Annual Capital Expenditures Survey Methodology

Survey Overview

Survey Abbreviation: ACES

Purpose:

The Annual Capital Expenditures Survey (ACES) provides data on capital spending for new and used structures and equipment by U.S. nonfarm businesses with and without employees. The economy wide level of capital expenditures is an important component in the overall assessment of national productivity. Data on capital expenditures improves the quality of current economic indicators of business investments, provides facts about trends in capital expenditures useful for identifying business opportunities, and informs fiscal and monetary policy.

Capital expenditures are expenditures for structures or equipment chargeable to assets accounts for which depreciation or amortization accounts are ordinarily maintained. Structures expenditures are those to acquire, construct, and prepare a structure for its intended use. Equipment expenditures are those for machinery, furniture, motor vehicles and other expenditures used in the production and distribution of goods and services or office functions.

Frequency of Data Products:

Data products are produced annually.

Legal authority:
Title 13, U.S. Code, authorizes this survey and requires businesses and other organizations to complete this survey to respond to the U.S. Census Bureau. Notifications mailed to a home address are still applicable to the business, and responses are still required.

Under Title 13, U.S. Code, responses are confidential. Responses may only be seeing by persons sworn to uphold the confidentiality of Census Bureau information. Responses are only used for statistical purposes. Copies retained in your files are immune from legal process, and reported data are exempt from request made under the Freedom of Information Act.

Primary customers and users:

ACES statistics are an important input for federal agencies constructing composite national economic measures. This includes the Bureau of Economic Analysis' estimates of private fixed investments, a major component of gross domestic product, the Bureau of Labor Statistics to assist in creating estimates of capital stocks for productivity analysis, and the Federal Reserve Board's Flow of Funds accounts. Industry analysts and businesses use ACES for market analysis, economic forecasting, product development, and business planning. ACES data is also available at select research data centers, such as the University of Michigan, for empirical social and economic measurement.

Key statistics produced:

ACES produces annual estimates for major industry sectors and ACES industry categories in millions of current dollars. Estimates of total capital expenditures for major industry sectors are made, as well as the percentage change of total capital expenditures between the reference period to the prior period, and the prior period to the preceding period. For ACES industries, which are three digits and selected four digit NAICS industries, estimates are created of total expenditures, as well as estimates of new and used structures, and new and used equipment. All estimates have an estimate of sampling variation in associated tables.

In select years, ACES also collects details on investment of structures and equipment. These years of expanded collection have been 1998, 2003, 2008, 2012, and 2017. The next reference period with expanded collection is expected to be 2022. Expansion year estimates include estimates of investment for structures by major industry group and major structure type for companies with employees, as well as investment of structures by type and classification for companies with employees. Expansion year estimates also include estimates of equipment by major industry group and major equipment type for companies with employees and investment for equipment by type and classification for companies with employees.

Geographic detail:

This survey provides estimates at the national level for all domestic non-farm businesses, including agricultural non-farm and businesses without employees. It excludes foreign operations of U.S. businesses, businesses in U.S. territories, government operations, agricultural production companies, and private households. There are no subnational estimates. ACES collects information at the firm level about activities across all of its geographic operations within the scope of the survey.

Publication level of detail:

Estimates up to 1998 used the Standard Industrial Code (SIC). Estimates from 1999 to the present use the North American Industrial Code System (NAICS). ACES currently uses 2007 NAICS to create ACES specific industry groups. The ACES specific industry groups are comprised primarily of three digits and selected four digit industries. These industries combinations were developed through consultation with data users. A category was created for structures and equipment expenditures serving multiple industries, for example, headquarters, regional offices, and central research laboratories.

Data dissemination:

Data products are generally available 14 months after the end of the reference year. Each release also revises the information from the prior period. Data products may be later if the reference year is an expanded collection year that involves collecting details of types of structures and equipment investment. Those more detailed tables are excluded from revision. The next reference period that collects those expanded details is expected to be 2022. ACES data is available for each year from 1994 to the current reference year.

Data is available online at the Census Bureau website at https://www.census.gov/programs-surveys/aces.html

Release and Revision Schedule:

Data for the 2016 reference period is expected to be available online in February 2018. This will include revised estimates for the 2015 reference period.

In the event of an unplanned data revision, the revised data will be available online at the Census Bureau website https://www.census.gov/programs-surveys/aces/news-updates/updates.html. The revised data will replace all previously released versions. Primary stakeholders will be notified of the update.

History of Survey Program

Congress provided the initial funding for the survey in fiscal year 1991. At that time, the U.S. Census Bureau developed and conducted a feasibility survey to collect 1991 data from a sample of approximately 4,400 nonfarm companies. The purpose was to test the clarity of questions and instructions, and to determine the ability of companies to report the requested data. The Bureau selected a sample of 11,200 nonfarm companies in 1992. While primarily used to evaluate content, form and instructions, selected results were published in May of 1994.

After evaluating the 1992 survey results, it was determined that the annual collection of detailed expenditures on the types of structures and equipment purchased was overly burdensome. A cycle of differing collections was developed to repeat every 5 years. The 5-year cycle included conducting annually a basic survey that collects total capital expenditures for new and used structures and equipment from companies with five employees or more and biannually a survey of companies with fewer than five employees, including those with no employees. Collection of detailed information on types of structures and equipment would occur once during the 5-year cycle with structures information collected in 1994 and equipment in 1996.

A proposal to revise this plan was approved by the Office of Management and Budget beginning with the 1996 survey. The new plan included a mail sample of all small companies annually to provide an improved time series estimate of total and new capital expenditures by all companies. Additionally, detailed information on types of structures and equipment would be collected in the 1998 survey from companies with employees and every 5 years thereafter or as determined by the 1998 survey results. This pattern continued with detail information collected in the 2003 and 2008 reference years. In 2012, the detailed information was collected a year early to coincide with the 2012 Economic Census. This new pattern is expected to continue with the expanded collection of 2017 data to coincide with data collected from the 2017 Economic Census.

Survey Design

Target population:

The scope of ACES is to capture investment by all domestic, private, non-farm businesses, including agricultural non-farm business and businesses without employees. Investment made after applying for an Employer Identification Number (EIN) from the Internal Revenue Service (IRS) but before having any payroll or receipts is also included.

Major exclusions are foreign operations of U.S. businesses, businesses in the U.S. territories, government operations (including the U.S. Postal Service), agricultural production companies and private households.

Sampling unit:

The survey collects information at the company level. The records of how the company invests are maintained at the headquarters level, and not at the location of each physical operating location.

Sampling design and sampling frame:

ACES collects information at the company level. The records of how the company invests are maintained at the headquarters level, and not at the location of each physical operating location. Companies may elect to have divisions within the company report, but the sampling unit and tabulation unit will be the company. A company's importance to the survey depends on their employment, payroll, and business activity. The greater the number of employees or the larger the payroll, the more likely, in general, a company is to be selected in the sample. The influence that the amount of payroll has on the likelihood of selection is adjusted by the business activity such that two companies with similar payroll but in different business activities will not have the same likelihood of selection.

The estimates come are from two distinct samples drawn using different sampling designs from two distinct frames. The first frame collects in-scope companies with employees. Companies sampled from this frame will receive an ACE-1 instrument, and the frame and sample are the ACE-1 frame and sample, respectively. The second frame collects companies without employees, also called nonemployer companies. Companies sampled from this frame will receive an ACE-2 instrument, and the frame and sample are the ACE-2 frame and sample, respectively.

Administrative records data in the Census Bureau's establishment-based database called 'the Business Register' (BR) is used to create the 2016 ACE-1 frame. Administrative data about the previous year concerning companies with employees is not of sufficient quality to use for sampling until about 9 months after the end of the target year. Final 2015 administrative data was available in October 2016, and was used to create the 2016 ACE-1 sampling frame. The administrative records are for each physical business entity, called 'the establishment', located within the United States. This information includes company ownership and contact information, and data such as employment, payroll and a single NAICS code associated with each establishment that describes its major business activity. Establishments that are out of scope due to location, ownership, or major business activity based on their assigned NAICS codes, are removed from the sampling frame.

During the creation of the ACE-1 frame, significant parts of the ACE-2 frame are also identified. These include the 3C and 3F stratum of the ACE-2 strata. The companies in these ACE-2 strata are included because while they are not considered active businesses in many economic surveys, they have been shown to make some capital investments. However, response rates for these companies are often lower than other parts of the survey, since many of these are truly inactive. For ACES 2016, a filter was put into place to try to remove companies that had a higher probability of being out of business, while retaining other companies in the hope of retaining the bulk of the investment of these groups.

There were about 4.5 million applicants in 2015 for an EIN (Employer Identification Number) that did not have any further administrative data, however only 1.4 million passed a filter for being a possible active business. These 1.4 million cases were moved from the ACE-1 frame for companies with employees into the ACE-2 frame for companies lacking employees. These transferred EIN applicants are known as the 3F stratum in the ACE-2 frame. There were about 9.5 million establishments that previously had either annualized payroll or first quarter employment but no longer have nonzero data based on the 2015 administrative records data. However just 1.0 million cases passed the filter to only keep cases that had a high probability of being in business. These 1.0 million companies were moved from the ACE-1 frame into the ACE-2 frame, and are known as the 3C stratum.

In the ACES-1 frame, company level records are created by aggregating establishments with the same ownership. The majority of companies, about 5.7 million in 2016, are only a single establishment and no aggregation is needed. These are called single-unit companies (SU). A company that has more than a single establishment, which number about 170,000 in 2016, are called multi-unit companies (MU).

For single unit companies, the business activity classification is the NAICS classification already assigned to its sole establishment in the BR. For multi-unit companies, business activity classification is assigned based on an examination of its constituent establishments. The payroll data for each establishment is collected by its assigned NAICS industry. The multi-unit company is assigned a code based on an algorithm that first assigns the company to the trade area in which it had an active establishment with the most payrolls (e.g., manufacturing, construction, etc.). Within that trade area, the three digit NAICS with the highest payroll is selected. The subsector within that sector with the highest payroll is selected, with the process continuing until the multi-unit company has a six-digit NAICS code. Each company, multi-unit and single unit will have a 2012 NAICS code. This NAICS code is then recoded to an Annual Capital Expenditures Survey specific industry code (ACE codes). These ACE codes are composed primarily of three digits and select four digit NAICS codes.

All of the companies in the 2016 ACE-1 sampling frame, both single unit and multiunit companies, are then partitioned into two major portions: the certainties and noncertainties. Certainties are those companies that will be selected in the sample based solely on the magnitude of their administrative data. The data collected from certainties will only represent themselves, while noncertainties have their data magnified to represent other similar companies that were selected.

The certainty portion of the ACE-1 sampling frame includes 3981 single unit companies and 15,152 multi-unit companies, for a total of 19,133 certainty companies. The criteria for a company in the ACE-1 frame to be a certainty is having 500 or more employees based on their 2015 administrative data. This is a constant across every business activity, and many ACE codes will have far more certainties selected than other ACE codes. The number of certainties selected does not necessarily reflect how many businesses are in that that ACE code, nor the relative amount of investment that companies in that ACE code account for.

The remaining ACE-1 frame companies may or may not be chosen based on their selection probability and then a probabilistic mechanism to compare against that sampling unit's probability. This portion is called the noncertainty portion of the frame and is still about 5.7 million cases, and those that are selected are the noncertainty companies in the sample. All companies in this portion of the ACE-1 frame show indications of having between 1 and 499 employees or at least some payroll based on 2015 administrative data.

The noncertainty employer portion of the ACE-1 frame were stratified into one of the ACE codes based on their business activity. Each of these ACE codes were further divided into four substrata based on 2015 administrative payroll using a modified Lavallee Hidiroglou method. The design can be considered a stratified simple random sample, with the main strata on ACE code, and substrata within those based on payroll. The exact payroll values of the substrata and sample required in each 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. ACE codes that contribute more investment will be prioritized for having more of the available sample. In addition, there are constraints that help to ensure any company selected does not represent too many other companies, which helps with the reliability of the estimate and in comparing estimates year to year. Companies selected in ACE codes with many companies or in ACE codes with relatively low investment or even in ACE codes where the estimate was more precise than required will have higher weights than those in other ACE codes. In the 2016 ACES, the 5.7 million companies in the ACE-1 noncertainty frame are represented by 30,865 companies selected as noncertainty companies in the sample. Together with the certainties already selected, the entire ACE-1 frame is represented by a sample of 49,998 selected companies.

One important deficiency in the sample design ACE-1 frame is the weak relationship of the administrative data such as the number of employees or payroll to the amount of capital expenditures. While many multiunit companies will be in successive cycles, even their value of capital expenditures in the previous cycle is not a reliable predictor of investment in the current cycle. Knowledge of business activity and size is still helpful in reducing the error due to sampling, and still allows for a reasonable sampling design.

The sample design and the sampling frame for the second ACE-2 group, which are those companies that are not expected to have employees in the current period, is also a form of stratified simple random sample, but it is notably different from the ACE-1 design. The ACE-2 sampling frame is a composite frame of four categories of small businesses, each treated as an independent stratum and randomly sampled within those strata. While both ACE-1 and ACE-2 designs are stratified, the ACE-2 design has only four strata, and there are no substrata based on payroll since payroll no longer is applicable for businesses without employees. The primary strata for the ACE-1 frame is based on a reliable estimate of business activity, but that data is not considered reliable for the ACE-2 frame and is not used in sampling. Instead, the primary and only stratification in the ACE-2 frame is based on either its legal form of organization, or its origin from creating the ACE-1 frame.

The first two of those four strata are described above as the 3C and 3F stratum, with 1.0 million and 1.4 million companies. They are created during the creation of the ACE-1 frame from cases that are within the scope of ACES, have administrative data on the final 2015 BR, but are not seen as companies with employees.

The last two strata of the ACE-2 sampling frame are from a separate 2015 nonemployer specific database. This database includes companies other surveys would likely consider active nonemployer businesses. The 2015 administrative data for this 2015 BR database is not final until June of 2017, so a

preliminary 2015 dataset available in January of 2017 is used. Most companies are sole proprietorships without employees, ~ 20.0 million, which form the 3E strata. The remaining companies, the 3D strata, are nonemployer corporations and partnerships, ~ 3.3 million. Collectively, these four strata comprise an ACE-2 frame of about 25.7 million records. Simple random samples taken from the four categories result in an ACE-2 sample of 20,000 selected companies.

The ACE-2 sample allocates the sample based on an examination of the prior years' contributions of each stratum to the estimate of total capital expenditures. The size of the substrata populations, the amount of capital expenditures and variability of the contributions in the strata are examined over the prior three cycles using a Neyman allocation to reduce the sampling variability in the estimate. This is also modified by a cost value based on how many in each substratum respond or are considered out of scope in the survey period. The three years of data is averaged to reduce the sensitivity in allocation. A simple random sample in each substratum of the appropriate substratum sample size is done using the modified Neyman allocation.

Deficiencies in the ACE-2 frame include that many of the companies are out of business, identify as companies without employees, and have no measure of their size or business activity. Some percentage of cases in both the 3C and 3F strata will show evidence that they were out of business in the reference year and are removed from scope during processing. However, many companies that are out of business may not result in a successful contact as the business is no longer there to respond. Others in these two strata will be found with employees in the reference year, which is not out of scope, but means the data is not collected in the ACE-1 frame where it would have been more appropriate. The size of some of the strata, particularly the 3E strata which has about 20.0 million records but generally modest investment, creates the risk that each responding company represents thousands of other unselected companies. Under these conditions, a company with an unusual investment reporting pattern could be included in the ACE-2 sample. The result is a risk for high sensitivity in both the estimate from companies without employees, and in the measure of quality about estimates which could be relatively high, or too low and understate the issues in the precision of the estimate.

The two samples, ACE-1 and ACE-2, are combined into a single sample. Together with the certainties and the noncertainties, the ACE-1 sample includes 49,998 selected companies. The ACE-2 sample adds another 20,000 companies, for a total ACES 2016 sample of 69,998 companies.

Frequency of sample redesign:

The ACES sample is reselected annually. The two frames are recreated each year from the best available administrative data. The analysis on where to shift sample to and from and redone each cycle based on the best available survey and administration data. After these processes, the samples are redrawn each year for both the ACE-1 and the ACE-2 portions. No units are deliberately held over consecutive cycles, but some units may meet the selection criteria in successive cycles. This is particularly (but not exclusively) true in the ACE-1 certainty portion, where larger companies may be in multiple ACES cycles. No units are deliberately excluded from the sample due to having been selected in a previous cycle.

Sample maintenance:

ACES resamples each year based on rebuilding the sampling frames and sampling analysis. There is no need for sample maintenance performed during the year to adjust the sample or sampling units. Some maintenance is done for those companies that were sampled but request to report as several separate

units. In addition, a sampling unit is not considered a respondent unless all of its reporting units have sufficiently responded.

Data Collection

Data items requested and reference period covered:

Example forms can be found at https://bhs.econ.census.gov/bhs/aces/form.html. The forms will be for the reference period then being collected only. For forms from prior periods, refer to historical data at https://www.census.gov/programs-surveys/aces/library/publications.html

Key data items:

The key data items for ACES are investment items. These include total capital expenditures, under the domestic depreciable asset data question or capital expenditures question, or reporting a total for either investment in structures or equipment in the capital expenditures question.

Type of request:

This survey is mandatory. Title 13, United States Code, requires businesses and other organizations that are asked to complete this survey to answer the questions and submit their responses to the U.S. Census Bureau.

Frequency and mode of contact:

Businesses and other organizations asked to complete this survey are notified by mail. This notification includes instructions to access a computer self-administered questionnaire online. The instructions also include a user identification number and a password.

Reporting units may be contacted at any time by telephone. Telephone contacts may inquire about responses or the lack of a response in general. Eight weeks after the initial mailing of notifications there is a second mailing, with a second follow-up eight weeks after that to reporting units that have not yet responded. A third follow-up mailing may also be conducted.

Data collection unit:

ACES is a company level survey, using the company as both the sampling and the tabulating unit. This means that companies are asked to report data for the company as a whole, and that data is used as a single component into the estimates from the company as a whole. Companies are asked to report for all domestic operations, including subsidiaries and divisions. This includes reporting for establishments of subsidiary companies where more than 50% ownership is held, as well as establishments of other companies where the reporting company has the power to direct or cause the direction of management and policies. Holding companies are asked to report for the entire corporation, including all subsidiaries under their ownership.

Most companies report for the company as a whole. This is usually because the company is comprised only of a single establishment, but generally true of multi establishment companies as well. However, companies may elect to report as units less than the entire company does, which fully partition the data for the company as a whole. In 2016, ACES sampled 49,998 ACE-1 companies and 20,000 ACE-2 companies.

Special procedures

Data is reported for the calendar year. If calendar year book figures are not available except at considerable cost, reasonable estimates are acceptable. If reasonable estimates on a calendar basis are not available, fiscal year data is accepted. Respondents should indicate on the questionnaire the exact dates the data represent if they are not from the calendar year. If a company's fiscal year is used and it ends in January, February or March, the company reports for the fiscal year ending in 2016. Otherwise, companies report for the fiscal year ending in 2016.

Estimates are acceptable. The data requested may not correspond to company accounting records. In that circumstance, carefully prepared estimates are acceptable. If a company does not have capital expenditures for the reporting period, it enters "0" in the appropriate cells.

In the event of a merger or acquisition, special actions are taken depending on the circumstance. If a company merges with or acquires another domestic company during the period, it is asked to include the domestic capital expenditures made by the merged or acquired company since the date of acquisition. The cost to a company for structures and equipment previously owned by the acquired company at the time of the merger or acquisition is to be reported as expenditures, but only if treated as capital expenditures by that company. Discontinued operations are treated in the same manner as continued operations. If during the survey year, a company had discontinued operations that were held for sale, it is asked to report data for the discontinued operations for the period of time the reporting company owned them.

Compilation of Data

Edits/other adjustments:

Subject matter analysts review the reported data. These edits refer to reported data from similar units in either the current or the prior cycles of the data. Common adjustments are reconciling internal inconsistencies in the reported data such as details not equating to reported totals, reporting data in the wrong units, or misunderstandings in the definitions or in the questions.

Tabulation units from the ACE-1 frame that were chosen during the probability-sampling portion have a sampling weight. This weight is a multiplier that allows the response data for statistical inference to similar units that were not selected in the sample. In some cases, the response data is both correct but not representative of other similar units that were not selected in the sample. Some of these units may have their sampling weight reduced marginally as outliers. This will improve the representativeness of the sample and improve the quality of this portion of the estimate.

Tabulation units in the ACE-2 frame from the 3C and 3F component are companies without employees that were removed from the ACE-1 frame for cause. Using updated administrative data prior to publication, some of these cases will be found to have employees in the survey period. They will be grouped separately from the others in substrata, a process called poststratification. This will improve the representativeness of the sample and improve the quality of this portion of the estimate.

Nonresponse and Imputation:

Capital expenditures have low correlation with administrative or collected data. There is not a reliable model for predicting capital expenditures from nonresponse based on the administrative or collected data. If a sampled company does not report, called unit nonresponse, the assumption is that the company has capital expenditures in direct proportion to its payroll measured against other reporting companies in the same ACES specific industry and payroll group. ACES will use a weight adjustment to reporting companies to account for companies that do not report. If a company reports capital expenditures sufficiently to be a respondent to the survey, but does not answer a particular question, called item nonresponse, then either a survey analyst will estimate the missing data or no correction will be made.

Nonresponse adjustment:

Capital expenditures have low correlation with administrative or collected data. There is not a reliable model for predicting capital expenditures from nonresponse based on the administrative or collected data. If a sampled company does not report, which is known as unit nonresponse, the assumption is that the company has capital expenditures in direct proportion to its payroll measured against other companies that did report in the same ACES specific industry and payroll group. ACES uses a weight adjustment to reporting companies to account for companies that do not report. The weight adjustment for nonresponse is made to the original sampling weight, so that the sample can be used to make inference back to the entire population.

If a company reports capital expenditures sufficiently to be a respondent to the survey, but does not answer a particular question, known as item nonresponse, then either a survey analyst will either estimate the missing data or no correction will be made.

The nonresponse weighting adjustment differs between the two samples. Every company in the ACE-1 sample, which are all determined to have employees, has payroll. In addition, every company in the ACE-1 sample has been put into a group of businesses with similar activities. If a company fails to respond to the survey, and that company is still considered to have been active and in scope during the survey cycle, then its payroll is added to other companies in that ACES industry group and substrata that also did not report. Each ACES industry has at least one to as many as five substrata. Each substratum should have sufficient number sampled to ensure there is at least one respondent in the substrata, or separate action has to be taken. The fraction of the sum of the payroll of all sampled companies over the sum of payroll of just responding companies is calculated for each substratum. This value, which must be greater than or equal to one, is multiplied by the sampling weight of each responding company to compute a nonresponse adjusted weight.

Companies in the ACE-2 sample are assumed not to have employees or payroll. Further, the business activity codes associated with them are not considered reliable enough to use for estimation. The compensating fraction for a nonresponse adjustment will be the number of sampled but still eligible cases over the number of eligible reporting cases. This value, which must be greater than or equal to one, is multiplied to the sampling weight of each responding company to compute a nonresponse adjusted weight.

The nonresponse-adjusted weights for both the ACE-1 and ACE-2 samples are used in estimation. Use of the nonresponse weight adjustments assumes companies that do not report are not different from those that do with respect to capital expenditures. The adjusted weights also provide enough influence to responding companies for adequate inference to the population. The higher weight values also increase estimation variability, which should occur with larger non-observation due not only to sampling, but also to decreased response rates.

Imputation:

This survey does not use imputation other than that provided by the survey analysts. Missing data is handled with the nonresponse weight adjustment as detailed above.

Other macro-level adjustments:

Tabulation units from the ACE-1 frame that were chosen during the probability-sampling portion have a sampling weight. This weight is a multiplier that allows the response data for statistical inference to similar units that were not selected in the sample. In some cases, the response data is both correct but not representative of other similar units that were not selected in the sample. Some of these units may have their sampling weight reduced marginally as outliers. This will improve the representativeness of the sample and improve the quality of this portion of the estimate.

Outlier detection is done initially through calculating the Mahalanobis distance between investment in structures and equipment. While ACES is primarily concerned with total capital expenditures, the lack

of a reliable covariate makes outlier detection on total capital expenditures difficult. The focus is then on the relative values of structures and equipment investment. This is examined by ACES specific industry. The reporting companies with the largest Mahalanobis distance for each ACES specific industry are identified. The survey analysts to double check the validity or representativeness of the reported data then review these as potential outliers. Based on the survey analysts' review and recommendation, select companies are inspected against similar companies as well as their impact on the estimates of structures or equipment expenditures in that ACES industry. A company passing those tests is considered an outlier. Its nonresponse-adjusted weight will be reduced in a manner to reduce the sampling variation. .

Tabulation units in the ACE-2 frame from the 3C and 3F component are companies without employees that were removed from the ACE-1 frame for cause. Using updated administrative data prior to publication, some of these cases will be found to have employees in the survey period. They will be grouped separately from the others in substrata, a process called poststratification. This will improve the representativeness of the sample and improve the quality of this portion of the estimate.

Updated administrative data comes from using the final BR datasets available about 12 months after the frame was first created. The data is inspected to see which reporting cases did have payroll in the survey year, as well as inspecting the reference period frame. Post-strata of both the frame and the sample are created, and new sampling weights are calculated. This implies new nonresponse adjusted weights as well.

Tabulation unit:

ACES uses the company as the tabulation unit. The sampling unit is also the company, and in all but a limited number of large companies, the company is also the reporting unit.

Estimation and sampling variance:
Each sampled company has an initial sampling weight to reflect how it will be used in making inferences back to the population. Each sampled company becomes a respondent, a nonrespondent, or declared an out-of-scope case. A sampled company can be defined as out-of-scope if it is found to have been out of business prior to the survey year or is a duplicate to another record in the sample. Companies that went out of business during the survey year are still in-scope, and efforts are made to collect data for the portion of the survey year that the company was active. Companies with mergers or acquisitions during the survey year are handled by attempting to capture all the data in a single record.

A company that receives the ACE-1 collection instrument is considered a survey respondent if the companies reports some value of capital expenditures in either item 1A, row11 (total Capital Expenditures), or in some rows in item 2 (more detailed Capital Expenditures). The company is considered a respondent even if the reported value is zero. A company that receives the ACE-2 collection instrument is considered a respondent if it reports data for some rows in item 1 (Capital Expenditures). Similarly, that company is considered a respondent even if the reported value is zero.

To account for those companies that did not respond, or the nonrespondents, the sampling weights of respondents were increased. The increase to respondent weight values is called the nonresponse weight adjustment. This is done so that the set of respondent companies can be used to make an inference back to the entire population. This increase is made to the original sampling weights that all the sampled companies (both respondents and nonrespondents) had initially.

The nonresponse weight adjustment for ACE-1 respondents is based on the outstanding payroll of nonrespondents within their ACES industry and substrata. The ACE-1 respondents are all companies with employees, and have payroll as well as reliable indications of their business activity. The adjustment for ACE-2 respondents is based solely on the percentage of companies not reporting, regardless of size, within their strata. The ACE-2 companies are those without employees, and their strata is based on general categories only. In addition, companies who are deemed 'extreme outliers' may have their weights further adjusted to minimize the mean squared error of the estimates.

The following paragraphs describe how this nonresponse adjustment and how the estimates are created.

ACE-1 Nonresponse Adjustment. The following discussion assumes 660 substrata, which is the approximate number given about 132 ACE code divisions of business activity, and 5 divisions of each business activity. Those divisions include the one substrata of the certainty cases based on having 500 or more employees, then four other noncertainty substrata divided by the size of their payroll. The substrata are designated h = 1, 2. . . 660. In any given year, the total number of substrata is usually a bit smaller than 660, as it can be more efficient to collapse some of the four payroll substrata at times.

A company in the certainty stratum has a sampling weight of one, which means its data only represents itself. Those companies selected in the other strata usually have weights greater than one to represent themselves and many other similar companies so that the sample can be used to make inference back to the entire population. In some situations, these sampling weights can grow very large (up to several hundred), which can become problematic during estimation. 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 adjusted within their particular substrata 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 hth substratum

W_h: substratum-sampling weight of the hth 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 Nonresponse Adjustment. The ACE-2 segment initially was stratified into four strata based on four categories. Two of these groups are defined by being companies without employees and their legal form of organization, while the other two are derived from being possible companies without employees based on assembly the ACE-1 frame. These two latter strata are further divided, or poststratified into four groups after survey data collection, based on updated administrative record data

showing the presence or absence of payroll. The population sizes, sample sizes, response counts, and adjusted weights for the four new strata resulting from the poststratification are modified accordingly. This combination of four sampling strata, with two of those split into four poststratification strata, results in six ACE-2 strata (strata designation h = 1, 2…6) for nonresponse adjustment.

The ACE-2 stratum weights (W_h) are adjusted to compensate for nonresponse based on number of respondents in that stratum:

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

Where,

$W_{h\,(adj)}$: adjusted stratum weight of the hth stratum

W_h : stratum weight of the hth stratum

N_h : population size of the hth stratum

n_h : sample size of the hth stratum

r_h : number of respondents in the hth stratum

Publication Estimation

Publication cell estimates are computed by obtaining a weighted sum of reported values for in-scope respondents. The adjusted respondent weight values are the sampling weights increased to account for eligible sampled units that were nonrespondents. In this way, estimates can make inference back to the entire population, 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 hth substratum

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

N.B. ACES allows a company to report capital expenditures in various business activities. Although a 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. The data for all reported industries are inflated by the weight of 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 initially assigned to as sampling business activity.

ACE-2 Estimation: The ACE-2 estimates $\hat{X}_j$ are (with k=6) 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 hth stratum

$X_{(j),i,h}$: value attributed to the ith 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, which has estimates only at the economy wide level. Therefore, j becomes a constant and the estimate will always represent an economy wide 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, based on a set of rules and parameters, that 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. Sensitive cells are also called primary suppressions. The process of not publishing a sensitive cell in a table is called cell suppression. Allowing the value of a sensitive cell to be closely estimated would be called a disclosure.

In addition to not publishing table cells that contain sensitive data, other cells are also suppressed to protect the sensitive cell from being analytically determined to within a proscribed amount. Since most tables have relationships between table cells or marginal totals, there can be a complicated pattern of suppressions created to protect the sensitive cell. These additional suppressions are called complimentary suppressions. The process of not publishing other table cells to protect a sensitive cell from being estimated too closely is called protecting the sensitive cell, or providing adequate protection for the sensitive cell. The process used to avoid 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. The suppression scheme is selected using linear programming techniques.

Accuracy, Reliability, and Limitations

Sampling error is a measure of the uncertainty in the estimate due to only observing a subset of observations from the population. The subset of observations must be chosen using probability sampling to enable sampling error to be estimated using statistical theory.

The published estimates are derived from sample data, and will differ from results derived from other samples or a census of the population. A sample and a census both experience errors classified as nonsampling errors, which often introduce systematic bias into the results. Bias is the difference between the estimate and the true value being estimated, averaged over all possible samples of the same design and size. These types of errors are not explicitly measured. Samples have sampling errors, but censuses do not. For a probability sample, this type of error can be explicitly measured. However, for any particular estimate, 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 must 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. Note that this definition of a 90 percent confidence interval only considers the effects of sampling, and not any of the other sources of error 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 below are not from any particular table or cycle, but are provided 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:

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

The 90-percent confidence interval can be constructed by multiplying this SE by 1.645 to create the margin of error (MOE), and then 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 capital expenditures is:

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

This results in a 90 percent confidence interval from $183,550 million to $216,450 million.

Therefore, there is 90% confidence that this interval contains the true value for capital expenditures in this sector by enterprises with paid employees in 2016.

b. Calculating a confidence interval for a percent change of an estimate between two survey years:

A confidence interval for the percent change of an estimate between two survey years can be calculated using estimates from Tables 2a and SEs from table 2b. The 90-percent confidence interval can be constructed by multiplying the SE of the percent change by 1.645 to create the MOE, and then adding and subtracting the MOE to the estimate. For example, from Table 2a, the a sector's total capital

expenditures estimated percent change from 2015 to 2016 is a positive 15.0 percent and from Table 2b, the standard error of this estimate is 10.0 percent.

$$15.0\% \pm (1.645 * 10.0\%) = 15.0\% \pm 16.45\%$$

$$= (15.0\% - 16.45\%)\ to\ (15.0\% + 16.45\%)$$

This results in a confidence interval of -1.45% to 31.45%.

By probability theory, 90-percent of all samples should produce an estimate of the percent change in this sector that contains the true unknown percent change. In this one observed sample, the estimate corresponds to a confidence interval of -1.45 percent to 31.45 percent. Since this confidence interval contains zero, there is *insufficient* evidence at the 90-percent confidence level to conclude that the estimated percent change was statistically different from zero, or that the change is positive. In other words, this sector showed a not statistically significant change for capital expenditures, even though the estimate of change is 15.0 percent. However, the interval is quite large, and had the estimate been slightly higher, or the standard error slightly smaller, the confidence interval might not have contained zero and shown a significant difference at the 90% confidence level.

Confidence intervals also do not consider any additional issues due to nonsampling errors (e.g. measurement errors or nonresponse biases). However, ACES confidence intervals are impacted by nonresponse due to the weight adjustments, as discussed above. If this particular industry had a higher response rate, its estimate of change may have been statistically significant.

Examples of Calculating Differences and Percent Changes

Data for the current year along with revised data for the prior year are provided in this publication. Data users can calculate a difference, $\hat{d}_j$ and a percent change, $\widehat{PC}_j$ between the current year and prior year estimates using data on tables where the difference and percent change are not expressly given. These estimates, along with the corresponding confidence intervals, are calculated using the following equations.

The difference is calculated as:

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

Where,

$\hat{X}_t$: Current year estimate of interest.

$\hat{X}_{t-1}$: Prior year estimate of interest.

The MOE for a 90-percent confidence interval on this difference is approximately:

$$MOE(\hat{d}_j) = 1.645 * \sqrt{\sigma^2(\hat{X}_t) + \sigma^2(\hat{X}_{t-1})}$$

As an example, consider an estimate for 2016 total expenditures for an industry from table 4a which is $150,000 million with a companion RSE on Table 4c, of 4.0%. The revised 2015 estimate for the same industry from Table 4b is $130,000 million with an RSE, found in Table 4d, of 9.0%. The difference is estimated as:

$$\$150{,}000\ million - \$130{,}000\ million = \$20{,}000\ million$$

The MOE for the 90-percent confidence interval of the year-to-year change is estimated as follows, including translating the RSEs into variances by dividing the RSE by 100, multiplying by the estimate, and squaring:

$$= 1.645 * \sqrt{\left[\left(\left(\frac{4.0}{100}\right) * \$150{,}000\ million\right)^2 + \left(\left(\frac{9.0}{100}\right) * \$130{,}000\ million\right)^2\right]}$$

$$= 1.645 * \sqrt{\left[\left((0.040) * \$150{,}000\ million\right)^2 + \left((0.090) * \$130{,}000\ million\right)^2\right]}$$

$$= 1.645 * \sqrt{\$36{,}000{,}000 + \$136{,}890{,}000\ million}$$

$$= 1.645 * \sqrt{\$172{,}890{,}000\ million}$$

$$= 1.645 * \$13{,}149\ million$$

$$= \$21{,}630\ million$$

The 90-percent confidence interval for the difference between the two years is:

$$\$20{,}000\ million \pm \$21{,}630\ million$$

$$(\$20{,}000 - \$21{,}630\ million)\ to\ (\$20{,}000 + \$21{,}630\ million)$$

This results in a CI of -$1,630 million to $41,630 million.

Therefore, we are 90-percent confident that the difference between the 2015 estimate and the 2016 estimate is between negative $1,630 million and $41,630 million. Since zero is in this interval, this is not sufficient evidence for a statistically significant change.

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

Continuing with the example from above,

$$\widehat{PC}_j = 100 * \left(\frac{\hat{d}_j}{\hat{X}_{t-1}}\right)$$

$$= 100 * \frac{\$20{,}000\ million}{\$130{,}000\ million}$$

$$= 15.4\%$$

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

$$MOE(\widehat{PC_j}) = 1.645 * \left(\frac{\hat{X}_t}{\hat{X}_{t-1}}\right) * \sqrt{\left(\frac{RSE_{\hat{X}_t}}{100}\right)^2 + \left(\frac{RSE_{\hat{X}_{t-1}}}{100}\right)^2}$$

$$= 1.645 * \frac{\$150{,}000\ million}{\$130{,}000\ million} * \sqrt{\left[\left(\frac{4.0}{100}\right)^2 + \left(\frac{9.0}{100}\right)^2\right]}$$

$$= 1.645 * (1.1538) * \sqrt{0.040^2 + 0.090^2}$$

$$= 1.645 * (1.1538) * \sqrt{0.0097}$$

$$= 1.645 * (1.1538) * (0.0985)$$

$$= 1.645 * 0.1136$$

$$= 0.1869$$

$$= 18.7\%$$

The 90-percent confidence interval for the percent change between the two years is:

$15.4\% \pm 18.7\%$

$(15.4\% - 18.7\%)\ to\ (15.4\% + 18.7\%)$

This gives a CI of -3.3% to 34.1%.

Because the 90-percent confidence interval contains zero, we *cannot* conclude that the percentage change from 2015 to 2016 is a statistically significant increase at the 90-percent confidence level.

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.

One type of nonsampling error is the error due to nonrespondents to the survey being different from respondents in substantial ways. This leads to nonresponse bias. Direct measurement of nonresponse bias, as for any bias, is difficult. Instead, some measures have been created that serve as general indicators of when nonresponse bias may be large enough to affect statistical inference.

The unit response rate (URR) is a quality measure defined as the percentage of all eligible companies that responded to the survey. If every eligible unit responded, the URR would be 100%. Companies thought to be eligible for the survey at the time of sampling may not be eligible, for instance, if the company went out of business before the start of the survey year. The URR treats all eligible companies, no matter how large or small, equally. For the 2016 ACES, the URR was 73.6% for the entire survey, and 80.2% for the ACE-1 (companies with employees) portion of the sample.

The following expression is used to calculate the URR:

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

URR: Unit Response Rate

R: total number of eligible companies that responded to the survey

S: total number of eligible 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 companies equally. A company's impact on the estimates in the TQRR measure varies with its sampling weight and reported data. Each sampled unit has a sample weight reflecting other unselected companies in the population. Sampled companies in the same substratum have identical weights. Smaller weights indicate a sampled company that represents few other companies not included in the sample. Larger weights, which can be several thousand, reflect a sampled company whose reported data is used to represent the data of many similar companies that were not sampled.

In addition to sampling weights, the respondents' weights are also increased to account for companies that did not respond to the survey, which is the ACES method of adjusting for nonresponse. The proportion of the published estimates coming from respondent data using only their original unadjusted-for-nonresponse sampling weights in the total quantity response rate. In 2016, this value was 88.5% across the entire survey, and 91.6% for the ACE-1 (companies with employees) part of the sample.

The following expression is used to calculate the TQRR:

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

TQRR: total quantity response rate

W_h: substratum-sampling weight of the hth substratum

$X_{i,h}$: total capital expenditures value attributed to the ith responding company of substratum h.

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

A third quality measure used during processing is the coverage rate. The coverage rate measures the capital expenditures reported by companies with paid employees, or. The percentage of payroll in the sample accounted for by the respondents. The coverage rate for the 2016 ACES was 93.6%.