

## Estimation:

### **Tabulation unit:**

ACES uses the company as the tabulation unit.

### **Estimation and sampling variance:**

Each sampled company has an initial sampling weight which may then be adjusted based on characteristics such as activity status, response status, and employment status. Each sampled company 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. Companies that went out of business during the survey year are still in-scope, and efforts are made to collect data for the period the company was active.

A company that receives the ACE-1 form is considered a respondent if they return a form in which they report some value of capital expenditures in item 1A, row11 (total Capital Expenditures), or they report data for some rows in item 2 (more detailed Capital Expenditures). A company that receives the ACE-2 form is considered a respondent if they report data for some rows in item 1 (Capital Expenditures).

To account for nonresponse, the sampling weights of respondents were raised to better represent the entire in-scope population. The adjustment for ACE-1 respondents is based on the outstanding payroll of nonrespondents by ACES industry by substrata. The adjustment for ACE-2 respondents is based solely on the percentage of companies not reporting, regardless of size. In addition, companies who are deemed 'extreme outliers' may have their weights further adjusted to minimize the mean squared error of the estimates.

**ACE-1 segment.** The following discussion assumes 660 substrata. The substrata are designated  $h = 1, 2, \dots, 660$ . The 660 comes from 132 ACES industries, each containing five strata, although some industries with relatively few companies may have fewer than five strata. The five strata are four noncertainty strata and the one certainty stratum. A company 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 ACE-1 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  $h^{\text{th}}$  substratum

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

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

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

**ACE-2 segment.** The ACE-2 segment initially was stratified into four strata based on four small business categories as discussed above. Two of these strata are further divided, or poststratified, after survey data collection based on updated administrative record data showing the presence of payroll. The stratum consisting of “companies without payroll in the prior year or employment on March 12<sup>th</sup> of the prior year, but had paid employees in the past and some IRS activity in the last 5 years” was poststratified into two strata. The stratum “companies that applied for an EIN in the last 2 years, but still have no payroll, employment, or receipts” was poststratified into two strata. This method resulted in six ACE-2 strata (strata designation h = 1, 2...6). The stratum population sizes, sample sizes, response counts, and stratum weights for the four new strata resulting from the poststratification were modified accordingly, while the other two strata retained their original weights.

The ACE-2 stratum weights ( $W_h$ ) were also adjusted to compensate for nonresponse based on number of respondents:

$$W_{h(adj)} = W_h \left( \frac{n_h}{r_h} \right) = \frac{N_h}{r_h}$$

where,

$W_{h(adj)}$  : adjusted stratum weight of the h<sup>th</sup> stratum

$W_h$  : stratum weight of the h<sup>th</sup> stratum

$N_h$  : population size of the h<sup>th</sup> stratum

$n_h$  : sample size of the h<sup>th</sup> stratum

$r_h$  : number of respondents in the h<sup>th</sup> stratum

### 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.

**ACE-1 Estimation:** The ACE-1 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<sup>th</sup> substratum

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

N.B. ACES allows an company to report capital expenditures in various business activities. Although an company is assigned to and sampled from a single ACES industry thought to be its primary business activity, it can report capital expenditures in several ACES industries. Larger companies 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 companies could influence a common industry differently based on which business activity they were assigned to as sampling business activity initially.

**ACE-2 Estimation:** The ACE-2 estimates,  $\hat{X}_j$ , are (with  $k=7$ ) are 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  $h^{\text{th}}$  stratum

$X_{(j),i,h}$  : value attributed to the  $i^{\text{th}}$  company in stratum  $h$ , where  $j$  is the publication cell of interest

N.B. There are no industry level estimates from the ACE-2 companies. Companies without employees only contribute to Table 1. Therefore,  $j$  becomes a constant and the estimate will always represent a national-level estimate.

### **Benchmarking:**

This survey is not benchmarked to any other program.

### **Seasonal adjustment:**

This survey does not use seasonal adjustment.

### **Disclosure avoidance:**

The Census Bureau operates under Title 13, U.S. Code, which prohibits the Census Bureau from making “any publication whereby the data furnished by any particular establishment or individual under this title can be identified”. This rule prohibits the Census Bureau from publishing a summary table that enables a data user to derive detailed information from an individual respondent. To ensure our tables do not violate disclosure rules implied by Title 13, they must be subjected to an analytical procedure referred to as disclosure avoidance.

Estimates are examined based on the individual contributions and contributors to that estimate. When the estimate for a specific data item might disclose sensitive information for an individual company, the

estimate is called sensitive. A sensitive cell is an estimate which a data user could use to calculate another company's data. Sensitive cells are not published, referred to as cell suppression. In addition, other cells will be suppressed to protect the sensitive cell from being analytically determined to within a proscribed amount. These additional suppressions are called complementary suppressions. They are necessary since users could frequently derive the values of sensitive cells since most data items are published in additive tables. Complementary cells must be suppressed so that the sensitive cell cannot be estimated closely, called protection. The process used to avoid sensitive disclosures used in ACES does not change the overall totals or otherwise affect the integrity of the published data. The effect is only to have a set of estimates withheld from publication. Suppressed estimates are replaced by the symbol "D".

An estimate is considered a sensitive cell if it fails a primary suppression rule. For ACES, the primary suppression rule is the p% rule. The p% rule attempts to protect the largest contributor's value, and therefore all contributors' values, in a given estimate from being estimated to within p%. Primary suppressions are identified during estimation. Complementary suppressions are identified based on information about sensitive cells and table structure in a separate process where all sensitive cells are adequately protected. There may be multiple choices for selecting complimentary cells. Different choices can be measured against their suppression cost, which is the sum of the data values suppressed as complimentary suppressions. In general, the suppression scheme with the lowest suppression cost but adequate protection will be chosen.