Sampling and Estimation Methodologies

The estimates in this 2011 data release were based on a stratified simple random sample of 46,408 companies with paid employees, as determined by having nonzero payroll in the previous year, 2010.

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 2011 sampling frame was developed using 2010 administrative data in the Census Bureau's establishment-based database, the Business Register (BR). The 2010 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 2010 payroll.

In creating the ICTS sampling frame, company level records were created for companies that had more than a single establishment, called multi-unit or multi-establishment companies, by consolidating their establishment level data. Companies comprised of only a single establishment are called single unit or single establishment companies. Collectively, the multi-establishment and single establishment companies resulted in a 2011 sampling frame of 5.7 million companies with paid employees based on 2010 administrative data. For single establishment companies, the business activity classification is the classification already assigned to its establishment in the BR. For multi-establishment companies, 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 System[1] (NAICS) industry in the BR. The multi-establishment company was then assigned to the economic sector in which it had an active establishment with the most payroll (i.e.,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 company, multi-establishment and single establishment, once having a 2007 NAICS industry code, was recoded to an ICT industry category code.

The 5.7 million companies in the 2011 ICTS sampling frame were partitioned into two portions; certainties and noncertainties. The certainty portion was a group of 16,792 companies with 500 or more employees. These companies were considered large enough for automatic inclusion in the sample as certainty cases. The remaining companies, all having between 1 to 499 employees were stratified (grouped) into one of the 132 industry categories. Each of these industries is subdivided into four strata. Each stratum was sampled from to build the rest of the sample, adding another 29,436 companies. Since noncapitalized expenditures data were not available on the sampling frame, 2010 payroll was used as the stratification variable. The stratification methodology resulted in minimizing the sample size subject to a desired level of reliability for each industry. The expected relative standard errors (RSEs) ranged from 1 to 3 percent.

ESTIMATION

The quality measure called the sampling unit response rate (URR) is the percentage of all mailed eligible companies that responded, which was 77.9%.

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

R : total number of companies that responded to the survey

S : total number of companies sampled

A company's impact on the estimates varies with their sampling weight and their reported data. Each sampled company had a sample weight reflecting other unselected companies in the population. Sampled companies in the same substratum had identical weights, which ranged from one, only represents itself, to several thousand. Respondents' weights were further increased to widen their representation to account for companies that would have been represented by nonrespondents. Final estimates used these increased weights. The coverage rate was a quality measure that is the percentage of the estimate of total noncapitalized and capitalized expenditures from respondents using only their original sampling weight. The coverage rate for ICTS was 90.8%.

Sampling Weights and Weight Adjustment for Nonresponse

Each sampled company had an initial sampling weight which may then be adjusted based on activity and response status. The goal was to have the in-scope responding sample reflect the entire in-scope frame. Each sampled company became a respondent, a nonrespondent, or out-of-scope due to being not in business during the survey year or a known duplicate to another record. Companies that went out of business during the survey year were still in-scope, and efforts were made to collect data for the period the company was active.

A company was considered a respondent or nonrespondent based on whether the company provided sufficient data in items 2 or 3 of the survey form asking about expenditures. Respondents had their sampling weights adjusted upwards to account for nonrespondents, such that the weighted respondents represented the entire in-scope population's total activity. The adjustment for respondents was based on the outstanding payroll that nonrespondents account for in each stratum. This adjustment may bias the estimates, since it was assumed that nonresponse is a purely random event, in that the relationship of payroll to expenditures does not differ in the aggregate from respondents to nonrespondents. No attempt was made to estimate the magnitude of any such bias due to the nonresponse weight adjustment.

In addition, companies who were deemed 'extreme outliers' may have had their weights reduced to minimize their impact on the estimates and the mean squared error of the estimates.

ICTS segment. The following discussion assumes 660 strata (strata designation h = 1, 2 . . . 660) which are based on 132 industries, each normally containing five strata (including the certainty stratum). When there was insufficient sample size to justify distinct strata, they are collapsed together. The original stratum weights (W_h) are adjusted to compensate for nonresponse. The adjusted weight is computed as follows:

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

where,

$W_{h(adj)}$ was the adjusted stratum weight of the hth stratum

$W_h = \frac{N_h}{n_h}$ was the original stratum weight of the hth stratum

N_h was the population size of the hth stratum

n_h was the sample size of the hth stratum

P_{hr} was the sum of total company payroll for respondent companies in stratum h

P_{hn} was is the sum of total company payroll for nonrespondent companies in stratum h

Publication Cell Estimation

Publication cell or point estimates were computed by obtaining a weighted sum of reported values for companies treated as respondents.

ICTS segment. The ICTS estimates were derived as follows. Each estimated cell total, $\hat{X}_j$ is of the form

$$\hat{X}_j = \sum_{h=1}^{660} \sum_{i \in h} \left(W_{h(adj)} \times X_{(j)ih} \right)$$

where,

$W_{h(adj)}$ was the adjusted weight of the hth stratum

$X_{(j)ih}$ was the value attributed to the ith company of stratum h, where j is the publication cell of interest.

Note: Although a company was assigned to and sampled in one ICTS industry, it may have reported expenditures in multiple ICTS industries. When this occurred, the reported data for all industries were inflated by the weight in the sample industry.

RELIABILITY OF ESTIMATES

The values shown in this report were estimates from a sample and can differ from the data which could have been obtained from a different sample or a complete census. Two types of possible errors are associated with any estimates based on data from sample surveys: sampling errors and nonsampling errors. The accuracy of a survey result depends not only on the measurable sampling errors but also on the nonsampling errors that are not explicitly measured. For any particular estimate, the total error may considerably exceed the measured sampling error.

Sampling Variability

The sample selected was only one of many possible samples that could have been selected using the sampling methodology described earlier. Each of these possible samples could yield different results. The relative standard error (RSE) is a measure of the variability among the estimates from all possible samples using the same methodology. The RSEs were calculated using a delete-a-group jackknife replicate variance estimator. The RSE accounts only for sampling variability, and does not account for any nonsampling error or systematic biases in the estimates. A bias is the difference, averaged over all possible samples of the same design and size, between the estimate and the true value being estimated.

The RSEs presented in the tables can be used to derive the standard error (SE) of the estimate. The SE can be used to derive interval estimates with prescribed levels of confidence that the interval includes the average results of all samples:

a. intervals defined by 1 standard error above and below the sample estimate will contain the true value about 68 percent of the time.

b. intervals defined by 1.645 standard errors above and below the sample estimate will contain the true value about 90 percent of the time.

c. intervals defined by 2 standard errors above and below the sample estimate will contain the true value about 95 percent of the time.

The SE of the estimate is calculated by multiplying the RSE presented in the tables by the corresponding estimate. Note, the RSE is the measure of variability presented for all estimates in this publication except for the estimates of percent changes presented in Table 2a, for which we provide the SE as the measure of variability (refer to Table 2b). Also note that RSEs in this publication are in percentage form. They must be divided by 100 before being multiplied by the corresponding estimate.

Examples of Calculating a Confidence Interval:

a. For a data value: using data from Table 3a and Table 3c the SE for 2011 total Nondurable Manufacturing noncapitalized expenditures would be calculated as follows:

$$\hat{\sigma}(\hat{X}_j) = \left[\frac{RSE(\hat{X}_j)}{100}\right] \times \hat{X}_j = \left(\frac{2.1}{100}\right) \times \$5{,}356\ million = \$112.5\ million$$

The 90-percent confidence interval can be constructed by multiplying 1.645 by the SE to create the margin of error (MOE), and then adding and subtracting this MOE to the estimate to create the upper bound, and subtracting it from the estimate to create the lower bound.

$$\hat{X}_j \pm \left[1.645 \times \hat{\sigma}(\hat{X}_j)\right]$$

Using the same data from Table 3a, the 90-percent confidence interval for the estimate of total Nondurable Manufacturing noncapitalized expenditures is :

$$\$5{,}356\ million \pm 1.645 \times (\$112.5\ million) = \$5{,}356 \pm \$185\ million$$

$(5{,}356 - \$185\ million)$ to $(5{,}356 + \$185\ million)$

This infers that if there were repeated samples taken of the population and similar confidence intervals created each time, 90 percent of those confidence intervals would contain the true population value, and 10 percent would not, if only uncertainty due to sampling is taken into account. This is generally translated to having 90 percent confidence that the interval of $5,171 million to $5,541 million also contains the actual total for Nondurable Manufacturing noncapitalized equipment expenditures, subject to further nonsampling errors.

b. For percent change: using data from Table 2a and Table 2b, the 90-percent confidence interval for percent change can be constructed by multiplying 1.645 by the SE of the percent change, adding this value to the estimated percent change to create the upper bound, and subtracting it from the estimate to create the lower bound. For example, for the noncapitalized expenditures in the Health Care and Social Assistance sector, the estimated percent change from 2010 to 2011 is 8.0 percent (from Table 2a), and the standard error of this estimate is 6.1 percent (from Table 2b)

$$8.0\% \pm [1.645 \times 6.1\%] = 8.0 \pm 10\%$$

$(8.0\% - 10\%)$ to $(8.0\% + 10\%)$

The 90 percent confidence interval is then -2% to 18% change in this sector for noncapitalized expenditures. Since this interval does contain zero (0), there is not sufficient evidence to conclude that the estimated percent change was statistically different from no change, i.e., the percent change is not statistically significant.

Examples of Calculating Absolute Differences and Percent Changes

Data for the current year along with revised data for the prior year are presented in this publication. Two numbers of interest for many data users may be the difference between the prior year and the current year, and the percent change from the prior year to the current year.

The difference is calculated as:

$$\hat{d}_j = [\hat{X}_t - \hat{X}_{(t-1)}]$$

where,

$\hat{X}_t$ is current year estimate of interest.

$\hat{X}_{t-1}$ is prior year estimate of interest.

While the 90 percent confidence interval on this difference is estimated as:

$$CI(\hat{d}_j) = \hat{d}_j \pm 1.645 \times \sqrt{[\sigma^2(\hat{X}_t) + \sigma^2(\hat{X}_{(t-1)})]}$$

For example, on Table 3a and 3c the 2011 estimate and RSE for total noncapitalized ICT expenditures in the Retail trade sector is $4,848 million and 1.9 percent, respectively. The 2010 revised estimate from Table 3b is $5,413 million with the RSE found in Table 3d as 11.3 percent. The difference between 2010 and 2011 is:

$$\hat{d}_j = [\$4{,}848\ million - \$5{,}413\ million] = -\$565\ million$$

The 90-percent confidence interval is estimated as:

$$CI(\hat{d}_j) = -\$565 \pm (1.645 \times \sqrt{[(.019 \times \$4{,}848)^2 + (.113 \times \$5{,}413)^2]})$$

$$= -\$565 \pm (1.645 \times \$618.6)$$

$$= -\$565 \pm \$1{,}017.5$$

$$= (-\$565 - \$1{,}017.5) \text{ to } (-\$565 + \$1{,}017.5)$$

$$\sim (-\$1{,}582.5 \text{ to } \$452.5)\ million$$

The 90-percent confidence interval is approximately -$1,582.5 million to $452.5 million.

The percent change is calculated as 100 multiplied by the ratio of the difference divided by the prior estimate.

Continuing with the previous example,

$$PC_j = 100 \times \left[\frac{\hat{d}_j}{\hat{x}_{(t-1)}}\right]$$

$$= 100 \times \frac{-\$565}{\$5{,}413}$$

$$= -10.4\ percent$$

The 90-percent confidence interval on this percent change is estimated as:

$$CI(PC_j) = PC_j \pm 1.645 \times 100 \times \left(\frac{\hat{X}_t}{\hat{X}_{(t-1)}}\right) \sqrt{\left[\left(\frac{RSE(\hat{X}_t)}{100}\right)^2 + \left(\frac{RSE(\hat{X}_{(t-1)})}{100}\right)^2\right]}$$

$$= PC_j \pm 1.645 \times 100 \times \left(\frac{\$4{,}848}{\$5{,}413}\right) \times \sqrt{(.019)^2 + (.113)^2}$$

$$= PC_j \pm 1.645 \times 100 \times 0.1026259$$

$$= PC_j \pm 16.9\ percent$$

$$= -10.4\ percent \pm 16.9\ percent$$

The 90-percent confidence interval is -27.3 percent to 6.5 percent.

Nonsampling Error

All surveys and censuses are subject to nonsampling errors, which are uncertainties in the estimates due to reasons other than taking a sample of the population for measurement. Nonsampling errors can be attributed to many sources, including: inability to obtain information about all companies in the sample; inability or unwillingness on the part of respondents to provide correct information; response errors; definition difficulties; differences in the interpretation of questions; mistakes in recording or coding the data; and other errors of collection, response, coverage, and estimation for nonresponse. In addition, the sampling frame may have inaccuracies such as not including in-scope cases. These coverage errors may have a significant effect on the accuracy of estimates for this survey. The businesses that are on the sampling frame may also have outdated or inaccurate data, such as inaccurate payroll, which can influence the estimates.

Explicit measures of the effects of these nonsampling errors are not available. However, to minimize nonsampling error, all reports were reviewed for reasonableness and consistency, and every effort was made to achieve accurate response from all survey participants.

A more detailed profile on the quality of the Information and Communication Technologies Survey is available on request. Please contact the Business Investment Branch of the Company Statistics Division 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 < http://www.census.gov/eos/www/naics/>.