

**Department of Commerce  
U.S. Census Bureau  
OMB Information Collection Request  
Annual Retail Trade Survey  
OMB Control No. 0607-0013**

**PART B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS**

**1. Universe and Respondent Selection**

**Target Population**

The target population of the Annual Retail Trade Survey (ARTS) consists of all United States firms with paid employees that are primarily engaged in retail trade, as defined by the 2012 North American Industry Classification System (NAICS). Firms without paid employees, or nonemployers, are included in the estimates through imputation or administrative data provided by other federal agencies.

**Sampling Frame**

The ARTS sampling frame has two types of sampling units represented: single-establishment firms and multiple-establishment firms. The information used to create these sampling units was extracted from data collected via the 2012 Economic Census and from establishment records contained on the Census Bureau's Business Register (as of October 2015). The Business Register (BR), which is the Census Bureau's master business list, contains basic economic information for more than 7.4 million employer businesses and over 22.5 million nonemployer businesses. It is updated through direct data collections and administrative records from other federal agencies.

To create the sampling frame, the Census Bureau extracts the records for all employer establishments located in the United States and classified in the retail trade sector, as defined by the 2012 NAICS. For these establishments, the agency extracts sales, end-of-year inventories, payroll, employment, name and address information, primary identifiers, and associated employer identification numbers (EINs).

To create the sampling units, the agency sums the establishment data for all retail establishments associated with the same firm identifier. No aggregation is necessary to put single-unit establishment information on a firm basis. Thus, the sampling units

created for single-unit firms simultaneously represent establishment and firm information.

### **Stratification**

The sample for ARTS uses a stratified, one-stage design with primary strata defined by industry. There are 85 primary strata. The primary strata are sub-stratified into 4, 7, 10, or 13 annual sales size strata. Sampling units expected to have a large effect on the precision of the estimates are selected “with certainty.” This means they are sure to be selected and will represent only themselves (i.e., have a selection probability of 1 and a sampling weight of 1). Within each industry stratum we determine a substratum boundary (or cutoff) that divides the certainty units from the noncertainty units, based on a statistical analysis of data from the 2012 Economic Census. We also use this analysis to determine the number of size substrata for each industry stratum and to set preliminary sampling rates needed to achieve specified sampling variability constraints on sales estimates for different industry groups.

### **Sample Selection**

Sample selection is a two-step process and begins by identifying the firms selected with certainty. If a firm's annual sales or end-of-year inventories are greater than the corresponding certainty cut-off, that firm is selected into the ARTS sample with certainty. In the second step, all firms not selected with certainty are subjected to sampling. To be eligible for the initial sampling, a firm had to have nonzero payroll in 2014. Within each noncertainty stratum, we select a simple random sample of firms without replacement.

For the ARTS, a new sample is selected approximately every five to seven years. The ARTS introduced a new sample for survey year 2016. The survey will continue to use this sample throughout this three-year clearance period. The initial sample consisted of roughly 16,500 firms. This count changes over the course of a sample due to sample maintenance activities, which are discussed in the Sample Maintenance section below.

### **Sample Maintenance**

The sample is updated on a quarterly basis to represent EINs issued since the initial sample selection. These new EINs, called births, are EINs recently assigned by the Internal Revenue Service (IRS) that have an active payroll filing requirement on the IRS Business Master File (BMF). An active payroll filing requirement indicates that the EIN is required to file payroll for the next quarterly period. The Social Security

Administration (SSA) attempts to assign an industry classification to each new EIN. EINs with an active payroll filing requirement on the IRS BMF are said to be “BMF active,” and EINs with an inactive payroll filing requirement are said to be “BMF inactive.” The Census Bureau samples EIN births on a quarterly basis using a two-phase selection procedure. To be eligible for selection, a birth must either have no industry classification or be classified in an industry within the scope of the Service Annual Survey (SAS), the Annual Wholesale Trade Survey (AWTS), or the ARTS. Additionally, it must meet certain criteria regarding its quarterly payroll. In the first phase, the agency stratifies births by broad industry groups and a measure of size based on quarterly payroll. A relatively large sample is drawn and canvassed to obtain a more reliable measure of size (consisting of revenue in two recent months) and a new or more detailed industry classification code. Births are contacted by telephone if they have not returned their electronic questionnaire within 30 days. Using this more reliable information, in the second phase, the Census Bureau subjects the selected births from the first phase to probability proportional-to-size sampling with overall probabilities equivalent to those used in drawing the initial ARTS sample from the original October 2015 BR. Because of the time it takes for a new employer firm to acquire an EIN from the IRS and the time needed to accomplish the two-phase birth-selection procedure, the Census Bureau adds births to the sample approximately nine months after they begin operation.

Similarly, each quarter the agency checks the BR to determine if any EINs on the survey have become BMF inactive. Typically, the Census Bureau does not canvass BMF inactive EINs during the reference year. Likewise, if any EIN on the survey was BMF inactive in a previous reference year or was part of an inactive sampling unit in the survey and is now currently BMF active on the BR, it is included in the canvass again. In both cases, the ARTS only tabulates data for the portion of the reference year that these EINs reported payroll to the IRS.

The sample is also updated to reflect mergers, acquisitions, divestitures, splits, and other changes to the business universe throughout the year. (The Census Bureau learns about these changes through a variety of means, including directly from respondents, and updates the sample accordingly as soon as the information is obtained. For example, the ARTS asks about organizational changes on each of its questionnaires. When the Census Bureau analyst in charge of a company becomes aware of a structural change, he or she will immediately take the appropriate sample maintenance actions to account for this change.) In general, any new establishments that a firm acquires, even if under new or different EINs, are included in the sample with the same sampling status as the original firm (i.e., with the same initial sampling weight). For noncertainty firms, additional evaluation may be done in some instances to determine the feasibility of adding the new establishments (by evaluating the effect of the new establishments on the industry estimates).

## Response Rates

The unit response rate is defined as the percentage of active 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. It is an indicator of the performance of data collection for obtaining usable responses. The ARTS calculates and monitors the unit response rate during each processing cycle. To be classified as a response, the respondent for the reporting unit must have provided sufficient data, and the data must satisfy all the edits. Specifically, to be considered a respondent for the ARTS a retailer must provide sales, end-of-year inventories, purchases, e-commerce (if the company is purely e-commerce), or total operation expenses.

Since the current sample was launched in the 2016 survey year, there have been three ARTS releases (years 2016-2018). Currently, the 2019 ARTS data are being collected and processed. The reporting instrument for the 2019 ARTS will be deactivated on November 30, 2020. The table below lists the unit response rate for the entire retail trade sector (sectors 44-45) for each year. (For the 2016-2018 ARTS, the value represents the unit response rate when the data were released to the public for that survey year. For the 2019 ARTS, it is the value as of August 28, 2020.)

2019 <sup>1</sup>	2018	2017	2016
57.3%	64.1%	63.8%	72.3%

<sup>1</sup>As of 9/25/20 (collection ends on 11/30/20)

Response rates for 2019 have been particularly challenging due to the impact of the COVID-19 pandemic on businesses and specifically many different types of retailers whose doors have been closed during this crisis. Although some stores were permitted to remain open, the corporate offices that are responsible for completing the ARTS have been closed or their staff are unable to access critical records to complete the survey. Data collection efforts for the ARTS will continue through November 30, 2020 and Census Bureau staff are likely to receive a few more responses as analysts finalize their industries through December. As many states, have recently begun to lift closure mandates, we fully expect the final response rate at the time of release to be in excess of 60%. Census Bureau staff will continue to monitor these rates, make adjustments to the data collection strategy and if response rates do not rebound for the ARTS then a nonresponse bias analysis may be taken into consideration.

For the ARTS, the Census Bureau also computes and monitors total quantity response

rates. This type of response rate 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 total quantity response rate is an item-level indicator of the “quality” of each estimate. In contrast to the unit response rate, these weighted response rates are computed for individual data items. Thus, there are several total quantity response rates per statistical period and release. The total quantity response rate is a weighted measure that takes the size of the tabulation unit into account as well as the associated sampling parameters. To compute the total quantity response rate 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 reported-equivalent data sources, or imputed.

## **2. Procedures for Collecting Information**

In addition to the following information, please refer to the Sampling Frame, Stratification, Sample Selection, and Sample Maintenance segments of the Universe and Respondent Selection section above.

The initial letter to firms respondents identifies the survey name, due date, instructions for accessing the survey, authority for collection, and burden estimate. It also provides a telephone number for those needing assistance. The initial and follow-up mailings instruct firms to access the Census Bureau’s Respondent Portal (<https://portal.census.gov>). The Census Bureau will also periodically send an e-mail reminder, or eBlast, to firms. These e-mail reminders also provide the same instructions. Once the user arrives at <https://portal.census.gov>, the individual will create a new account (i.e., register) or sign in to an existing account. The Respondent Portal allows users to add and view all current Census Bureau surveys assigned to them. Using the Respondent Portal, firms can access the online reporting system called Centurion (by clicking “Report Now” for the appropriate survey). A telephone follow-up is scheduled for nonrespondents as well. For the ARTS, electronic reporting is the only advertised reporting option. Nevertheless, the Census Bureau will accept data submitted through other methods. For example, if a firm does not have access to the Internet, the Census Bureau can arrange for the business to provide its data to an analyst via telephone.

The quarterly birth/death process, removal of out-of-scope cases, and restructuring of reporting units to account for organizational changes (e.g., mergers, acquisitions, etc.) will inevitably cause the number of respondents to fluctuate over time. Historically, the number of respondents typically decreases each year within a given sample. This trend has held true for this sample so far (survey years 2016, 2017, 2018, and 2019). The table

below shows the number of firms at the time of the initial mailing for the current sample by year.

2019	2018	2017	2016
16,813	17,971	18,401	20,067

### **Analysis**

Census Bureau analysts review the company-level and industry-level data for accuracy. The analysis performed includes (but is not limited to) comparing a company's data for the current year against its historical data, investigating differences between the industry-level estimates generated from the Monthly Retail Trade Survey (MRTS) and the ARTS, and reviewing records that are displaying unusual behavior (e.g., operating expenses that are greater than sales). When analysts are unable to validate certain information through their own research, they sometimes have to contact respondents to obtain an explanation.

### **Nonresponse**

The Census Bureau imputes data using survey data and administrative data for unit nonresponse, item nonresponse, and for responses that fail computer or analyst edits.

### **Estimation Procedures**

Total estimates are computed using the Horvitz-Thompson estimator (i.e., the sum of weighted reported or imputed data) for all selected sampling units that meet the sample canvass and tabulation criteria. The weight for a given sampling unit is the reciprocal of its probability of selection into the ARTS sample. These estimates are then adjusted through a benchmarking process. The Census Bureau estimates variances for published statistics using the method of random groups.

### **Linking Samples**

As previously mentioned, the current sample was introduced with the 2016 ARTS. This sample is designed to produce estimates based on the 2012 NAICS. All published estimates from the 2015 ARTS were restated from 2007 NAICS definitions to 2012 NAICS definitions.

In order to maintain the time series for each industry, an operation is performed to link estimates from the prior and new samples. For the linking operation to occur, two years of data were collected (2015 and 2016) from units in the new sample.

Sales estimates from the new sample for reference year 2015 and subsequent years are linked to the restated prior sample estimates by multiplying the Horvitz-Thompson estimates from the new sample by a ratio. The ratio is calculated as follows:

- The numerator is the 2015 published, census-adjusted (based on the 2012 Economic Census) sales estimate for employer firms for the industry restated on a 2012 NAICS basis from the prior sample.
- The denominator is the 2015 Horvitz-Thompson sales estimate for employer firms for the industry on a 2012 NAICS basis from the new sample.

The resulting sales estimates (called “modified” sales estimates) are implicitly benchmarked to 2012 Economic Census results via this linking procedure.

The following method is used to produce modified estimates for end-of-year inventories, purchases, and sales tax. First, the sales ratio described above is multiplied by the Horvitz-Thompson estimate for the given item for 2015 and subsequent years. Then, the published estimates for 2010 through 2015 from the prior sample are input into the benchmarking program. Using this program, the estimates for 2011 through 2015 for each detailed industry are revised in a manner that:

- uses the benchmarked estimate for 2010 from the prior sample as a constraint, resulting in no revision to the 2010 estimate.
- uses the modified estimate for 2015 from the new sample as a constraint.
- minimizes the sum of squared differences between the year-to-year changes of the input and revised estimates for 2011 through 2015.

A similar method is used to produce "modified" estimates for total operating expenses and e-commerce. E-commerce benchmarks with mostly 3-digit NAICS aggregate industry levels, but splits out a few extra levels of detail, most notably NAICS 4541 (Electronic Shopping and Mail-Order Houses). For total expenses, the benchmarked estimate for 2012, instead of 2010, from the prior sample is used as a constraint.

Modified merchandise lines sales and e-commerce estimates within the NAICS 4541 industry group are obtained in a similar way to e-commerce, with additional raking to total sales and e-commerce for NAICS 4541 to ensure the lines properly sum to the totals.

Modified estimates at aggregate industry levels are computed by summing the modified estimates for the appropriate detailed industries comprising the aggregates.

The ARTS estimates will also be benchmarked to the 2017 Economic Census results in the upcoming 2019 release.

## **Benchmarking**

Results of the 2017 Economic Census are used to benchmark the ARTS estimates. Due to the uncertainties of a new sample, some details described herein could change.

Sales estimates are input to the benchmarking program and are revised in a manner that:

- uses the 2017 Economic Census sales total as a constraint, along with the existing 2012 modified sales estimate, which is already linked to the 2012 Economic Census.
- minimizes the sum of squared differences between the year-to-year changes of the input and revised estimates for 2013 through 2019.

A similar method to the one for adjusting sales is used to adjust estimates for end-of-year inventories, purchases, operating expenses, e-commerce, and sales tax. Each of these items are revised in the following manner:

- 2012 and 2017 modified estimates are multiplied by the ratio of benchmarked sales divided by modified sales for the same year.
- Modified estimates for each item are input into the benchmarking program using the two constraints calculated above.
- The benchmarking program minimizes the sum of squared differences between the year-to-year changes of the input and revised estimates for 2013 through 2019.
- For merchandise lines sales estimates within the NAICS 4541 industry group, the 2013 through 2019 modified merchandise lines sales estimates are multiplied by the ratio of benchmarked sales divided by modified sales for the same year. For merchandise lines e-commerce estimates, the 2013 through 2019 modified merchandise lines e-commerce estimates are multiplied by the ratio of benchmarked e-commerce estimates divided by modified e-commerce for the same year.

Benchmarked estimates at aggregate industry levels are computed by summing the benchmarked estimates for the appropriate detailed industries comprising the aggregates.

### **Quality Suppressions**

Estimates can be suppressed from publication for quality reasons. In particular, an estimate with a coefficient of variation greater than 30 percent, with a total quantity response rate less than 50 percent, or with other concerns about data quality is suppressed from publication (unless the estimate has consistently been published for prior years and the coefficient of variation and total quantity response rate are acceptably close to the thresholds). A suppressed estimate and its corresponding measure of sampling variability are replaced with an "S" in the published tables when this occurs. For a description of the



Census Bureau's Standards for Releasing Information Products, see <https://www.census.gov/about/policies/quality/standards/standardf1.html>.

### **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.

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 values with a symbol ("D"). If the suppressed cell values were known, one would be able to estimate an individual survey unit's data too closely.

The cells that must be protected are called primary suppressions. To make 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.

The process of suppression does not usually change the higher-level totals. 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 ARTS will use automated cell suppression for disclosure avoidance with the 2019 release scheduled for January of 2021.

### **3. Methods to Maximize Response**

Firms are notified of their obligation to report via a mailed letter. This letter provides a telephone number that companies can call if they have any questions or concerns. They receive at least 30 days to complete the ARTS. If needed, businesses can request time extensions through the Respondent Portal. They can send secure messages to the Census Bureau using the Respondent Portal as well.

Approximately four weeks after the initial mailing, the Census Bureau provides nonrespondents with a due date reminder letter. One week after this, a due date reminder e-mail is transmitted to companies that have not reported. Once the due date has passed, delinquent firms receive their first follow-up letter. This occurs about three weeks after the due date reminder e-mail and about two weeks after the due date. About two weeks after the follow-up letter is distributed, the first follow-up e-mail is sent to nonresponse

cases. Three weeks later, outstanding firms receive their second follow-up letter. The Census Bureau sends the second follow-up e-mail to delinquent firms roughly four weeks after this. The agency then attempts to reach the nonrespondents via telephone about one week after the second follow-up e-mail. The telephone follow-ups are conducted over a period of approximately four weeks. One week after the telephone follow-ups have concluded, the third e-mail follow-up is sent. Two weeks later, a third follow-up letter is sent to respondents. Delinquent firms receive their fourth follow-up e-mail approximately three weeks later. The fifth follow-up e-mail is transmitted roughly four weeks after the fourth follow-up e-mail.

Throughout the year, the Census Bureau monitors the survey response rates (previously discussed in the Universe and Respondent Selection section) to ensure publication standards are satisfied. If the response rates fail to meet publication requirements, additional follow-ups, such as analyst phone calls or follow-up e-mails, may be performed.

As discussed in Part A (Project Schedule section), the Odyssey companies, which are a subset of the ARTS respondents, do not share the same follow-up schedule as the cases outlined above. Nevertheless, the general follow-up approach will be the same for these companies (i.e., letters, e-mails, and telephone calls used throughout the survey year to acquire responses). The only exception to this are Full-Service Account Manager (FSAM) companies who will only receive phone calls from FSAM members.

Samples of some of the correspondence mentioned in this section can be viewed in Attachment 2. (Please note that these are samples only at this time and the correspondence is subject to change.)

#### **4. Tests of Procedures or Methods**

The Census Bureau checks the procedures employed in every phase of the ARTS – from mailout and data capture to data editing and publication. Employees test the systems used to ensure they are programmed as expected and meet expected specifications. The staff also regularly reviews the guiding logic behind various aspects of the ARTS to confirm that logic is still relevant in the current economy.

The ARTS will continue to use the same sample and collect virtually the same information as the previous three-year clearance window with the exception of the removal of the accounts receivable information. The information received from respondents on all other portions of the questionnaire have not revealed issues regardless, the staff still check the worksheets for accuracy, relevance, and layout each year. During this three-year window, Centurion will still be the online reporting tool for respondents.

This system is tested internally each survey cycle to make sure the correct content is displayed. Additionally, the staff ensure the data are properly fed into the internal processing system. The data edits, imputation, and estimation are also tested annually to confirm they are functioning properly. The staff reviews the output from the ARTS publication tables program to determine if the tables are displaying the correct values and suppressions/disclosures. The general methodology employed for this three-year period will also be the same as the methodology used in the previous three-year clearance window.

## **5. Contacts for Statistical Aspects and Data Collection**

For questions regarding the planning and implementation of this survey, please contact Chris Savage, Chief of the Retail Trade Branch (Economy-Wide Statistics Division, U.S. Census Bureau), at 301-763-4834 or [John.C.Savage@census.gov](mailto:John.C.Savage@census.gov).

For inquiries about the methodology for this survey, please contact Deanna Weidenhamer, Chief of the Retail and Wholesale Surveys Statistical Methods Branch (Economic Statistical Methods Division, U.S. Census Bureau), at 301-763-7186 or [Deanna.L.Weidenhamer@census.gov](mailto:Deanna.L.Weidenhamer@census.gov).

## **Attachments to the Supporting Statement**

1. Bureau of Economic Analysis Support Letter
2. Initial Letters and Follow-up Correspondence Examples
3. Centurion (Reporting Instrument) Screenshots
4. ARTS Worksheets
5. Legal Authority for the ARTS - Title 13, United States Code