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 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. 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. 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 2019 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 2018 administrative data was available in October 2019, and was used to create the 2019 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 2019, 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 4.7 million applicants in 2018 for an EIN (Employer Identification Number) that did not have any further administrative data, however only 1.5 million passed a filter for being a possible active business. These 1.5 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 10.0 million establishments that previously had either annualized payroll or first quarter employment but no longer have nonzero data based on the 2018 administrative records data. However just 1.2 million cases passed the filter to only keep cases that had a high probability of being in business. These 1.2 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 6.1 million in 2019, 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 is more than 150,000 in 2019, 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 2019 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 4,100 single unit companies and 16,172 multi-unit companies, for a total of 20,272 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 2018 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 6.1 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 2018 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 2018 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 2019 ACES, the 6.1 million companies in the ACE-1 noncertainty frame are represented by 29,711 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 49,983 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 1.2 million and 1.5 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 2018 BR, but are not seen as companies with employees.

The last two strata of the ACE-2 sampling frame are from a separate 2018 nonemployer specific database. This database includes companies other surveys would likely consider active nonemployer businesses. The 2018 administrative data for this 2018 BR database was not final until June of 2019, so a

preliminary 2018 dataset available in January of 2020 is used. Most companies are sole proprietorships without employees, about 22.1 million, which form the 3E strata. The remaining companies, the 3D strata, are nonemployer corporations and partnerships, about 3.5 million. Collectively, these four strata comprise an ACE-2 frame of about 28.4 million records. Simple random samples taken from the four categories result in an ACE-2 sample of 20,001 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 22.1 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 49,985 selected companies. The ACE-2 sample adds another 20,001 companies, for a total ACES 2019 sample of 69,986 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.