

ICTS Sampling and Estimation Methodologies

The scope of the survey included all private, nonfarm, domestic companies. Major exclusions from the frame were government-owned operations, including the U.S. Postal Service; foreign-owned operations of domestic companies; establishments located in U.S. Territories; establishments engaged in agricultural production (not agricultural services); and private households.

The 2013 sampling frame was developed using 2012 administrative data in the Census Bureau's establishment-based database, the Business Register (BR). Enterprise level records were then created by consolidating the establishment level data. The 2012 BR contained records for each physical business entity with payroll located within the United States. Records included company ownership information and current-year administrative data, such as 2012 payroll. Enterprises comprised of only a single establishment are called single unit or single establishment enterprises. Collectively, the multi-establishment and single establishment enterprises resulted in a 2013 sampling frame of 5.7 million enterprises with paid employees based on 2012 administrative data.

For single establishment enterprises, the business activity classification is the classification already assigned to its establishment in the BR. For multi-establishment enterprises, business activity classification was assigned based on an examination of its constituent establishments. The employment and payroll data for each of these establishments were gathered using that establishment's assigned 2007 six-digit North American Industry Classification System¹ (NAICS) industry in the BR. The multi-establishment enterprise was then assigned to the economic sector in which it had an active establishment with the most payroll (e.g., manufacturing, construction, etc.). Following that, subsector within that sector, industry group within that subsector, and industry within that industry group were subsequently determined in the same fashion. Each enterprise, multi-establishment and single establishment, once having a 2007 NAICS industry code, was recoded to ICTS specific industry code.

The several million enterprises in the 2013 sampling frame were partitioned into two major portions: the certainties and noncertainties. The certainty portion was a group of 17,880 enterprises that had 500 or more employees based on 2012 administrative data. These enterprises were considered large enough for automatic, or certain, inclusion in the sample. The remaining enterprises in the 2013 sampling frame, or noncertainty portion, had between 1 and 499 employees based on 2012 administrative data. They were stratified into one of the ICTS industry codes. Each of these ICTS industry codes were further divided into four substrata based on 2012 administrative payroll. The exact payroll values of the substrata were determined by minimizing the overall sample size needed to achieve a desired level of reliability based on sample estimation of the known frame value of administrative payroll. Samples were chosen from each of these ICTS industry codes and their four substrata from the noncertainty part of the sampling frame.

The estimates in this 2013 data release were based on a stratified simple random sample of 46,642 companies with paid employees, as determined by having nonzero payroll in the previous year, 2012.

ESTIMATION

Estimation is the process of creating statistics from the reported data from surveys and censuses. In a survey, unlike a census, only some of the eligible units will be queried for information. For ICTS, the unit is the enterprise. Not all enterprises will respond, and this nonresponse must be

accounted for in estimation. This is discussed in more detail in sections below, but the measures below are different metrics for quantifying nonresponse.

The unit response rate (URR) is a quality measure defined as the percentage of all eligible enterprises that responded to the survey. If every eligible unit responded, the URR would be 100%. Enterprises thought to be eligible for the survey at the time of sampling may not be eligible, for instance, the business went out of business before the start of the survey year. The measure treats all eligible enterprises, no matter how large or small, equally. For the 2013 ICTS, the URR was 73.8%.

$$URR = \frac{R}{S} * 100$$

URR: Unit Response Rate

R : total number of companies that responded to the survey

S : total number of companies sampled

The total quantity response rate (TQRR) is a quality measure defined as the percentage of the estimated total from reported or 'equivalent to reported' data. If every eligible unit responded, the TQRR would be 100%. Unlike the URR, TQRR does not treat all eligible enterprises equally. An enterprise's impact on the estimates, as in the TQRR measure, varies with their sampling weight and their reported data. Each sampled enterprise has a sample weight reflecting other unselected enterprises in the population. Sampled enterprises in the same substratum have identical weights. Smaller weights reflect a sampled enterprise's data is not increased to represent other similar enterprises, since those similar enterprises were also sampled. Larger weights, which can be several thousand, reflect a sampled enterprise whose reported data will be used to reflect the data of many similar enterprises that were not sampled. In addition to sampling weights, the respondents' weights are further increased to widen their representation to account for enterprises that did not respond to the survey. This is the ICTS method of adjusting for nonresponse. The proportion of the published estimates coming from respondent data using only their original unadjusted-for-nonresponse sampling weights is the total quantity response rate. In 2013, this value was 90.0% for total capitalized expenditures and 88.8% for total noncapitalized expenditures.

$$TQRR = \frac{\sum_{h=1}^k \sum_{i \in h} (W_h * X_{i,h})}{\hat{X}_{tot}}$$

TQRR : total quantity response rate

W_h : substratum sampling weight of the h^{th} substratum

$X_{i,h}$: total capital expenditures value attributed to the i^{th} responding enterprise of substratum h .

\hat{X}_{tot} : published estimate for total capital expenditures for all enterprises

A third measure used during processing as a quality measure of an estimate is the percentage of payroll in the sample accounted for by the respondents. The coverage rate for the 2013 ICTS was 89.7%.

Sampling Weights and Weight Adjustment for Nonresponse

As discussed above, each sampled enterprise has an initial sampling weight which may then be adjusted based on characteristics such as activity status, response status, and employment status. Each sampled enterprise becomes a respondent, a nonrespondent, out-of-scope (if it is found to have been out of business prior to the survey year), or a duplicate to another record. Enterprises that went out of business during the survey year are still in-scope, and efforts are made to collect data for the period the enterprise was active.

An enterprise that receives the ICTS form is considered a respondent if the enterprise reported data in items 2 or 3 of the survey asking about expenditures.

To account for nonresponse, the sampling weights of respondents were raised to better represent the entire in-scope population. The adjustment for ICTS respondents is based on the outstanding payroll of nonrespondents by ICTS industry by substrata. In addition, enterprises who are deemed 'extreme outliers' may have their weights further adjusted to minimize the mean squared error of the estimates.

ICTS segment. The following discussion assumes 660 substrata. The substrata are designated $h = 1, 2, \dots, 660$. The 660 comes from 132 ICTS industries, each containing five strata, although some industries with relatively few enterprises may have fewer than five strata. The five strata are four noncertainty strata and the one certainty stratum. An enterprise in the certainty stratum has a sampling weight of one, while those in the other strata have weights usually greater than one, sometimes several hundred. All substratum sampling weights, W_h , are the same within each substratum h , equaling the ratio of the substratum population size, N_h , to its sample size, n_h . The ICTS respondent sampling weights are then adjusted for nonresponse based on payroll in the following way:

$$W_{h(adj)} = W_h * \frac{P_{hr} + P_{hn}}{P_{hr}}$$

where,

$W_{h(adj)}$: adjusted substratum weight of the h^{th} substratum

W_h : substratum sampling weight of the h^{th} substratum

P_{hr} : sum of total enterprise payroll for respondents in substratum h

P_{hn} : sum of total enterprise payroll for nonrespondents in substratum h

Publication Estimation

Publication cell estimates were computed by obtaining a weighted sum of reported values for in-scope respondents. The weights at this time have been adjusted (increased) to account for eligible sampled units that were nonrespondents. In this way, the entire population can still be estimated, although the precision of those estimates is reduced.

ICTS Estimation: The ICTS estimates, \hat{X}_j , are (where substrata $h = 1$ to k , and $k=660$) calculated as:

$$\hat{X}_j = \sum_{h=1}^k \sum_{i \in h} (W_{h(adj)} * X_{(j),i,h})$$

where,

$W_{h(adj)}$: adjusted weight of the h^{th} substratum

$X_{(j),i,h}$: value attributed to the i^{th} enterprise of substratum h , where j is the publication cell of interest.

N.B. ICTS allows an enterprise to report capital expenditures in various business activities. Although an enterprise is assigned to and sampled from a single ICTS industry thought to be its primary business activity, it can report capital expenditures in several ICTS industries. Larger enterprises commonly do so. Reported data for all reported industries are inflated by the weight in the sample industry of the respondent. This means that similar sized enterprises could influence a common industry differently based on which business activity they were assigned to as sampling business activity initially.

RELIABILITY OF THE ESTIMATES

The published estimates are derived from sample data, and will differ from results derived from data from other samples or from a complete census of the population. A sample and a census will both experience errors classified as nonsampling errors, which often introduce systematic bias into the results. Bias is the difference, averaged over all possible samples of the same design and size, between the estimate and the true value being estimated. These types of errors are not explicitly measured. Only samples have sampling errors, the error from only observing a subset of the population. With a probability sample, this type of error can be explicitly measured. For any particular estimate though, the total error from sampling and nonsampling error may considerably exceed the measured error.

Sampling Variability

The sample selected is only one of the many possible samples that could have been selected with that same design and size, with each possible sample producing possibly different results. The relative standard error (RSE) is a measure of the sampling variability among all these possible estimates from all these possible samples, relative to the estimates. These are calculated using a delete-a-group jackknife replicate variance estimator. The RSEs in the tables can be used to derive the standard error (SE), which can then be used to create interval estimates with prescribed levels of confidence, called confidence intervals (CI). Note that the SE is in the same units as the estimate, while the RSE is a unit-less number.

The SE of the estimate is calculated by multiplying the RSE by its corresponding estimate. Note that the RSE is the measure of variability presented for all estimates in this publication except for the estimates of percent change. RSEs are also given as a percentage, and need to be divided by 100 before being used to calculate the SE.

In general for a probability sample, intervals defined by 1.645 standard errors above and below the sample estimate will contain the true population value about 90 percent of the time. This only considers the effects of sampling, and not any of the other issues that can affect an estimate. Sampling variability is also important in determining which year to year changes are statistically significant. The Census Bureau quality standard is that a 90% confidence interval for an estimate of change must not include zero to be considered statistically significant.

Examples of Calculating a Confidence Interval (CI)

N.B. The estimates used in the examples are not from any particular table or cycle. They are examples to illustrate the concept.

a. Calculating a confidence interval for a specific estimate within a single survey year:

Consider an estimate for a sector in a particular year from table 4a is \$200,000 million and its companion RSE from table 4c is 5.0. The SE for that estimate would be calculated as follows

$$\begin{aligned}\hat{\sigma}(\hat{X}_j) &= \left(\frac{RSE(\hat{X}_j)}{100} \right) * \hat{X}_j \\ &= \left(\frac{5.0}{100} \right) * \$200,000 \text{ million} = \$10,000 \text{ million}\end{aligned}$$

The 90-percent confidence interval can be constructed by multiplying this SE by 1.645 to create the margin of error (MOE), and adding and subtracting the MOE to the estimate. The value of 1.645 corresponds to using the Census Bureau standard of 90% confidence intervals. The 90-percent confidence interval for the estimate for this sector's total noncapitalized ICT expenditures is:

$$\begin{aligned}\hat{X}_j \pm (1.645 * \hat{\sigma}(\hat{X}_j)) \\ &= \$200,000 \text{ million} \pm (1.645 * \$10,000 \text{ million}) = \$200,000 \pm \$16,450 \text{ million} \\ &= (\$200,000 - \$16,450 \text{ million}) \text{ to } (\$200,000 + \$16,450 \text{ million})\end{aligned}$$

Which gives a 90 percent confidence interval of \$183,550 million to \$216,450 million.

So there is 90% confidence that the interval from \$183,550 million to \$216,450 million contains the actual true value for noncapitalized ICT expenditures in this sector by enterprises with paid employees in 2013.

By probability theory, 90-percent of all samples should produce an interval estimate that contains the true unknown percent change. Confidence intervals also do not consider any additional issues due to nonsampling errors, e.g., measurement errors or nonresponse biases. However, ICTS confidence intervals are impacted by nonresponse due to the weight adjustments as discussed above. Had this particular industry had a higher response rate, its estimate of change may have been statistically significant.

Nonsampling Error

All surveys and censuses are subject to nonsampling errors. Nonsampling errors can be attributed to many sources, including: inability to obtain information about all enterprises in the sample; inability or unwillingness on the part of respondents to provide correct information; difficulties in defining concepts; differences in the interpretation of questions; mistakes in recording or coding the data; and other errors of collection, response, coverage, and estimation for nonresponse. Explicit measures of the effects of these nonsampling errors are not available. However, to minimize total nonsampling error, all reports were reviewed for reasonableness and consistency, and every effort was made to obtain accurate responses from all survey participants. Coverage errors, meaning errors from not including enterprises that are in-scope of the survey or mistakenly including those that are out-of-scope as eligible, may have a significant effect on the accuracy of estimates for this survey. The Business Register, a subset of which forms the sampling frame, may not contain all in-scope businesses, or have incorrect values of payroll that then affect how they are sampled and the impact of their responses through their sampling weights.

A more detailed profile on the quality of the Annual Capital Expenditures Survey is available on request. Please contact Economy Wide Statistics at 301-763-3324.

¹*North American Industry Classification System (NAICS) – United States, 2007.* For sale by National Technical Information Service (NTIS), Springfield, VA 22161. Call NTIS at 1-800-553-6847 or go to <www.census.gov/epcd/www/naics.html>.