2024 Annual Integrated Economic Survey (AIES) Methodology

Survey Design

Target Population

The target population of the Annual Integrated Economic Survey (AIES) covers all private sector, non-farm employer companies located in the U.S. (50 states and the District of Columbia) as defined by the North American Industry Classification System (NAICS). The survey excludes foreign operations of U.S. companies, most government entities (including the U.S. Postal Service), agricultural production companies, private households, and companies located in U.S. territories.

Key Data Items

The key data items for the AIES are:

- Sales, value of shipments, or revenue (Revenue)
- Annual payroll
- First-quarter payroll
- Number of employees
- Total operating expenses

Coverage

The 2024 AIES is based on the 2017 NAICS, which includes coverage of the following NAICS sectors of the U.S. economy:

- 22 - Utilities
- 31-33 - Manufacturing
- 42 - Wholesale Trade
- 44-45 - Retail Trade
- 48-49 - Transportation and Warehousing
- 51 - Information
- 52 - Finance and Insurance
- 53 - Real Estate and Rental and Leasing
- 54 - Professional, Scientific, and Technical Services
- 56 - Administrative and Support and Waste Management and Remediation Services
- 61 - Educational Services
- 62 - Health Care and Social Assistance
- 71 - Arts, Entertainment, and Recreation
- 72 - Accommodation and Food Services

- 81 - Other Services (except Public Administration)

The following industries are out-of-scope for the 2024 AIES. However, companies that have establishments with business activities in these industries were asked to provide key data items limited to revenue, annual payroll, first quarter payroll, and number of employees. Companies with establishments in a subset of these out-of-scope industries will also be asked to provide a group of supplemental data items, including company-level business characteristics and company classification (formerly collected in the Report of Organization). Capital expenditures and robotics data (formerly collected in the Annual Capital Expenditures Survey [ACES]) were only collected from in-scope units for SY24.

Key Data Items Only Collection

- 111 - Crop Production
- 112 - Animal Production and Aquaculture
- 491 - Postal Service
- 525 - Funds, Trusts, and Other Financial Vehicles
- 54112 - Offices of Notaries
- 814 - Private Households
- 92 - Public Administration

Supplemental Items Collection

- 113 - Forestry and Logging
- 114 - Fishing, Hunting and Trapping
- 482 - Rail Transportation
- 6111 - Elementary and Secondary Schools
- 6112 - Junior Colleges
- 6113 - Colleges, Universities, and Professional Schools
- 8131 - Religious Organizations

The following industries have a limited scope for the AIES. Data were collected from companies in these industries for a limited set of items, including company-level business characteristics and company classification (formerly collected in the Report of Organization); and key data items limited to revenue, annual payroll, first-quarter payroll, number of employees, and total operating expenses.

- 115 - Support Activities for Agriculture and Forestry
- 21 - Mining, Quarrying, and Oil and Gas Extraction
- 23 - Construction
- 55 - Management of Companies and Enterprises
- 81393 - Labor Unions and Similar Labor Organizations
- 81394 - Political Organizations

Sampling Frame

The AIES sampling frame is constructed from single-establishment and multiple-establishment employer companies. These companies are extracted from the Census Bureau's Business Register. The Business Register is a multi-relational database that contains a record for each known establishment located in the U.S. or one of its territories and has paid employees. An establishment is a single physical location where company transactions take place and for which payroll and employment records are kept.

To create the sampling frame, the records are extracted from the Business Register (as of December 2023) for all employer establishments located in the 50 states and the District of Columbia that are classified in the NAICS sectors in-scope to the AIES. For these establishments, revenue, payroll, employment, name, and address information are extracted from the Business Register. Other Business Register information, along with data from the 2022 Economic Census, is also added to the sampling frame for sample analysis purposes.

Sampling Unit

The sampling unit for the AIES is the company. To create the company sampling unit for multiple-establishment companies, the extracted establishment data are aggregated to the company level by tabulating the establishment data for all establishments associated with the company. For single-establishment companies, no aggregation is necessary to put the establishment on a company basis.

Sample Design

The measure-of-size (MOS) variable used for the AIES sample design is based on administrative annual payroll. Payroll data are available for most establishments on the AIES sampling frame and generally correlate positively with the AIES key data items.

Approximately 385,000 companies are selected in the sample. Of the sampled companies, approximately 48,500 are identified as self-representing (included with probability equal to 1) based on the diversity of their operations. These self-representing companies are defined as certainties.

For sectors 11, 21, 23, and 55, sampling occurs at the sector level. The remaining companies on the frame are stratified into broad strata defined by the 3-digit NAICS subsector. A zero MOS stratum is also defined for each sector level, containing companies active in the previous Economic Census but not active on the most recent vintage of the Business Register. For select sectors, the broad strata are further defined using:

- Type of operation (only in Wholesale Trade)
- Tax-exempt status (only in Educational Services; Health Care and Social Assistance; Arts, Entertainment, and Recreation; and Other Services).

Companies are further sub-stratified within each broad stratum by nesting geographic groups as follows:

- One multi-geography stratum for each unique broad stratum, containing all multi-unit companies that operate in two or more geographic categories within that broad stratum.
- Twenty-seven single-geography strata: one stratum for each unique broad stratum, cross-classified by geographic categories.

The geographic categories for stratification are specified in the table below. Each direct-use state defines a unique geographic stratum. Each "Balance of Region" stratum contains companies that operate in one or more of the states included in the collapsed geographic group.

Geographic (Strata) Categories for the AIES Sampling Design

Region	Total States	Direct-Use States	Balance of Region States
Northeast	9	Massachusetts, New York, New Jersey, Pennsylvania	5
Midwest	12	Illinois, Indiana, Michigan, Minnesota, Missouri, Ohio, Wisconsin	5
South	16 + DC	Florida, Georgia, Maryland, North Carolina, Tennessee, Texas, Virginia	9 + DC
West	13	Arizona, California, Colorado, Oregon, Washington	8

Sample Size Allocation

Allocation is performed independently by sector. The AIES requires a minimum stratum sample size of 5 and targets a maximum sampling fraction of 0.30. The maximum sampling fraction allowed is 0.50. Consequently, each geographic stratum with less than 15 companies is automatically designated as a certainty stratum with inclusion probability equal to 1.

In each zero MOS stratum, all companies are sampled with equal probability. A small sample is selected with a sampling rate of approximately 0.005. If 5 or fewer companies are in the zero MOS stratum, then all units will be sampled with probability equal to 1.

The remaining within-sector strata allocations are obtained in two steps. First, proportional allocation is used to allocate the sector sample size to the unique broad stratum based on the proportion of companies in the sector that fall in the broad stratum. Then, a power allocation is used to allocate the broad stratum sample size to each geographic substratum.

Once sample size allocation is complete, the inclusion probabilities for each company are calculated based on each company's contribution to the national industry (typically 6-digit NAICS) level estimates and geographic by 3-digit NAICS estimates. For small companies with similar MOS and inclusion probabilities, a data-driven weight trimming algorithm is used to determine the threshold for the minimum inclusion probability in each substratum.

Sample Selection

The AIES uses an unequal probability stratified sequential sampling design, also known as Chromy's method. This stratified sequential sampling is a without-replacement probability proportional to estimated size sampling method that allows for a within-strata sort, thereby implicitly offering an additional level of sub-stratification. To perform the sort, each company on the frame is assigned a single stratification industry corresponding to the company's lowest unique level of NAICS classification. Within each AIES sampling stratum, companies are sorted by stratification industry, state, and MOS. The estimated size is a company's inclusion probability.

Matrix Sampling for the Manufacturing Sector

With the AIES, there will be an ongoing focus on finding the right balance between satisfying user demands for expansive (and detailed) data without overburdening the sampled companies during collection. A matrix sampling design is a commonly used survey methodology technique designed to reduce respondent burden while allowing unbiased estimation (with a measurable "cost" of increased sampling variance). In a matrix (or split panel) sample, sets of questions are randomly assigned to different sampled units, so that each question has a known inclusion probability.

Companies in the AIES sample with establishments conducting business in the manufacturing sector are asked to provide information on a suite of items for each establishment. To reduce burden on noncertainty companies in the AIES sample (mostly single-establishment companies), the AIES introduces a matrix sampling design.

For the matrix design, the AIES certainty companies (included in the AIES with probability equal to 1) with establishments in the manufacturing sector are asked to provide data for all items for each establishment. For the remaining companies in the AIES sample with manufacturing activity, the previously defined strata are maintained for the matrix sample design. Manufacturing sector strata with fewer than 15 noncertainty (selected with probability less than 1) companies selected for the AIES collection are defined as certainty strata for the matrix design. Strata with more than or equal to 15 noncertainty companies selected in the AIES sample are split into two panels by conducting a simple random sample with a 1 in 2 sampling rate. The companies selected for the first panel will be asked the full set of manufacturing questions for each establishment, while the second panel will only be asked to provide data on four items: annual payroll, total revenue, first quarter payroll, and March 12 employment.

There are approximately 43,500 companies in the AIES sample with activity in the manufacturing sector. Of these, 33,000 companies will receive the full set of manufacturing items for each establishment, and the remaining 10,500 companies will only be asked for information on the previously listed four items at each establishment.

Frequency of Sample Redesign

The 2023 inaugural sample will be redesigned after two years. Subsequent sample redesigns are expected to occur every year.

Sample Maintenance

Periodically throughout the life of the AIES sample, updates to the sample are needed to account for new companies that come into existence (referred to as births), existing companies that become inactive or go out of business (referred to as deaths), companies or parts of companies that change ownership or association, and companies that go out of business and then become active again (referred to as reactivations). This effort helps ensure that the AIES sample properly represents the target population for any time in which estimates are produced.

Birth Sampling

A birth occurs when a new company comes into existence. Birth companies are identified using available administrative, classification, and payroll data on the Business Register. A sample of births is selected and added annually to the AIES sample using similar sampling methodology as was applied to the initial sample.

A birth frame of all active, new, and in-scope employer companies is constructed from the Business Register. This frame is the input to the birth sampling process.

For birth sampling, a similar sampling strata structure is used that is adapted for the small scale of the birth sample. The birth sampling geographic strata for single-establishments are generally defined by 3-digit NAICS and tax status with a flexible hierarchical collapsing geographic component that is initially defined by NAICS, tax status, type of operation code and division.

Depending on the number of companies in a stratum, the final geography chosen could be division, region, or none (i.e., the whole NAICS and tax status collapsed into one stratum that includes all geographies). The selection of the three geography levels was based on historic data.

Once the final sampling strata are defined, initial sector sample sizes are determined, based on approximately 2 percent of a sector's birth frame. Then, proportional and power allocation, like initial sampling, is performed independently for each sector. This allocation ensures that the minimum stratum sample size is 5, and the maximum sampling fraction in a noncertainty stratum is 0.50. Once sample size allocation is complete, like initial sampling, the inclusion probabilities for each company are calculated, and Chromy sampling is used to sample the births based on their inclusion probability. Note: However, unlike initial

sampling, we will not apply the "take-all" rule to small strata during birth sampling. While initial sampling classifies strata with 15 or fewer companies as certainties, birth sampling excludes this rule because the number of new births is significantly lower than the number of original sampling units.

Change of Status

Administrative data from the Business Register are also used to identify and extract status updates or changes for sampled units. Types of status changes for an existing sampled company may include deaths or deletes, reactivations, and ownership or structure changes such as divestitures, mergers, and spin-offs. Changes in status can result in additions or deletions of companies in the sample.

Experimental Tables

Experimental tables are released for the AIES and follow a separate methodology. To access these tables, please visit AIES Data page.

Data Collection

Data Items Requested

Depending on the company's industry and characteristics, AIES collects information in the following areas:

- Company characteristics, including employment, operating status, organizational change, and ownership information
- Company classification, including business activity
- Revenue, revenue by class of customer, source of revenue, taxes, contributions, gifts, grants, products, and e-commerce activity
- Operating expenses, including purchased services, payroll, benefits, rental payments, utilities, interest, resales, equipment, materials and supplies, research and development, and other detailed operating expenses (In SY24, research and development was collected at the company level.)
- Assets, including capital expenditures, robotic equipment, inventories, and depreciable assets

Additional information regarding the AIES content can be found in the AIES Content Summary.

Type of Request

Response is required by law. The AIES OMB control number is 0607-1024. The AIES collection is authorized by Title 13 U.S.C. Sections 131 and 182. Response to the AIES is mandatory per Sections 224 and 225 of Title 13 U.S.C.

Frequency and Mode of Contact

The AIES is collected annually.

Companies are mailed with instructions to provide responses online via a spreadsheet. Due date reminders and nonresponse follow-ups occur by mail or e-mail during the collection period. Phone calls are also utilized to follow up with companies that fail to respond.

Data Collection Unit

All single-establishment companies and multi-establishment companies are asked to report data at the company, industry and establishment levels. To ensure precision across these levels—especially for complex firms—the AIES utilizes the Kind of Activity Unit (KAU).

The KAU is a statistical subdivision of an enterprise that aggregates all entities (or parts of entities) engaged in a specific, homogeneous economic activity.

In AIES, the KAU serves as the reporting unit when data must be partitioned by industry rather than strictly by geography. This structure allows multi-industry firms to accurately allocate receipts and expenses, ensuring that reported data reflects what is being produced or sold, even when a company's reporting units span multiple industrial classifications.

Compilation of Data

Editing

Data captured in the AIES must be edited to identify and correct reporting errors. After an establishment has been assigned a valid industry code, the data edits evaluate the response data for consistency and validity, for example, ensuring that employment data are consistent with payroll or revenue data. To ensure accuracy, additional checks compare current year's data to data reported in previous censuses, annual surveys, or from administrative sources.

Nonresponse

Nonresponse is defined as the inability to obtain requested data from an eligible survey unit. Researchers often differentiate between two types of nonresponse. Unit nonresponse is the inability to obtain any of the substantive measurements about a unit. In most cases of unit nonresponse, the Census Bureau was unable to obtain any information from the survey unit after several attempts to elicit a response. Item nonresponse occurs either when a question is unanswered or unusable.

Nonresponse Adjustment and Imputation

To adjust for nonresponse, the AIES performs imputation, which is the procedure for determining a value for a specific data item where the response is missing or unusable. The AIES utilizes a variety of imputation methods, including regression, substitution, and nearest-neighbor hot deck imputation.

Tabulation Unit

A tabulation unit contains the data used for tabulation creation. For the AIES, the tabulation unit will vary. Estimates will be created using either the establishment or the Kind of Activity Unit (KAU) level tabulation unit.

Estimation and Sampling Variance

Estimation

Total estimates are computed using the Horvitz-Thompson (HT) estimator, which is the sum of weighted reported and imputed data for all selected sampling units that meet the sample canvass and tabulation criteria. These data come from all in-scope establishments using reported data collected from the AIES, plus administrative records data or imputed data for nonrespondents. The weight for a given sampling unit is the reciprocal of its probability of selection into the AIES sample.

Sampling Error

The sampling error of an estimate based on a sample survey is the difference between the estimate and the result that would be obtained from a complete census conducted under the same survey conditions. This error occurs because characteristics differ among sampling units in the population and only a subset of the population is measured in a sample survey. The sample used in this survey is one of many samples of the same size that could have been selected using the same sample design. Because each unit in the sampling frame has a known probability of being selected into the sample, it is possible to estimate the sampling variability of the survey estimates.

Common measures of the variability among these estimates are the sampling variance, the standard error, and the coefficient of variation (CV), which is also referred to as the relative standard error (RSE). The sampling variance is defined as the squared difference, averaged over all possible samples of the same size and design, between the estimator and its average value. The standard error is the square root of the sampling variance. The CV expresses the standard error as a percentage of the estimate to which it refers. For example, an estimate of 200 units that has an estimated standard error of 10 units has an estimated CV of 5 percent. The sampling variance, standard error, and CV of an estimate can be estimated from the selected sample because the sample was selected using probability sampling. Measures of sampling variability, such as the standard error and CV, are estimated from the sample and are also subject to sampling variability. It is also important to note that the standard error and CV only measure sampling variability. They do not measure any systematic biases in the estimates.

The Census Bureau recommends that individuals using these estimates incorporate sampling error information into their analyses, as this could affect the conclusions drawn from the estimates.

Estimates of variances for the HT estimates are obtained via the doubled half bootstrap (DHB) method, see Antal and Tillé, 2014 (view at Springer). Bootstrapping requires repeatedly sampling with replacement from the survey sample to generate an approximation of the population. For each bootstrap replicate sample, the DHB first takes a Poisson sample based on the design inclusion probabilities. If more units are needed to meet sample size, then half of the remaining units are randomly selected using simple random sampling. The resulting bootstrap sample consists of the units in the Poisson sample being included once and those in the simple random sample being included twice.

For the AIES, four hundred DHB replicate samples are taken independently within stratum. Each of them yields an estimate of the domain total. Thus, the bootstrap variance estimate is obtained as the empirical variance of the replicate estimates:

$$\hat{s}_d^2 = \frac{1}{400-1}\sum_{b=1}^{400}\left(\hat{Y}_{db}^{Dir} - \hat{\bar{Y}}_{db}\right)^2, \qquad b = 1, \ldots, 400$$

Where $\hat{Y}_{db}^{Dir}$ is the bth bootstrap replicate estimate of the domain total and $\hat{\bar{Y}}_{db} = \frac{1}{400}\sum_{b=1}^{400}\hat{Y}_{db}^{Dir}$.

The variance of ratio estimates is calculated using this same method by substituting ratios of replicate estimates.

Confidence Interval

The sample estimate and an estimate of its standard error allow us to construct interval estimates with prescribed confidence that the interval includes the average result of all possible samples with the same size and design. To illustrate, if all possible samples were surveyed under essentially the same conditions, and an estimate and its standard error were calculated from each sample, then:

- Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average estimate derived from all possible samples.
- Approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average estimate derived from all possible samples.

The margin of error (MOE) associated with the 90 percent confidence interval is the product of 1.645 and the estimated standard error.

Nonsampling Error

Nonsampling error encompasses all factors other than sampling error that contribute to the total error associated with an estimate. This error may also be present in censuses and other nonsurvey programs. Nonsampling error arises from many sources: inability to obtain

information on all units in the sample; response errors; differences in the interpretation of the questions; mismatches between sampling units and reporting units, requested data and data available or accessible in respondents' records, or regarding reference periods; mistakes in coding or keying the data obtained; and other errors of collection, response, coverage, and processing.

Although no direct measurement of nonsampling error is obtained, precautionary steps are taken in all phases of the collection, processing, and tabulation of the data to minimize its influence.

Precise estimation of the magnitude of nonsampling errors requires special research or access to independent data, and, consequently, the magnitudes are often unavailable.

The Census Bureau recommends that individuals using these estimates factor in this information when assessing their analyses of these data, as nonsampling error could affect the conclusions drawn from the estimates.

Quality Measures

CVs are used as one of the quality measures. Associated flags indicate the quality of data. Specifically, the flags are:

- "v" to indicate an estimated coefficient of variation between 30.0 percent and 60.8 percent
- "w" to indicate an estimated coefficient of variation exceeding 60.8 percent, where the estimate is not significantly different from zero at a 90 percent confidence level.

The **Unit Response Rate (URR)** is defined as the percentage of reporting units in the statistical period, based on unweighted counts, that were eligible for data collection or of unknown eligibility that responded to the survey. Revenue or payroll at the establishment level must be reported for the unit to be counted as a respondent. URRs are indicators of the performance of data collection for obtaining usable responses.

The **Total Quantity Response Rate (TQRR)** is defined as the percentage of the estimated (weighted) total of a given data item reported by the active tabulation units in the statistical period or from sources determined to be equivalent-quality-to-reported data. The TQRR is an item-level indicator of the “quality” of each estimate. In contrast to the URR, these weighted response rates are computed for individual data items, so that a survey may produce several TQRRs per statistical period and release. The TQRR is a weighted measure that takes the size of the tabulation unit into account as well as the associated sampling parameters. To compute the TQRR for a particular estimate, it is necessary to determine the source of the final tabulated value of the associated data item for each tabulation unit. This value could be directly obtained from respondent data, indirectly obtained from other equivalent quality data sources, or imputed.

For additional information about response rates, see the Census Bureau's Statistical Quality Standard D.3.,
<https://www.census.gov/about/policies/quality/standards/appendixd3b.html>

2024 URR and TQRR Metrics by Sector

2017 NAICS Sector	Title	2024 URR	2024 TQRR Revenue Range
11	Agriculture, Forestry, Fishing and Hunting	80.4	50% to less than 60%
21	Mining, Quarrying, and Oil and Gas Extraction	66.2	50% to less than 60%
22	Utilities	78.6	50% to less than 60%
23	Construction	68.6	70% to less than 80%
31-33	Manufacturing	58.4	50% to less than 60%
42	Wholesale Trade	66.2	60% to less than 70%
44-45	Retail Trade	61.6	50% to less than 60%
48-49	Transportation and Warehousing	60.8	50% to less than 60%
51	Information	71.3	50% to less than 60%
52	Finance and Insurance	84.6	70% to less than 80%
53	Real Estate and Rental and Leasing	64.0	50% to less than 60%
54	Professional, Scientific, and Technical Services	65.5	50% to less than 60%
55	Management of Companies and Enterprises	71.8	80% to less than 90%
56	Administrative and Support and Waste Management and Remediation Services	67.0	50% to less than 60%
61	Educational Services	64.4	70% to less than 80%
62	Health Care and Social Assistance	64.1	50% to less than 60%
71	Art, Entertainment, and Recreation	65.6	70% to less than 80%

72	Accommodation and Food Services	62.5	70% to less than 80%
81	Other Services (except Public Administration)	58.0	70% to less than 80%
AIES	Overall	65.9	60% to less than 70%

Linking Samples

The collection of point estimates for a particular measure across time for a given industry level is referred to as a time series. To maintain and link the time series of estimates for each industry from the legacy surveys to the AIES estimates, a sample linking operation is performed. For most industries, revenue estimates from the new sample for reference year 2022 and subsequent years are linked to the legacy survey estimates by multiplying the estimates from the AIES by a ratio. This ratio, called the carry-forward factor, is calculated as follows:

- The numerator is the 2022, census-adjusted (based on the 2022 Economic Census) revenue estimate, for a given industry level, from the legacy surveys.

- The denominator is the 2022 revenue estimate from the new AIES sample for the same industry level.

- Note that prior to calculating the carry-forward factors, estimates from the legacy surveys are converted from millions of dollars to thousands of dollars.

- The carry-forward factor for a given industry level is then applied to the 2022 and 2023 AIES revenue estimates, ensuring the 2022 AIES revenue estimates match the final published 2022 revenue estimates from the legacy surveys. For most industries, the new AIES revenue estimates are implicitly benchmarked to the 2022 Economic Census because they are linked to previously benchmarked estimates from the prior sample.

- For the remaining key data items within a given industry, the same carry-forward factor is applied to the AIES estimates to preserve the relationship between the item and revenue, while still reflecting the revised revenue level derived from sample linking. Exceptions are noted below.

    - For NAICS 485111, 485112, 485113, 485119, 488510, and 541840, revenue estimates from the AIES will not be linked to revenue estimates from the legacy surveys. However, a carry-forward factor based on the 2022 Economic Census revenue estimate will be calculated for each industry level as follows:

        - The numerator is the published 2022 Economic Census revenue estimate for the given industry level.
        - The denominator is the 2022 revenue estimate from the new AIES sample, for the same industry level.
        - The carry-forward factor for the given industry level is then applied to the 2022 and 2023 AIES revenue estimates, as well as most remaining key data items.

    - For all sectors, the carry-forward factor will not be applied to capital expenditures or robotics-related items. Capital expenditure and robotics estimates from the AIES will not be linked to legacy survey estimates due to differences in scope.

    - For Newspaper Publishers (NAICS 511110), there exists revenue detail data items for which new time series will be established beginning with the 2023 AIES. Due to questionnaire differences, the AIES estimates will not be linked to the time series of these items from the legacy surveys. However, the revenue carry-forward factor for the industry will be applied to the 2023 revenue detail estimates. This includes:

        - Revenues for Advertising Space and Breakdowns of Revenue by Media Type.

    - For Wholesale (Sector 42) and Retail (Sector 44-45), the carry-forward factor used to sample link inventories estimates is calculated as follows:

        - The numerator is the 2022 end-of year inventories for a given industry level, from the legacy survey.
        - The denominator is the 2023 beginning inventories estimate from the new AIES sample for the same industry level.
        - The carry-forward factor for the given industry level is then applied to the 2023 and 2024 AIES end-of-year inventories estimates as well as to other detail inventories items.

- Additional sample linking processing was performed on select items in sectors in Wholesale (Sector 42), Retail (Sector 44-45), and Health Care and Social Assistance (Sector 62) in order to adjust historical estimates for 2018 through 2022 and maintain a time series. For these items, prior year data was not available in AIES to use in the sample linking process. Instead, a geometric mean was used to calculate a revised 2022 value by using the relationship of the item to revenue from both samples. This geometric mean was computed at the previously published detail NAICS levels as the square root of the product of two ratios. The first ratio was the 2022 legacy survey estimate for the item divided by the 2022 legacy revenue. The second ratio was the 2023 AIES estimate for the item divided by the 2023 AIES revenue. This geometric mean was then applied to the 2022 census-adjusted revenue estimate to get a revised 2022 value for the item.

- Then the estimates for 2017 through 2022 from the legacy sample are input into the benchmarking program. Using this program, the estimates for 2018 through 2022 for each detailed industry are revised in a manner that:

    - Uses the 2017 legacy survey estimate, along with revised 2022 value calculated using the geometric means as constraints.
    - Uses the Sequential Least Squares Programming optimization method to minimize the sum of squared differences between the year-to-year changes of the legacy estimates and revised estimates for 2018 through 2022.
    - Estimates at aggregate industry levels are computed by summing the resulting estimates for the appropriate detailed industries comprising the aggregate.

Benchmarking

Benchmarking adjusts the level of a given series to the levels (referred to as benchmarks) from a less frequent data source that is considered to be of better quality, while attempting to minimize revisions to the period-to-period changes from the more frequent series. In this way, we produce consistent time series and attempt to reduce the effects of sampling and nonsampling errors in the original, more frequent series.

Historical estimates from the legacy annual programs were benchmarked to the 2022 Economic Census on a 2017 NAICS basis. For information on how AIES estimates are tied to these legacy levels, refer to the Linking Samples section. Estimates from AIES will be benchmarked to the 2027 Economic Census when it becomes available.

Disclosure Avoidance

Disclosure is the release of data that reveals information or permits deduction of information about a particular survey unit through the release of either tables or microdata. Disclosure avoidance is the process used to protect each survey unit's identity and data from disclosure. Using disclosure avoidance procedures, the Census Bureau modifies or removes the characteristics that put information at risk of disclosure. Although it may appear that a table shows information about a specific survey unit, the Census Bureau has taken steps to disguise or suppress a unit's data that may be at risk of disclosure while ensuring the results are still useful.

Cell suppression is a disclosure avoidance technique that protects the confidentiality of individual survey units by withholding cell values from release and replacing the cell value with a symbol, usually a "D". If the suppressed cell value was known, it would allow one to estimate an individual survey unit's response too closely.

The cells that must be protected are called primary suppressions.

To be sure the cell values of the primary suppressions cannot be closely estimated by using other published cell values, additional cells may also be suppressed. These additional suppressed cells are called complementary suppressions.

Values for cells that are not suppressed remain unchanged. Before the Census Bureau releases data, computer programs and analysts ensure primary and complementary suppressions have been correctly applied.

The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data (Project No. P-7529180, Disclosure Review Board (DRB) approval number: CBDRB-FY26-0304).

For more information on disclosure avoidance practices, see FCSM Statistical Policy Working Paper 22.

Legacy Surveys

For more information on the legacy surveys, please visit https://www.census.gov/programs-surveys/aies/about.html