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 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 their 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. This is done to improve the quality of the estimates from any particular ACES specific industry code for business activities.

Estimates are from two distinct samples from distinct frames. The first frame collects in-scope companies with employees. Companies sampled from this frame will receive an ACE-1 form, 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 form, and the frame and sample are the ACE-2 frame and sample, respectively.

Administrative record data in the Census Bureau's establishment-based database, the Business Register (BR) is used to create the 2015 ACE-1 frame. Administrative data about companies with employees is not of sufficient quality to use for sampling until about 9 months after the end of the target year. Final 2014 administrative data, available in October 2015, is used to create the 2015 ACE-1. The administrative records are for each physical business entity, the establishment, located within the United States, company ownership information, and data such as with 2014 payroll and a NAICS code associated with each establishment. Establishments that are out of scope for reasons of geography, ownership, or business activity based on their assigned NAICS codes are removed. Applicants for an EIN unit but do not have any further administrative data, about 4.4 million in 2015, are placed into the ACE-2 frame. These are the 3F stratum in the ACE-2 frame. Establishments that previously had annualized payroll or first quarter employment but no longer do, about 5.8 million in 2015, are also placed into the ACE-2 frame. These are the 3C stratum in the ACE-2 frame.

Company level records are created by aggregating establishments with the same ownership. The majority of companies, about 5.6 million in 2015, are only a single establishment. These are called single-unit companies. A company that has more than a single establishment, which number about 200,000 in 2015, is a multi-unit company.

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. The NAICS code is recoded to an Annual Capital Expenditures Survey specific industry code. These ACES codes are composed primarily of three digits and select four digit NAICS codes.

The 5.8 million companies in the 2015 ACE-1 sampling frame are 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 criterion for ACES is having 500 or more employees based on 2014 administrative data. The data collected from certainties will only represent themselves. The certainty portion is 3758 single unit companies and 14,856 multi-unit companies.

The remaining companies in the 2015 ACE-1 sampling frame are eligible for selection based on some probabilistic mechanism. This portion is called the noncertainty portion. All companies in this portion of the ACE-1 frame have 1 and 499 employees based on 2014 administrative data. They were stratified into one of the ACES specific industry codes. Each of these ACES industry codes were further divided into four substrata based on 2014 administrative 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. Samples were chosen from each of these ACES industry codes and their four substrata from the noncertainty part of the ACE-1 sampling frame to achieve a reliable estimate of the payroll in that substrata. Additional sample was allocated to reduce sampling weights, which will also increase the quality of the estimates. In the 2015 ACES, this resulted in an additional 26,284 companies selected.

A deficiency in the ACE-1 frame is the weak relationship of employment and payroll to the variable of interest, capital expenditures. While many multi-unit companies will be in successive cycles, their value of capital expenditures in a previous cycle is not a reliable predictor of future investment. While neither employment nor payroll is reliable predictors of investment, the administrative data is available, reliable, and allows for a reasonable sampling design.

The 2015 ACE-2 sampling frame is a composite frame of four categories of small businesses, each treated as an independent stratum. The first two categories are described above as the 3C and 3F stratum. 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 2014 BR, but are not companies with employees. The last two categories comprising the ACE-2 sampling frame are from a separate 2014 nonemployer specific database. This database is not final until June of 2015, so a preliminary dataset available in January of 2015 is used. Most companies are sole proprietorships without employees, about 18.6 million. This is the 3E strata. The remaining companies, the 3D strata, are nonemployer corporations and partnerships, about 3 million. Collectively, these four categories create the ACE-2 frame of about 31.5 million records. Simple random samples taken from the four categories resulted in an ACE-2 sample of 30,000 selected companies.

Deficiencies in the ACE-2 frame include out of business, coverage and in their size alone. A large percentage of cases in both the 3C and 3F cases will be considered out of business in the reference year. Sampling these strata often results in no successful contact as the business is no longer there to respond. Others in these two strata will be found to have had employees in the reference year, which is not out of scope, but will mean 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, creates the risk that each responding company represents thousands of other unselected companies. This creates high sensitivity in both the estimate from companies without employees, and in their measure of quality.

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. Companies may elect to have divisions within the company report, but the sampling unit and tabulation unit will be the company.

Sample design:

ACES selects an independent sample from both the ACE-1 and ACE-2 frames. Each sample is a stratified simple random sample. The ACE-1 sample is stratified by business activity then by payroll. It has an approximate sample size of 45,000. The ACE-2 sample is stratified by other criteria such as legal form of organization and reason for being removed from the ACE-1 frame. It has an approximate sample size of 30,000. The sample design uses administrative and past ACES data to allocate the two samples across the two frames to produce a set of estimates with lower sampling variability than other designs.

The ACE-1 sample comes from the ACE-1 frame. The 5.8 million companies in the sampling frame are 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 criterion for ACES is having 500 or more employees based on 2014 administrative data. The data collected from certainties will only represent themselves. The certainty portion are 3758 single unit companies and 14,856 multi-unit companies.

The remaining companies in the ACE-1 sampling frame are eligible for selection based on some probabilistic mechanism. This portion is the noncertainty portion. The size of the noncertainty portion of the sample is constrained by the number of certainties and the total approximate size of the ACE-1 sample. All companies in this portion of the ACE-1 frame have 1 and 499 employees based on administrative data. The first level of stratification is into one of the ACES specific industry codes. Each of these ACES industry codes are divided into four substrata based on administrative payroll. The exact payroll values determining the substrata and the sample size 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. This is done using a cumulative square root of rule followed by a modified Lavallee Hidiroglou method developed by Slanta and Krenzke in 1994 at the program's outset. This is supplemented by allocating sample where the number of establishments in a substratum is large. This will reduce the representative power of any responding unit, to lower the sensitivity of the estimate and its estimate of sampling variability. Samples are then selected using simple random sampling within each substratum based on the parameters. In the 2015 ACES, this resulted in an additional 26,284 companies selected.

The ACE-2 sample comes from the ACE-2 frame. The 32.5 million companies in the sampling frame are stratified into four groups based on either legal form or organization or reason for removal from the ACE-1 frame. The first groups are about 5.8 million companies that were in the ACE-1 frame until removed for lack of current payroll or employment. These are the 3C strata. The 3F strata are the 4.4 million companies that have applied for an EIN number from the IRS, but do not yet have administrative data elsewhere on the ACE-1 frame and were removed. The remaining companies were not removed from the ACE-1 frame, but are stratified by their legal form of organization. There are about 3.2 million partnerships and corporations in the 3D strata, and about 19.2 million sole proprietorships.

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 responded or were 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.

The two samples, ACE-1 and ACE-2, are compiled into a single sample.

Frequency of sample redesign:

The sample is reselected annually. This includes both samples from the ACE-1 and the ACE-2 frames. Information from the previous year, as well as administrative data, inform each new sample selection process. No units are held over consecutive cycles unless they meet selection criteria in each cycle. No units are excluded from the sample due to having been selected in a previous cycle.

Sample maintenance:

ACES resamples each year. There is no sample maintenance performed during the year to adjust the sampling or tabulating units. There is maintenance done for the reporting units if a multi-unit company requests to report as several separate units. A sampling unit is not considered a respondent unless all of its reporting units have also sufficiently responded.