

Frame Construction and Sample Maintenance for Current Economic Surveys¹

Katrina Washington, James Burton, and Ruth Detlefsen
Service Sector Statistics Division, U.S. Census Bureau, Washington, DC 20233

Abstract

The Service Sector Statistics Division of the U.S. Census Bureau conducts seven surveys on an annual, quarterly, or monthly basis. These surveys measure the nation's economic activity of the retail, wholesale, and service sectors. Three of these surveys are principal economic indicators. Two of these surveys contribute to a fourth economic indicator. To produce quality data about these industries in a reliable, timely and relevant manner, we must use a reliable sampling frame and have in place quality survey control procedures. This paper will provide an overview of the surveys, describe in detail the initial frame construction and sample selection for these surveys, discuss how the frame and samples are maintained and periodically updated, and discuss the quinquennial sample revision activities for these surveys, following each Economic Census.

Key words: frame construction, sample selection, frame maintenance, sample maintenance

1. Introduction

The U.S. Census Bureau conducts several business surveys that measure the nation's economic activity of the retail, wholesale, and service sectors on an annual, quarterly, or monthly basis. They are the Monthly Retail Trade Survey (MRTS), the Advance Monthly Sales for Retail and Food Services Survey (MARTS), the Monthly Wholesale Trade Survey (MWTS), the Quarterly Services Survey (QSS), the Annual Retail Trade Survey (ARTS), the Annual Wholesale Trade Survey (AWTS), and the Service Annual Survey (SAS). The MARTS, the MWTS, and the QSS serve as principal economic indicators. The MARTS provides an early indication of sales of retail and food service companies for the previous month. The data are released approximately nine days after the close of the reference month. The MWTS provides monthly estimates of sales and inventories of merchant wholesalers, except for manufacturers' sales branches and offices. The data are released six weeks after the close of the reference month. The QSS provides quarterly revenue and expenses for selected service industries. The data are released 75 days after the end of the reference quarter. A fourth economic indicator, the Manufacturing and Trade Inventories and Sales (MTIS) report is partially based on data from both the MWTS and the MRTS. While not designated as a principal economic indicator, data collected in the MRTS contribute to both the MTIS and the MARTS. The MRTS provides monthly estimates of sales at retail and food services stores and inventories held by retail stores. The data are released approximately six weeks after the end of the reference month.

¹ This report is released to inform interested parties of ongoing research and to encourage discussion. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

The ARTS provides detailed industry estimates of annual sales, sales from e-commerce, per capita sales, gross margins, year-end inventories held inside and outside the U.S., merchandise purchased, total operating expenses, gross margin/sales ratios, and accounts receivable balances. The AWTS provides detailed industry estimates on annual sales, year-end inventories held inside and outside the U.S., purchases, total operating expenses and e-commerce data for merchant wholesalers and manufacturers' sales branches and offices, as well as sales, commissions, and operating expenses data for agents, brokers, and electronic markets. In survey years ending in 2 and 7, both the ARTS and AWTS also provide detailed operating expenses. The SAS provides estimates of total revenue and total expenses for sector, sub-sector, and service industry groups, as well as industry specific data such as detailed revenue and detailed expenses for select industries.

E-commerce data provided by these business surveys is combined with e-commerce data provided by the U.S. Census Bureau's manufacturing surveys and is published in a combined E-Stats report, released in the spring of each year.

2. Frame Construction

Samples for the ARTS, the AWTS, and the SAS are selected approximately once every five years. The frame, from which the samples for the annual surveys are drawn, is developed using data extracted from the most recent Economic Census (collected every five years, in years ending in 2 and 7) and from information extracted from the Census Bureau's Business Register. We periodically reselect the samples in an effort to:

- Ensure each sample is representative of its target population
- Improve the efficiency of the sample
- Expand industry coverage of the survey
- Incorporate updates to the industry classification structure, and
- Redistribute burden for small and medium sized businesses.

Samples for the MRTS, the MWTS, and the QSS are selected at the same time the samples for the annual surveys are selected, as subsamples of their related annual survey's sample. The sample for the MARTS is selected approximately once every 2 ½ to 3 years as a subsample of the MRTS.

2.1 Universe Extraction

The initial step in developing the sampling frame is universe extraction. Universe extraction is the process used to create a list of business establishments from which the sampling frame will be created. This list of establishments is extracted from the Census Bureau's Business Register.

2.1.1 The Business Register

The Business Register is a multi-relational database that contains a record for each known establishment that is located in the United States or one of its territories and has paid employees. An establishment is defined as an economic unit, usually at a single physical location, where business transactions take place and for which payroll and employment records are kept. There are two types of establishments - singleunits and multiunits. A singleunit establishment is an establishment whose company entity owns or operates only that one establishment. A multiunit establishment is an establishment whose company entity owns or operates two or more establishments. The structure of an

establishment's primary identifier on the Business Register differs depending on whether it is a singleunit or a multiunit.

A singleunit's primary identifier is its Employer Identification Number (EIN). The Internal Revenue Service (IRS) issues the EIN for the establishment to use in reporting social security payments for its employees under the Federal Insurance Contributions Act (FICA). Because a singleunit company has only one establishment, there is a one-to-one relationship between the company and the EIN. Thus the company, the EIN, and the establishment all reference the same physical location and all three terms can be used interchangeably and unambiguously when referring to a singleunit company.

For multiunit establishments however, a different structure connects the company with its establishments via the EIN. A multiunit company is associated with a cluster of one or more EINs and each EIN is associated with one or more establishments. Therefore, a multiunit establishment is an establishment that is associated with a company that has 2 or more establishments, and one or more EINs. Each company is associated with at least one EIN and only one company can use a given EIN. However, one company may have several EINs. Similarly, there is a one-to-many relationship between EINs and establishments. Each EIN can be associated with many establishments but each establishment is associated with only one EIN. Because of the possibility of one-to-many relationships, we must distinguish between the company, its EINs, and its establishments. Thus, a multiunit establishment's primary identifier is a unique identifier number called the employer unit identification number, not the EIN.

To ensure that the list of business establishments contains the most up-to-date information for each establishment, the Business Register is periodically updated with data from administrative records from the IRS, the Social Security Administration (SSA), and the Company Organization Survey (COS). The universe extraction process also extracts data for each establishment from the most recent Economic Census.

Establishments are extracted based on whether the establishment is determined to be in-scope or out-of-scope to the aforementioned surveys. The scope of each establishment is based on characteristics such as its North American Industrial Classification System (NAICS) code, geographic code, business activity code, employment code, tax status, and wholesale type of operation code (TOC). The TOC identifies a wholesale establishment as a distributor, manufacturers' sales branch or office (MSBO), or agent or broker. All establishments considered to be out-of-scope are omitted from further extraction processing. Once establishments are determined to be in-scope, we extract sales, payroll, employment, identifying information such as name and address, and any associated EINs for multiunit establishments.

2.1.2 Measure of Size

In order to later assign sampling units to size based sampling strata, we next calculate a sales/revenue measure of size (MOS) and an inventory MOS for retail and wholesale, for each establishment. Multiple measures of size for sales are calculated for each establishment based on its most recent administrative payroll and/or revenue available from the Business Register, as well as payroll and/or revenue from the most recent Economic Census. Multiple measures of size for inventory are calculated for each establishment based on the most recent administrative inventory and/or revenue data, economic census inventory data, and stock-to-sales ratio data extracted from other

surveys. Best sales and inventory measures of size are determined, using methodology described in King (2005) to rank and compare the various measures calculated. Each chosen MOS is then assigned to the given establishment.

2.1.3 Edits

Once a sales (and inventory, in some cases) MOS is assigned to each establishment, editing is done at both the establishment level and at the sampling unit level to identify inconsistencies. The sampling unit level is either the Employer Identification Number (EIN) or company level, which are the levels from which samples are actually drawn.

2.1.3.1 Establishment Edits

Establishment edits are performed on both administrative and census payroll data, census sales data, the ratio of census sales data to the assigned sales MOS, the ratio of census inventory to census sales for wholesale, the assigned MOS to the remaining calculated measures of size, and other correlated data for specific industries. Parameters used in the edits are researched and developed prior to extraction. For the 2006 Business Sample Revision Universe (BSR-06) Extraction, there were initially over 3.8 million singleunit establishments with approximately 89,000 edit failures and over 1.6 million multiunit establishments with approximately 35,000 edit failures. Thus, approximately 2% of all establishments failed the establishment edits; however, this 2% was equivalent to over 100,000 establishments that the analysts needed to review.

2.1.3.2 Sampling Unit Edits

Sampling unit edits are used to identify sampling units that are adversely affecting the sample size. Due to the large number of edit failures produced by the establishment edits, and the limited amount of time allocated to review and correct the failures, sampling unit edits were put in place as an additional tool for survey analysts to use to determine which problem areas to direct their focus on. This methodology was introduced in the 2006 Business Sample Revision as described in King (2007).

Before the edits can be performed, the sampling units must be created. Section 3.1 details the creation of the sampling units, which is performed initially during universe extraction (and again during sampling) for the purpose of implementing the sampling unit edits. Once the sampling units are created, they are stratified as discussed in Section 3.2.

For noncertainty EINs, the generalized stratification program is used to determine the sample size for each sampling recode. Then, the sample size is recalculated after removing one EIN. This difference in sample size is a measure of the effect of that EIN on the sample for that sampling recode. The EIN is replaced and the edit is repeated for each noncertainty EIN in a given sampling recode.

For certainty companies, the Hidioglou-Berthelot edit to detect outliers is performed on the ratio of the product of administrative payroll and a regression factor (β) to the certainty unit's MOS. The regression factor used is a factor calculated from regressing Census revenue onto Census payroll by sample recode. The edit identifies certainty companies whose data exhibit the greatest differences between the numerator and denominator variables.

For both certainties and noncertainties, edit failures are prioritized. The larger the effect for noncertainty edit failures, and the greater the difference between the numerator and denominator variables for certainty edit failures, the higher the priority.

Survey analysts review sampling units based on edit failure priority. The higher the priority, the more important it is that the unit be reviewed. Once a sampling unit is identified for review, then the establishments that are part of that sampling unit should be reviewed individually as well, as needed. This method of review allows survey analysts to focus their attention on the units affecting the sample and reduces the number of establishments for them to review in the limited time allotted.

During the review period, survey analysts are allowed to update, modify or revise data for a given establishment that will affect sampling, such as payroll or revenue data used in the calculation of the MOS, or classification information such as NAICS code or TOC. Once the review period is complete, the edited universe of inscope establishments is used to construct the sampling frame from which the survey samples will be drawn.

3. Sampling

Once the sampling frame is developed, the initial sample selection process begins. To ensure each survey's CV constraints are properly met and that representative samples for each trade area are chosen, sample selection is done independently for wholesale, retail, and services.

3.1 Sampling Frame Creation

Sampling is done at the EIN and company levels. Inscope multiunit establishment level records are aggregated to the EIN and company levels to create EIN and company summary records. Each EIN and company record is then assigned a major kind of business (KB) based on the breakout of that company or EIN's total sales by NAICS. No aggregation is necessary to put singleunit establishment information on an EIN basis or a company basis. Thus, the sampling units created for singleunit companies simultaneously represent establishment, EIN, and company information, and are contained on both the EIN summary data sets and the company summary data sets.

These EIN and company summary records are then considered sampling unit records. They are assigned sampling recodes based on their major KB, tax status, and TOC. The sampling recode will be used as the primary stratification variable. Certainty status is also determined at this time. A sampling unit (EIN or company) is considered a certainty for its major KB if it has an aggregated MOS for any KB in which it operates that is greater than a predetermined cutoff for that KB. Survey analysts can also request for other companies, which do not meet the predetermined cutoff, to be certainties. EINs associated with a certainty company are removed from the noncertainty frame. All other EINs and companies are then considered noncertainties, and all noncertainty records are EIN level records. Therefore, if a company does not meet the certainty cutoffs for any of its industries, then each of its EINs are included individually on the sampling frame as noncertainty sampling units with a major KB assigned to each EIN. The sampling frame is then ready to be stratified.

3.2 Stratification

The primary stratification of the sampling frame is by industry group (i.e., sampling recode) based on the detail required for publication. For some service industries, tax status (taxable or tax-exempt) and government ownership (government or non-government) are also taken into consideration when defining the primary strata. TOC is also used to stratify the AWTS sampling frame. The primary strata are substratified into 4 or more substrata based on estimated annual sales. The largest sales size stratum within each primary stratum consists of companies whose MOS are above the certainty cutoff for that particular recode, all of which are selected with certainty, as described in section 3.1. All firms not selected with certainty are subjected to sampling on an EIN basis. For noncertainty EINs, a generalized stratification program is then used to determine the sample size for each sampling recode based on predetermined coefficient of variation (CV) constraints.

3.3 Sample Selection

The first step in the sample selection identifies firms selected with certainty. For retail and wholesale, any company selected with certainty into the annual sample is also included in the monthly sample with certainty. However, since the industry stratification for the QSS sample is broader than the industry stratification used for the SAS sample, then the stratum assignments, including certainty/noncertainty status, may differ between the two surveys.

Once the noncertainty EINs are stratified and the sample size determined, a simple random sample of EINs without replacement is selected for retail and wholesale, within each noncertainty stratum. The selected noncertainty EINs are assigned to one of two groups. One group is canvassed for both the monthly and annual surveys; the other group is canvassed only for the annual survey. For wholesale, the Manufacturer's Sales Branches and Offices, and Agents and Brokers samples are only canvassed in the annual survey. However, EINs selected into the Manufacturers' Sales Branch and Office or Agent and Broker sample are included in the monthly sample if that EIN had activity in scope to the monthly survey.

The SAS noncertainty sample is selected using the same method as retail and wholesale for only one group. Using the selected SAS sample, QSS sample selection is done independently within each size stratum using a systematic probability-proportional-to-size sampling procedure where the size used is the SAS sampling weight. It is possible that we select some units in the SAS sample at a lower sampling rate than desired for the Quarterly Services Survey sample. We include such a unit into the QSS sample with the SAS sampling weight.

We should note here that we exclude from the noncertainty sample selection operations particular noncertainty sampling units that were selected in the previous sample in an effort to redistribute the reporting burden for small and medium-sized businesses. We call this process controlled nonselection.

The MARTS sample is selected from the MRTS sample. The MARTS units are stratified by broader industry categories and substratified by monthly sales as measured in the MRTS. Sample sizes are calculated to meet hypothetical reliability constraints on estimated monthly sales totals for specified industries. Sample selection is done independently within each size stratum using a systematic probability-proportional-to-size procedure where the size used is the MRTS sampling weight. The MARTS sample

does not undergo sample maintenance procedures, as described in Section 5. Instead, the sample is redesigned for and reselected approximately every 2½ to 3 years.

For all other sample surveys discussed in this paper, a new sample is drawn approximately every five years. However, throughout the duration of the initial sample selection and ongoing canvassing operations, the makeup of the business world is ever changing. In an effort to keep up with these changes and maintain representative samples throughout, updates are made to the samples on a quarterly basis. These processes are part of our Sample Maintenance operations, outlined in Section 5. Note also that ongoing survey processes for the previous business sample revision (BSR) continue while frame creation and sample selection activities are taking place for the new sample.

4. Survey Control Files

In order to keep both an accurate record of the initial disposition of each company or EIN sampling unit during sampling frame development and sample selection, to maintain a record of changes to both the sampling frame and samples throughout their lifespan, and to ensure that the samples properly represent the target population at any time estimates are made, several databases and database tables are set up and updated throughout the life of the samples. These databases and database tables are collectively referred to as the survey control files. These files include the Sampling History Table, the EIN Universe, and the survey register files. The birth backlog, quarterly birth, annual sample update, and status change processes outlined in Section 5 are all processes used to update the survey control files.

4.1 Sample History Table

The sample history table resides on the Business Register and is principally used by the birth backlog and quarterly birth processes to identify EINs not already included in the sample. This table contains a sampling status code to indicate, for each EIN on the Business Register, how that EIN was represented in the initial (or birth) sampling process. EINs are assigned a sampling status code to indicate if the EIN was deemed out of scope for our surveys, excluded from the sampling process by controlled non-selection, or subjected to sampling.

Once the initial sample is selected for each survey, the results of the Universe Extraction and initial sampling processes are used to update the Business Register with the initial sampling status codes.

4.2 EIN Universe

The EIN universe is a repository that contains the sampling history for each EIN subjected to sampling. While the sampling history table contains limited information on all EINs on the Business Register, the EIN universe contains detailed information on each EIN subjected to sampling for the current sample revision. This information includes key data such as when the EIN was subjected to sampling, in which survey it was subjected, if it was selected, the classification information used to assign the EIN to a stratum, which stratum the EIN was sampled in, and the sampling weight assigned to the EIN.

The EIN universe is updated on an ongoing basis with sampling information from the quarterly birth process, with administrative updates from the Business Register, and with status updates from the survey registers. The EIN universe also includes an interactive system for making status updates. The EIN universe is initially created after the initial

sample selection. The initial universe consists of two parts, the select EIN universe and the nonselect EIN universe. The initial select EIN universe includes EINs selected into the sample and EINs associated with selected certainty companies. The initial nonselect EIN universe includes any noncertainty EINs subjected but not selected into the initial sample.

Following the birth backlog and quarterly birth processing operations outlined in Section 5, the nonselect EIN universe is updated to include EINs subjected but not selected in first phase birth sampling operations carried out during those processes as well.

The select universe is also updated to include:

- EINs subjected and selected in first phase birth sampling operations for the quarterly birth and birth backlog processes,
- Additional second phase birth sampling information on EINs selected in first phase birth sampling and subjected to second phase birth sampling,
- Changes to business activity, such as going out of business or reactivation, for EINs already on the select universe,
- Additional EINs not yet on the Business Register but found by analysts to be associated with in-sample certainty companies, and
- Other EIN updates as allowed by our status change rules.

4.3 Survey Registers

Survey registers are survey-specific files used for survey processing. These files contain both control information, such as name, address, NAICS code, sampling recode, and sampling weights; and data, such as survey items, data values for those items, and data flags associated with each value. The survey register contains all necessary data needed for mailing, data collection, analyst review, editing, imputation, and tabulation.

New survey registers are created after the initial sample selection. Data for sampling units reselected into a new sample is transferred from the previous survey register to the new register. Records on each survey register are usually at the sampling unit or reporting unit level. Reporting units are set up to facilitate reporting by a company, and may consist of reporting units set up by NAICS level, geographic area, or by the company's subsidiaries.

5. Sample Maintenance

In an effort to ensure that our samples properly represent the target population for any time for which estimates are produced, the samples, and control files, are updated quarterly to reflect the changes taking place in the universe. These updates are done so that the samples can reflect new EINs or companies that come into existence (referred to as births), existing EINs or companies that become inactive or go out of business (referred to as deaths), EINs or companies or parts of companies that change ownership or association, and EINs or companies that go out of business and then become active again (referred to as reactivations). The Service Monthly and Annual Sample Maintenance (SMASM) system is designed to identify, track and apply these changes, based mostly on administrative updates to the Business Register, and then eventually applied to the current samples and survey control files.

5.1 Quarterly Birth Process

Periodically throughout the life of the samples, we update the samples to represent new EINs appearing on the Business Register. These new EINs, called births, are EINs recently assigned by the IRS on the latest available IRS mailing list for FICA taxpayers and assigned an industry classification (if possible) by the SSA. EIN births are sampled on a quarterly basis using a two-phase selection procedure. To be eligible for selection, a birth must either have no industry classification or be classified in an industry within the scope of the AWTS, the ARTS, or the SAS, and meet certain criteria regarding its quarterly payroll.

In the first phase, births are stratified by broad industry groups and then by a measure of size based on quarterly payroll. A relatively large sample is selected using equal probability systematic sampling. The selected births are contacted to obtain a more reliable measure of size, based on sales in two recent months, a new or more detailed industry classification code, and company affiliation information. Using this more reliable information, the births that were sampled during the first phase are then subjected to a second phase probability-proportional-to-size sampling with overall probabilities equivalent to those used in drawing the initial sample for each annual survey. Because of the time it takes for a new employer firm to acquire an EIN from the IRS, and for us to accomplish the two-phase birth selection process, births are added to the samples approximately four to seven months after they begin operation. Because of the proximity of birth selection to the start date of operations for most EIN births, it is assumed that the majority of these birth EINs are singleunit EINs. The table below shows the number of EIN births for each phase of the quarterly process for eight quarters of the most recent sample revision.

Quarter	Extracted Births	First Phase Selects	Second Phase Selects
2008Q4	115,779	11,657	1,966
2009Q1	93,826	8,093	1,285
2009Q2	84,043	7,339	1,214
2009Q3	263,579	30,372	4,581
2009Q4	109,567	11,312	1,951
2010Q1	122,842	14,025	2,373
2010Q2	84,530	7,293	1,221

For information on how current surveys control for EIN and company sampling units deaths or reactivations, see Section 5.4, Status Changes.

Births are added to the annual survey samples in February, May, and August of the calendar year following the survey reference year. Births that are selected in the quarterly birth selection process in November of the survey reference year are added to the samples to be included in the initial mailing of the annual survey questionnaires in January of the following year.

For the monthly surveys, births are not represented in the surveys until they go through the two-phase selection process. Similar to the annual surveys, births are added to the

monthly surveys in February, May, August, and November. At the same time, deaths are removed from the monthly surveys. To minimize the effect of births and deaths on the month-to-month change estimates, these changes are phased into and out of the survey by incrementally increasing the sampling weights of the births and decreasing the sampling weights of the deaths simultaneously. For the first month that a birth is included in the sample (i.e., February, May, August, or November), they are tabulated at one-third their sampling weight, while the new deaths for that month are tabulated at two-thirds their sampling weight. For the following reference month, the same births are tabulated at two-thirds their sampling weight and the deaths at one-third their sampling weight. For the third reference month for these EINs, the births are tabulated at their full sampling weight and the deaths are dropped (i.e., sampling weight equal zero). This process is repeated in each of the other quarters.

Births are added to the QSS in the same manner, and at the same time intervals as the annual surveys. The births selected for the QSS sample are a subsample of the births selected for the SAS. They are selected using probability-proportional-to-size sampling with overall stratum probabilities equivalent to those used in drawing the initial QSS sample.

Each quarter, there exist some EINs whose classification information is insufficient to assign those EINs to sampling strata during second phase. These EINs are stored on the unsampled EINs file during each quarterly processing cycle. The unsampled EIN file is updated each quarter during the SMASM process with EINs selected in first phase which do not meet conditions to be included in second phase processing. Each quarter, the file of unsampled EINs is updated with administrative data and some unsampled EINs are remained for classification, measure of size, and company affiliation information. These EINs are then checked to see if they are now eligible for second phase. If eligible, they are then subjected to second phase birth sampling.

5.2 Birth Backlog

The birth backlog is a special quarterly process used to identify all new employer EINs and business activity changes for EIN sampling units selected into the samples between the frame construction and the initial mail of the surveys. Since the birth backlog covers several quarters of business activity, there are some features that are part of the birth backlog operations but are not part of the usual quarterly processing operations.

To meet the scope requirements for the birth backlog, an EIN must be geographically inscope, have NAICS classification that is either a full NAICS code and inscope, a partial NAICS code and inscope, or be unclassified, and be an active business and have some payroll activity in one of the two years canvassed by the annual surveys using the new sample. An exception to this requirement are some currently inactive EINs that showed payroll activity in one of the two years canvassed, but meet all of the requirements to have been included on the initial BSR sampling frame. These are included in the birth backlog, and are called ‘special inactives’.

Because of the volume of new employer EINs that meet the birth backlog criteria and because we know that many of the new employer EINs that are identified as inscope during the birth backlog are EINs that have already had a chance for selection in either

the initial or a birth sampling operation in the previous BSR, as a cost savings measure, efforts are taken to make optimal use of the information already gathered for these EINs.

For those EINs that were subjected to initial sampling during the previous BSR (and were inactive on the Business Register at the time of the creation of the current BSR sampling frame), we use their classification and measure of size information from the Business Register and directly subject these EINs to current second phase birth sampling.

For those EINs that were subjected to first phase birth sampling as part of the previous BSR's quarterly processes, we use the results of the previous BSR first phase processing. EINs subjected but not selected as part of the previous BSR's first phase birth sampling are considered subjected but not selected for the current birth backlog first phase processing. Those EINs that were subjected and selected as part of the previous first phase processing are considered subjected and selected for the current birth backlog first phase processing. By doing this, we are accepting some bias in order to be able to use the accumulated response data from the previous BSR for these selected and mailed EINs. The selected EINs are then subjected to second phase birth sampling.

If the EIN was not subjected to previous BSR sampling, either as part of initial sampling or as part of a quarterly sampling operation, then one of the following is true concerning the EIN:

- Either the EIN was an add to the Business Register since the previous BSR's most recent quarterly process,
- The EIN was already on the Business Register, but only has recently become active, or
- There was a change in scope between the previous and current BSRs, so that EINs previously out of scope are now considered in scope.

EINs in this group have not yet been subjected to sampling as part of either the previous or current BSR, nor have they been mailed as the result of the previous BSR's first phase birth sampling operations. These EINs are subjected to first phase birth sampling as part of the birth backlog operations, and selected EINs are mailed for classification, measure of size, and company affiliation information.

5.3 Annual Sample Update

While the quarterly processes described above account for sample updates primarily for singleunits, the Annual Sample Update (ASU) addresses sample updates for multiunits. The ASU uses information from the Business Register, updated to reflect the latest COS results, along with administrative data. The ASU is conducted to improve survey coverage and data quality by attempting to minimize error in the estimates arising from the multiunit portion of the sample. The ASU also makes use of information on reporting parts in the current annual surveys. The ASU is performed each year prior to the mailing of the annual surveys.

The ASU uses the Business Register and current annual survey registers in a series of matching procedures to create listings and data sets of:

- All inscope establishments of the originally selected multiunit certainty companies and their successors. This ensures that all inscope NAICS industries of these companies are accurately measured in the sample to ensure accurate industry coverage
- Companies in the non-certainty universe that are growing rapidly and are not well represented in the selected sample
- Large companies that are not represented at all in the current samples, and
- EINs that were singleunits at the time of sample selection, but are now associated with a certainty company, in an effort to minimize duplication.

The analysts use the listings and data sets to determine what changes need to be made to either the survey registers or Business Register.

5.4 Status Changes

Any type of change to a business' structure since its initial inclusion into the sample is considered a status change. In addition to starting or closing a business, other changes to the structure or nature of a company can have an impact on our survey samples throughout the life of the sample. Events such as the merging of two companies, the divesting of one company into two or more pieces, an acquisition of part of a company by another company, or a change in the type of business conducted could have an impact on how a company reports data in our surveys, or how we choose to represent that company in our sample. Processes outlined in sections 5.1 through 5.3 above identify some of these changes. This section outlines other rules that have been established to handle all other possible changes that could occur over the life of the sample. These rules are collectively known as status change rules. These rules, like the other sample maintenance activities, are put in place to help maintain the sampling frame and minimize deterioration of the selected samples. For more details on the status change rules, see Burton (2006).

In general, a selected EIN is to report for all establishments inscope to the survey while a selected company is to report for all inscope establishments associated with that company. Status change rules are set up differently for certainty versus noncertainty EINs and companies, as certainty units represent only themselves in sample whereas noncertainties represent other similar units. Thus, the effect of various status changes on the variance of the estimates can differ greatly depending on that entity's contribution to the tabulated estimates.

The following terms are used throughout this section:

Successor – An EIN or company post-status change, and

Predecessor – The original EIN or company

5.4.1 Status Changes for Certainty Units

For certainty units, the status change rules depend on survey analysts following up on detected changes and making appropriate adjustments to the sampling units on the survey registers. When a predecessor certainty EIN or company is bought out or absorbed by a successor, the successor automatically becomes a certainty EIN or company (depending on what was acquired) in the sample, and should report for all establishments that it acquired from the predecessor. If it becomes a certainty company, it should also report

for any additional establishments that it may already own that are also inscope. The predecessor will become inactive.

A certainty EIN or company is considered a death only if all of its establishments under that EIN or company are out of business. If this is the case, then the entire certainty unit is set to inactive. Note, however, that this rarely happens.

When a certainty EIN or company splits and creates two or more new companies, those companies are added to the sample as certainties. The predecessor no longer exists and becomes inactive.

When a certainty EIN or company remains in business but some of its establishments form a new company or EIN, this is considered a spin-off. When a certainty EIN or company remains in business but sells some of its establishments to other already established EINs or companies, this is considered a divestiture. In the case of either a spin-off or a divestiture, the successor becomes a certainty EIN or certainty company, respectively.

It should be noted that any birth EINs associated with certainty units should be removed from birth sampling, as they should be reported for under the certainty unit automatically by association.

5.4.2 Status Changes for Noncertainty Units

Status change rules for noncertainty units are a bit more complicated than certainty units because these units have been selected into the sample to represent themselves as well as other similar firms. Additionally, not all surveys treat all noncertainty status changes the same. The Business Master File (BMF) indicates the business activity of an EIN. An EIN is considered to be BMF active if that EIN is used to report payroll information for the employees of the business represented by that EIN. A selected EIN is mailed and tabulated for a survey as long as it is BMF active. Once it goes BMF inactive, it is no longer mailed a survey questionnaire and is considered a delete. If at any time the EIN becomes active again, a questionnaire should once again be mailed, and it is considered a reactivation. In this case, its previously assigned tabulation characteristics such as weight, NAICS code, etc. will be used.

If a noncertainty EIN merges into a certainty EIN or company, it is converted to certainty status. If a selected noncertainty EIN merges into another selected noncertainty EIN, the annual surveys will continue to tab the predecessor EIN at its original level. The monthly and quarterly surveys will tab the predecessor with zero data until it goes BMF inactive. If the noncertainty merges into an EIN that was not selected in the sample, the annual surveys will continue to tab the predecessor EIN at its original level; while the monthly and quarterly surveys will tab the predecessor at its original level until it goes BMF inactive.

A noncertainty EIN is said to have died if the EIN has gone out-of-business, sold all of its inscope establishments or has changed to a new EIN. For MRTS, MWTS, and QSS, the EIN is retained and tabbed with nonzero data until it becomes BMF inactive, at which time it is set to not be mailed or tabulated. For the annual surveys, the EIN will continue to be tabbed to reflect the time that it was in business for the current survey year.

Some EINs may appear to be out of business (i.e., deaths) due to zero payroll, but are actually still thriving companies that have simply chosen to use a Professional Employer Organization to handle their payroll, etc. In these cases, these EINs should continue to be mailed and tabulated. If they are being represented as a nonemployer, based on other Census Bureau files and are found to be included in the annual nonemployer estimates produced by Census, then they can be removed from the survey files during benchmarking processing. Otherwise, no changes should be made.

5.4.3 NAICS Code, Type of Operation Code (TOC), and Tax Status Changes

Changes to classification characteristics that are used as stratification variables during sample selection are also considered status changes and must be handled with care. Similar to the rules above, these rules are designed to be consistent with other survey processes so that they do not introduce any significant bias or variation into the sample estimates. When a change to the NAICS code, TOC, or tax status is made, agreement must be reached between both the current survey area and the Economic Census area. This agreement is essential to ensure that data consistency is maintained across the two program areas.

No change to noncertainty EIN NAICS codes, tax status or TOC are allowed. An additional tabulating unit may be set up when a new establishment of a different NAICS code with significant sales or inventory is added to the EIN.

Any birth EIN that is classified inscope when selected during first phase birth sampling must remain inscope. If the EIN is later found to be out-of-scope, it should be assigned the most fitting inscope NAICS code and subjected to second phase birth sampling. Once the EIN has been added to the sample, it should not be mailed but should continue to be tabulated for estimation purposes.

For certainties, NAICS code and TOC code changes are allowed during the initial introduction of the sample, prior to any final tabulations for the first tabulation period of the new sample. If the certainty unit is found to be out-of-scope after this initial period, it should continue to be tabulated, and then dropped or transferred to the correct NAICS.

Acknowledgements

During the years since Konschnik, Monsour, and Detlefsen (1985) wrote their paper “Constructing and Maintaining Frames and Samples for Business Surveys”, the Census Bureau has continued to update its methodology. This paper incorporates the changes in methodology since 1985, and acknowledges the contributions of the authors to the original paper, and indirectly, to this paper as well.

The authors would also like to acknowledge Frank Sell, Jim Hunt, Matt Nelson, Brenna Matthews, and Miriam Rosenthal for their reviews and/or contributions to this paper.

References

- Burton, J., “Birth Backlog Operations,” Specifications for the Sample Maintenance Team (SG-4) of the 2006 Business Sample Revision, Initial Draft, October 2005, internal memorandum.
- Burton, J., “BSR-06 Status Change Rules,” Specifications for the Sample Maintenance Team (SG-4) of the 2006 Business Sample Revision, Initial Draft #3, February 2006, internal memorandum.
- Davie, W., “Design and Selection of BSR-12 Samples for the Census Bureau’s Retail, Wholesale, and Service Surveys,” Specifications for the Sampling Team (SG-2) of the 2012 Business Sample Revision, Initial Draft, February 2010, internal memorandum.
- Isaki, C.T., K.M. Wolter, T.R. Sturdevant, N.J. Monsour, and M.L. Trager, (1976), “sample Redesign of the Census Bureau’s Monthly Business Surveys,” *Proceedings of the Section on Business and Economic Statistics*, American Statistical Association, 90-98.
- King, C.S., “BSR-06 Universe Extraction,” Specifications for the Universe Extraction Team (SG-1) of the 2006 Business Sample Revision, Final Version, August 2005, internal memorandum.
- King, C.S., (2007), “An Evaluation of Changes to Universe Extraction Editing for Current Business Surveys at the U.S. Census Bureau,” *Proceedings of the Third International Conference on Establishment Surveys*, American Statistical Association.
- Kinyon, D.L., D Glassbrenner, J. Black, and R.Detlefsen, (2000), “Improving the Design of Samples used for Business Surveys conducted by the United States Bureau of the Census,” *Proceedings of the Section on Survey Research Methods*, Joint Statistical Meetings, 388-393.
- Konschnik, C.A., N. Monsour, and R. Detlefsen, (1985), “Constructing and Maintaining Frames and Samples for Business Surveys,” *Proceedings of the Section on Survey Research Methods*, 113-122.
- Neft, M., “Creation of the Initial BSR-06 Feedback File,” Specifications for the Sample Maintenance Team (SG-4) of the 2006 Business Sample Revision, Initial Draft #2, June 2005, internal memorandum.
- Sell, F., “BSR-06 Creation and Update of the EIN Universe,” Specifications for the Sample Maintenance Team (SG-4) of the 2006 Business Sample Revision, Final Version #4, October 2008, internal memorandum.