected on the EI Universe. In contrast, for birth sampling, each EI which has establishments in two or three different trade areas (retail, wholesale, or services) is subjected to sampling in the trade area which gives it the highest chance of selection, and, if selected, it is

added to each of the surveys in which it has establishments. However, it does not often happen that a birth EI has establishments in more than one inscope trade area.

The survey registers are the files which are used for mailing and tabulating the surveys. There is one register for each of retail, wholesale, and services. These files contain name and address, geographic codes, and activity codes which indicate whether to mail a report form or tabulate a sampling unit for a given month, sampling weights, SIC codes for tabulation, the tabulated data for the sampling units for the latest 14 months, along with other data necessary for control, editing, imputation, and tabulation. For retail, the survey register contains all the retail establishments associated with the sampling unit, and data are requested and tabulated for both the sampling unit and either all or a subsample of its establishments.

IV.B. Setting Up the EI Control File, the EI Universe, and the Survey Registers

The EI Control File contains all BMF active EI numbers of employer businesses in the United States as well as many inactive ones. A recent count showed about seven million BMF active and nine million BMF inactive EI records on the EI Control File. Key fields on the EI Control File are: EI, alpha (if a submaster EI), indication of source of the EI for the file (e.g., SSA birth, 941 payroll file), SIC, date payroll first reported (SSA birth month is shown if payroll data are not yet reported), 941 and other tax forms filing requirements (from which we obtain BMF activity), expected employment, largest of the latest four quarters of payroll, date of latest quarter of payroll, and sampling status code.

After the initial sample selection from the SSEL-based frame, a new EI Control File is created. This is accomplished by blanking all sampling status codes on a duplicate copy of the EI Control File used for the previous samples. Then each EI on this EI Control File is updated with a sampling status code which indicates whether that EI has been subjected to sampling. A blank status code indicates an EI has not been subjected to sampling. Following this update, the EI Control File reflects only the initial sampling results. Also EIs which have been determined to be out of scope but are BMF active are assigned a unique sampling status code. This is done to freeze the SIC coding to the time of the initial sample selection and to prevent an EI which was out of scope initially from having its SIC changed and then coming through the birth sampling. The freezing of the SIC codes for the period that the samples are used (usually five years) results from our inability to identify SIC changes in a way that would lead to less bias in our estimates than freezing SIC codes.

The EI Universe, the file of all EIs subjected to sampling, is also established on the basis of the initial sample selection. The EI Universe differs from the EI Control File in that it contains only EIs which have been subjected to sampling in the current business surveys, whereas the EI Control File includes all EI numbers. Following the most recent sample revision's initial EI Universe construction, there were about 64,000 EIs selected in the sampling (or associated with certainty alpha sampling units) in all three trade areas, and about 2,040,000 EIs subjected but not selected.

Setting up the survey registers consists of forming the records which contain the necessary data items for mailing and tabulating the surveys. For many alpha sampling units, it is often necessary to separate the sampling unit into two or more reporting units. These reporting units are set up to facilitate reporting by the firm and may consist of reports by SIC or by geographic areas or by subsidiaries. As an example, the 1600 retail certainty alphas expanded to about 1900 reporting units in the most recent sample revision.

Because of the need to publish detailed SIC and geographic area data for retail, all establishments associated with the selected sampling unit are contained on the register, and the tabulating units (units from which the data are summed to form the estimates) for retail are the establishments. Some reporting units report data for all their establishments (these are called full list reporters). Most with many establishments are asked to report for only a subsample of their establishments. In the Monthly Retail Trade Survey (MRTS), data are requested both for the reporting unit total and the selected establishments. For reporting units which report for a subsample of establishments, data are adjusted to weight up to the reporting unit total. For wholesale and services where subnational estimates are not published, most sampling units are reporting units, though some sampling units are separated into several reporting units based on SIC. Other reporting units are broken out into separate tabulating units because the reporting unit is relatively large and includes more than one SIC. Otherwise, tabulating units and reporting units are identical in the wholesale and services surveys.

For sampling units which are reselected into the new sample (as for example, almost all of the certainty alphas, many of the certainty EIs, and many of the larger noncertainty EIs) data are transferred from the previous register into the new sample register. This includes all reporting parts and their associated historical data. In the most recent revisions, about 20 percent of the sampling units were reselected and transferred directly from the old to the new registers.

The EI Control File, the EI Universe, and the survey registers have been described in their state after the initial sample selection. The following sections will describe operations which update these files.

V. Updating of Frame and Samples

The frame and samples are updated quarterly. The reasons are basically threefold: first, new EIs come into existence - these are called birth

EIs or, simply, births; secondly, existing EIs become BMF inactive or go out of business - these are called deaths; and finally, EIs that were BMF active and subjected to sampling, subsequently become deaths and then become BMF active again - these are called reactivations. Updating procedures, beginning with the backlog birth sampling and continuing with quarterly birth and death processing, allow the samples and the list frame to represent as large a part of the universe of retail, wholesale, and services businesses as one reasonably can. As stated earlier, the area sample provides the remaining coverage.

V.A. Backlog Birth Sampling

Following the initial creation of the EI Control File with the results of the initial sample selection, those EIs which have not been subjected to sampling, but are BMF active and have SIC codes included in the retail, wholesale, or services surveys and satisfy certain other conditions, are subjected to a double sampling procedure. BMF active EIs with a "blank" sampling status code are in this initial backlog birth sampling if they are inscope, and have nonzero payroll in any of the latest four quarters, or expected employment of three or more. In birth sampling, we sample the unclassified EIs for which an SSA birth record was received but SSA was unable to assign an SIC code. If no SSA birth record was received for an EI, but 941 payroll was reported for the current or prior quarter and four quarters or more ago, we also sample the EI as an unclassified case.

This identification of the birth backlog contains about a year's worth of births, or about 400,000 EIs inscope to the business surveys or unclassified. Many of these EIs are birth EIs which have not been added to the main SSEL files which are updated annually. This birth backlog would also include any EIs which were inactive as of the time of the sampling but subsequently became active, or active inscope EIs which were missed in the initial sampling for whatever reason. The sampling of the birth backlog is the first of many birth sampling operations which take place on a quarterly basis until the initial samples are replaced.

All EIs in the birth backlog are matched against the birth sampling results from the previous samples. Many of these EIs would have been sampled in the first phase of previous quarterly operations. All EIs in the birth backlog which were subjected to birth sampling in the past two years are represented by EIs from this group which have been selected in the first phase of sampling. This is to take advantage of the prior mailing and data received from the respondent at that time, thus saving cost and respondent burden. These first phase selected EIs are then prepared for the second phase of sampling, using the SIC and sales data collected in the earlier mailing. The EI Universe from the prior sample is the key file used to identify this portion of the birth backlog since it contains the record of the earlier birth samplings.

For EIs not sampled as births in the most recent two years, a first phase sampling is done. A first phase sampling size stratum is assigned to each EI based on its SIC and quarterly payroll or expected employment, whichever assigns the EI to the stratum with the largest probability of selection. These EIs are then sampled with an equal probability systematic sampling scheme within SIC by size stratum. The selected EIs are mailed a birth classification form which requests two recent months of sales or receipts, data to further refine or correct the SIC, and type of operation information. The monthly sales are then converted to a measure of size for sampling in the second phase.

The second phase of sampling for the birth backlog combines both groups of first phase selects: those mailed and those from previous quarterly birth samplings. Many first phase selected EIs are not subjected to the second phase of sampling. These are cases that were mailed as unclassified but based on their responses were determined to be out of scope, for example. Also, wholesale EIs which are determined to be nonmerchant are dropped prior to the second phase of sampling, as are EIs determined to be already out of business, or while BMF active, have not yet started business. Using the appropriate SIC information and a pseudo sales measure of size based on the reported monthly sales data, the second phase of sampling is completed. It is a probability proportional to size systematic sample conducted over time (from quarter to quarter). This systematic sampling results in the same probability of selection that the EI would have had in the initial sample selection. More discussion of the birth sampling is contained in Wolter et al. (1976). Following the second phase sample selection the newly selected EIs are added to their appropriate survey register.

A BMF activity update is performed on the survey registers prior to the initial sample mailouts. This is necessary since the time lapse between the extractions from the SSEL and the start of survey canvassing is about one year. The BMF activity update is usually accomplished using the results of the second EI Control File update for the new samples following the birth backlog operations. Through this update, BMF inactive noncertainty EIs are dropped from the samples as deaths.

V.B. Quarterly Birth and Death Processing

On a quarterly basis beginning with the backlog birth operations, the EI Control File is updated with the quarterly 941 file, the BMF activity file, three months of SSA births, and a feedback file from the most recent first and second phase birth sampling. This feedback file updates the EI Control File with sampling status codes and SIC information from the birth classification form mailing. Just as in the birth backlog identification, inscope EIs with blank sampling status codes which satisfy the payroll or expected employment conditions mentioned earlier are subjected to the first phase of sampling. Each EI to be sampled for this

quarter is assigned a unique sampling status code to identify this most recent quarter's group. (This code is needed for the area sample processing discussed in section VII.) Also, each out of scope EI which is BMF active as of the update is assigned a nonblank sampling status code as part of the control over SIC changes.

The double sampling procedure for birth EIs is similar to birth backlog sampling. However, there is no corresponding match to the EI Universe or use of any prior response data for these birth EIs. Following the first phase sampling, mailing, and data collection, if an EI is active but is unable or refuses to provide sales data, a measure of size for sampling is imputed based on the EI's payroll, or lacking payroll, on the expected employment. If an EI is mailed but is a total nonresponse in the data collection phase, has not yet begun business, or has otherwise not provided information for sampling, the EI is retained on an unsampled first phase birth register. These EIs will be remailed after six months if they remain BMF active and have recently reported quarterly payroll.

Following the quarterly first and second phase sampling, feedback files are prepared for the succeeding quarter's EI Control File update, and the EI Universe is updated with the results of sampling. When updating the EI Universe, changes in BMF activity obtained from the EI Control File generate corrections to the survey registers. As part of the quarterly survey control operations, second phase selected birth sampling units and reactivations are added to the survey registers while noncertainty sampling unit deaths which have become BMF inactive are dropped. Certainty EI sampling units do not rely on the BMF activity for determining a death or a reactivation. Instead extensive follow-up of the certainty EI unit is done to ascertain the activity status and to pick up a certainty successor EI unit, if possible. This is discussed further in section VI which follows. A reactivation is defined to be a noncertainty EI sampling unit which was selected into the sample, went BMF inactive and was dropped from mailing and tabulations, then became BMF active again. For the current samples, the number of births and reactivations added generally is about the same as the number of sampling units dropped as inactive or deaths. Thus the overall sample sizes remain fairly constant.

VI. Status Changes for Selected Sampling Units

There are various natural and frequent changes in the business universe which affect the list sample frame and selected EI or alpha sampling units. Some of these events include the start of a new firm (birth), the going out of business (death) of an existing firm, the adding or dropping of establishments, or the merging of two firms. We have already seen that births are identified through the quarterly birth sampling operations and that deaths for selected noncertainty sampling units are determined by their BMF activity. For the other changes in status, rules must be established in order to maintain the list sample frame and minimize deterioration of the selected samples. These "status change rules" as they are called are different for certainty and noncertainty units. A principal reason for these differences is that dependence on the area sample to represent the large certainty firm status changes could seriously increase the variances of the estimates; consequently, there is an extensive follow-up of all certainty unit changes. Moreover, the high cost of following up on all sampling units leads us to depend rather on the BMF activity for noncertainty sampling units.

We will first discuss the status change rules for certainty sampling units, alphas, or EIs. A selected certainty sampling unit is instructed to report for all active establishments, and only active establishments which belong to the unit for the period of the survey. Thus data from newly added establishments are to be included in the report, while those from sold or closed establishments are not. For certainty sampling units, the status change rules depend upon following up on detected changes and making appropriate adjustments to the sampling units on the survey registers. Next we consider the various status changes and their accompanying rules.

Status change: All establishments of a certainty alpha or a certainty EI sampling unit are bought out or absorbed through a merger into another firm. Rule: The successor firm itself becomes a certainty alpha (if an alpha was absorbed) or a certainty EI (if an EI was absorbed), and if it was not already a certainty alpha. The new certainty EI or alpha is then to report for all the newly acquired establishments as well as those which had previously belonged to the successor firm. If the successor is made a certainty alpha, any prior EI sampling units selected into the sample for this firm are dropped from the sample since they are now to be covered by the certainty alpha sampling unit. The predecessor alpha or EI is treated as a death and dropped from the sample.

Status change: A certainty alpha or a certainty EI goes out of business with no successor. Rule: The alpha or EI is treated as a death and dropped out of the sample with no successor.

Status change: A certainty alpha or a certainty EI divests some but not all of its establishments, or while divesting all of its establishments, the establishments are taken over by two or more firms. Rule: (This rule depends on the trade area.) For retail, if the certainty firm or EI had 2 to 20 establishments in the previous month, and more than half of them have been divested to a single EI, convert this successor EI to certainty EI status. If the certainty alpha or EI has more than 20 establishments in the previous month and 11 or more establishments are divested, convert to certainty EI status any EI which purchased 11 or more establishments. Rules similar to these for retail apply for wholesale and services.

Noncertainty sampling units are likewise instructed to report for all active establishments and only active establishments associated

with the selected EI. In addition, the following situations and corresponding rules apply.

Status change: All establishments of a noncertainty EI sampling unit are bought out or absorbed into another firm. Rule: In retail, this change in EI number of a selected sampling unit results in the tabulation of its data at zero value for the months following the EI change. This is done to prevent duplicate tabulation of the data from the old EI's establishments. These data are represented by the area sample in the case where the new EI is not subjected to sampling or is BMF inactive, or by the list sample if the EI is subjected to sampling and BMF active. The old EI is dropped as a death when it goes BMF inactive. In wholesale, a noncertainty firm which indicates a change in EI number is retained and its data tabulated at a nonzero value (imputing if necessary) until the EI goes BMF inactive. Since there is no area sample to represent wholesale birth EIs, this upward biased procedure is used to offset the downward bias caused by the lag in representing new birth EIs in the birth processing operations. These biases are both small, and some investigations have shown that they are approximately equal in size.

Status change: A noncertainty EI goes out of business with no successor. Rule: For retail, the EI is tabulated at zero value until it goes BMF inactive. For wholesale the EI is imputed at a nonzero value to compensate for the birth lag until it goes BMF inactive.

Status change: A noncertainty EI divests some but not all of its establishments. Rule: No action is taken to follow up for successors. The selected EI reports only for the retained establishments.

Status change: A noncertainty EI divests all of its establishments into two or more successor EIs. Rule: For retail, the old EI is tabulated at zero until it goes BMF inactive. For wholesale, the old EI has nonzero data imputed for it until it goes BMF inactive.

The preceding rules apply for the Monthly Retail Trade Survey (MRTS) and Monthly Wholesale Trade Survey (MWTS) surveys. Somewhat different rules hold for the Monthly Advance Retail Trade Survey (MARTS) and the Monthly Retail Inventory Survey (MRIS) because they are fixed subsamples of the MRTS which do not use a quarterly birth and death procedure but rely on ratio estimates of sampling units which are not deaths. For these smaller samples, the certainty status change rules are similar to those of the MRTS and the noncertainty status change rules rely on follow-up procedures similar to those for the certainty units. For the annual retail and wholesale surveys, the status change rules are applied along lines similar to those of the monthly surveys. For the annual survey for services, rules similar to those for retail are used for those services SICs which use a supplementary area sample. For services SICs which do not use a supplementary area sample (such as the medical and legal services SICs) the rules employed are analogous to those used for wholesale.

One other aspect of survey control which is related to and may be considered as a status change deals with the problem of SIC changes and incorrect SICs of selected sampling units. When it is ascertained that an SIC is changed or is incorrect, a noncertainty EI sampling unit is requested to continue to report and is tabulated in the SIC in which it was selected. If the respondent objects to the reporting, the EI is placed on "reserve scope" status, held out from the mailing, and imputed in the tabulations. Note that these changes often are undetected since SIC information is not asked on the surveys' questionnaires. The use of the reserve scope concept is needed to reduce any bias in the estimates caused by not having subjected to sampling inscope EIs that were incorrectly assigned an out of scope SIC code or out of scope EIs that became inscope. In effect, we freeze the SIC for selected noncertainty sampling units. This is a companion operation to freezing the out of scope SICs for EIs on the sampling frame. Also, as part of controlling the SIC misclassification, in the birth sampling operations, an EI which is first phase selected with an inscope SIC is second phase sampled in that inscope SIC if its correct SIC is determined to be out of scope. If the EI is selected in the second phase of sampling, it becomes a "reserve scope" case if it is not willing to report in the survey. This reserve scope rule also applies to certainty sampling units but in this case somewhat more latitude on changes is allowed. In effect, inscope SIC changes are usually made for certainty sampling units. The reserve scope cases account for approximately 1 1/2 percent of the MRTS total U.S. sales estimate, about 5 percent of the MWTS sales, and about 5 percent of the MWTS inventory estimates for the total U.S.

VII. Area Sample Processing and Unduplicating the List Frame and Area Sample

As stated earlier, the area sample covers retail and services nonemployer businesses and new employer businesses (birth employers). Nonemployer businesses are not found on the universe lists from which the list samples are selected because of the expense and technical problems associated with maintaining a list of a large number of very small businesses, many of which are in operation for only a short time. Even though the universe lists are continuously augmented with birth employers, there is a lag in representing new or birth employer businesses by the list samples for the following reasons: first, new employers may be open for business before filing forms that would enter the businesses into the administrative or tax system; secondly, the time it takes for the business record to be processed and sent to the Census Bureau; and finally, the time it takes for the Census Bureau to conduct its own double sampling procedure that leads to the representation of the birth employers in the list sample.

In recent years, the area sample has accounted for from 5 to 10 percent of the monthly retail sales estimate for the United States as produced by the Monthly Retail Trade Survey.

The estimated sales for nonemployers are about 3 percent of the total sales estimate, varying between 2 1/2 and 4 percent. The birth employers in the area sample account for from 3 to 7 percent of the total sales estimate, depending on the length of the delay in representing the births in the list universe. As the delay increases, the percentage of sales represented by the births in the area sample likewise increases. The percentage contributions are similar for the services receipts estimates produced by the Services Annual Survey, except that nonemployers account for a larger percentage and birth employers a smaller percentage.

No area sample is used for the wholesale trade surveys for three reasons: first, there are relatively few wholesale birth EIs; secondly, nonemployers are out of scope to the wholesale trade surveys; and finally, the same area sample which is used for retail and services is not suitable for wholesale because of the different physical clustering characteristics which would lead to large sampling errors for wholesale estimates.

The area sample of retail stores and selected services establishments is being redrawn, since many of the sample segments were last drawn about twenty years ago for the largest metropolitan areas and more than thirty years ago for the rest of the country. This sample selection work, as well as the area sample itself, is described in Isaki et al. (1981).

Monthly surveys, such as the MRTS, which rely on both a list frame and an area sample and which maintain these over a five-year period, must have survey control procedures. Data from retail establishments located in the area sample segments, which are not associated with alphas in the sample and not associated with EIs subjected to sampling, are tabulated in the MRTS estimates. The EI Control File plays a key role in determining which establishments are tabulated. This file, updated on a quarterly basis with the latest results of birth sampling, BMF activity and other data, contains all EIs which have been subjected to sampling in any of the current business surveys as well as an indication of the latest BMF activity of these EIs. Thus, in practice, if an EI of an area sample establishment is found to be BMF active and subjected to sampling on the proper quarter's EI Control File, then the sales of that establishment would not be tabulated for the MRTS estimates. However, if the establishment's EI were not on the EI Control File, or if, while on the file, were either BMF inactive or not subjected to sampling, the sales, properly weighted, would be tabulated for the MRTS estimates. Of course, data from retail establishments determined to be nonemployers would be tabulated in the MRTS estimates. Similar use is made of the EI Control File for processing the SAS.

References

Daly, J., Kailin, H., Woodruff, R., and Shor, M. (1972), "Some Experience in the Large Scale Use of Administrative Lists in a Monthly Business Survey," The Statistician, Vol. 22, No. 1.

Impett, L. and Monsour, N. (1982), "Sample Reselection of the Census Bureau's Current Business Surveys," Proceedings of the American Statistical Association, Business and Economic Statistics Section, 245-250.

Isaki, C., Konschnik, C., and Monsour, N. (1981), "Reselection of an Area Sample for the Retail and Service Surveys," Proceedings of the American Statistical Association, Section on Survey Research Methods, 116-121.

Isaki, C., Wolter, K., Sturdevant, T., Monsour, N., and Trager, M. (1976), "Sample Redesign Aspects of the Census Bureau's Monthly Business Surveys," Proceedings of the American Statistical Association, Business and Economic Statistics Section, 90-98.

Office of Federal Statistical Policy and Standards (1977), Standard Industrial Classification Manual, 1977 Supplement. Washington, D.C.: Department of Commerce.

Office of Management and Budget (1972), Standard Industrial Classification Manual. Statistical Policy Division. Washington, D.C.: Office of Management and Budget.

U.S. Bureau of the Census (1979), Standard Statistical Establishment List Program, Technical Paper 44, Washington, D.C.

U.S. Bureau of the Census (1982), 1982 Industry and Product Classification Manual. Washington, D.C.: Department of Commerce.

Wolter, K., Isaki, C., Sturdevant, T., Monsour, N., and Mayes, F. (1976), "Sample Selection and Estimation Aspects of the Census Bureau's Monthly Business Surveys," Proceedings of the American Statistical Association, Business and Economic Statistics Section, 99-109.