

PART B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Universe and Respondent Selection

Sampling Frame: The sampling frame used for the AWTS has two types of sampling units: Employer Identification Numbers (EINs) and large, multiple-establishment firms. Both sampling units represent clusters of one or more establishments owned or controlled by the same firm. The information used to create these sampling units was extracted from data collected as part of the 2002 Economic Census and from establishment records contained in the Census Bureau's Business Register as updated to December 2004.

To create the sampling frame we extract the records for all employer establishments located in the United States and classified in the Wholesale Trade sector as defined by the 2002 NAICS. For these establishments we extract sales, payroll, employment, name and address information, wholesale type of operation code, as well as primary identifiers and, for establishments owned by multi-unit firms, associated EINs. We use the type of operation code to distinguish between different types of wholesale establishments. To create the sampling units for multi-unit firms, we aggregate the economic data of the establishments owned by these firms to an EIN level by tabulating the establishment data for all wholesale establishments associated with the same EIN. Similarly, we aggregate the data to a multi-unit firm level by tabulating the establishment data for all wholesale establishments associated with the same alpha number. In some cases, a multi-unit firm has establishments active in more than one wholesale type of operation (merchant wholesalers excluding manufacturers' sales branches and offices and agents, brokers, and electronic markets). In these situations, separate EIN level and firm level sampling units are created for each type of operation. No aggregation is necessary to put single-unit establishment information on an EIN basis or a firm basis. Thus, the sampling units created for single-unit firms simultaneously represent establishment, EIN, and firm information.

Sample Design: The sample for AWTS consists of three separate samples: (1) a sample of merchant wholesalers, excluding manufacturers' sales branches and offices, (2) a sample of manufacturers' sales branches and offices, and (3) a sample of agents, brokers, and electronic markets. The AWTS uses a stratified, one-stage design with primary strata defined by industry. There are 59 primary strata: 40 from the merchant wholesale sample, excluding manufacturers' sales branches and offices, 17 from the manufacturers' sales branches and offices sample, and 2 from the agents, brokers, and electronic markets sample. We further stratify the sampling units within industry group by a measure of size (substratify) related to their annual sales. 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. We base these cutoffs on a statistical analysis of data from the 2002 Economic Census.

Accordingly, these values are on a 2002 sales basis. We also used 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. The size substrata and sampling rates are later updated through analysis of the sampling frame.

Sample Selection: Sample selection is a two-stage process and begins by identifying the firms selected with certainty. If a firm's annual sales or end-of-year inventories were greater than the corresponding certainty cutoff, that firm was selected into the AWTS sample.

In the second phase, all firms not selected with certainty were subjected to sampling on an EIN basis. If a firm had more than one EIN, we treated each of its EINs as a separate sampling unit. To be eligible for the initial sampling, an EIN had to have nonzero payroll in 2003. The EINs were stratified according to their major industry and their estimated sales (on a 2002 basis). Within each non-certainty stratum, a simple random sample of EINs was selected without replacement. This process was done separately for each wholesale type of operation.

There were approximately 8,490 firms selected in the latest AWTS sample: 6,844 wholesale distributor firms; 1,132 manufacturers' sales branch and office firms; and 514 agent, broker, and electronic market firms. Of these, about 2,356 are large firms that were selected with certainty. The remaining 6,134 firms were selected with noncertainty.

Sample Maintenance: Periodically, we update the sample to represent new EINs appearing on the Business Register. These new EINs, called births, are EINs recently assigned by the Internal Revenue Service (IRS) on the latest IRS mailing list for FICA taxpayers and assigned an industry classification (if possible) by the Social Security Administration.

EIN births are sampled 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 AWTS, and it must meet certain criteria regarding its number of paid employees or quarterly payroll. In the first phase, births are stratified by broad industry groups and a measure of size based on quarterly payroll. A relatively large sample is selected using equal probability systematic sampling. The selected births are canvassed to obtain a more reliable measure of size, consisting of sales in two recent months, company affiliation information, and a new or more detailed industry classification code. Births that have not returned their questionnaire after 30 days are contacted by telephone.

Using this more reliable information, the selected births from the first phase are subjected to probability proportional-to-size sampling with overall probabilities equivalent to those used in drawing the initial AWTS sample from the December 2004 Business Register. Because of the time it takes for a new employer firm to acquire an EIN from the IRS, and

because of the time needed to accomplish the two-phase birth-selection procedure, births are added to the samples approximately nine months after they begin operation.

If a firm was selected with certainty and had more than one establishment at the time of sampling, any new establishments that the firm acquires, even if under new or different EINs, are included in the sample with certainty. However, if a single-unit firm was selected with certainty, only future establishments associated with that firm's originally-selected EIN are included in the sample with certainty; any new EINs that might later be associated with that firm are subjected to sampling through the quarterly birth-selection procedure. Single-unit EINs selected into the sample with certainty are not dropped from canvass and tabulation if they are no longer on the IRS mailing list. Rather, the firm that used the EIN is contacted, and if a successor EIN is found, it is added to the survey. For both inactive and reactivated EINs, data are tabulated for only the portion of the reference year that these EINs reported payroll to the IRS.

Non-Response: Data are imputed using survey data and administrative data as input for unit non-response, item non-response, and for responses that fail computer or analyst edits.

Estimation Procedures: Estimates are computed as the sum of weighted data (reported and imputed) for all selected sampling units that have activity in the reference year. The weight for a given sampling is the reciprocal of its probability of selection into the sample. The estimates are adjusted using the results of the most recent economic census. Variances are estimated using the method of random groups and are used to determine if measured changes are statistically significant.

Benchmarking: Sales estimates from the current sample are adjusted to the 2002 Economic Census by linking these estimates to the published census-adjusted estimates from the prior sample. For a given detailed industry based on the 2002 NAICS, the linking is performed by multiplying the sample-based sales estimate by a ratio. The numerator and denominator of the ratio are as follows:

- The numerator is the 2004 published census-adjusted sales estimate for the industry from the prior sample.

- The denominator is the 2004 sales estimate for the industry from the current sample.

For merchant wholesalers, excluding manufacturers' sales branches and offices, the 2006 expenses estimate for a detailed industry from the current sample was adjusted by applying the ratio described above. A method similar to the one for adjusting sales was used to link the end-of-year inventories and purchases estimates for merchant wholesalers, excluding manufacturers' sales branches and offices. First, the ratio described above is applied to the sample-based estimates for the given detailed industry for 2004 and subsequent years, resulting in adjusted estimates for these years. Then, the published adjusted estimates for 1998 through 2004 from the prior sample are input to the benchmarking program. Using this program, the estimates for 1999 through 2004 are revised in a manner that:

- Uses the published adjusted estimate for 1998 from the prior sample as a constraint, resulting in no revision to the published 1998 estimate.
- Uses the adjusted estimate for 2004 from the current sample as a constraint.
- Minimizes the sum of squared differences between the year-to-year changes of the input and revised estimates for 1999 through 2004.

For manufacturers' sales branches and offices, the following describes the method used to link the end-of-year inventories and expenses estimates for 2004 and subsequent years from the current sample. First, the ratio described above is applied to the sample-based estimates for the given detailed industry for 2004 and subsequent years, resulting in adjusted estimates for these years. Then, the difference between the 2004 adjusted estimate from the current sample and the 2004 published adjusted estimate from the prior sample is taken into account by applying a second ratio to the published adjusted estimates for 2002 and 2003 from the prior sample. The numerator and denominator of the second ratio are as follows:

- The numerator is the 2004 adjusted estimate for the industry from the current sample.
- The denominator is the 2004 published adjusted estimate for the industry from the prior sample.

Estimates at 2- and 3-digit NAICS levels are computed by summing the adjusted estimates for the appropriate detailed industries comprising the aggregate.

For agents, brokers, and electronic markets, estimates for 2004 and subsequent years are produced directly from the current sample and are not adjusted.

2. Procedures for Collecting Information

While the initial sample consisted of 8,490 firms, our mail count can vary. It can increase if we require, or firms request, that reporting responsibility be broken out by different locations or type of activity. It can decrease because we do not canvass companies that are out-of-business, out-of-scope, or merged with other firms. Typically the mail count

decreases each year of the sample. The following provides the mail count and unit response for the current sample:

<u>Survey Year</u>	<u>Mail Count</u>	<u>Unit Response</u>
2007	7,752	NA
2006	8,136	80.8%
2005	9,137	82.2%

Data for the AWTS are obtained at the 5-digit NAICS level with all intermediate and summary tables obtained by summation. Tabulations for merchant wholesalers, excluding manufacturers' sales branches and offices, will include sales, e-commerce, end-of-year inventories, inventory valuation method, purchases, and operating expenses. Tabulations for manufacturers' sales branches and offices will include sales, e-commerce, end-of-year inventories, inventory valuation method, and operating expenses. Tabulations for agents, brokers and electronic markets will include sales, commissions, and operating expenses. All estimates are published on the internet and provide information at broad industry levels, 4-digit NAICS industry groups, and one 5-digit industry. All other 5-digit industry levels are not published due to high sampling variance.

Statistical analysis of the annual data will be based on comparisons to prior year annual and monthly results; results of the most recent Economic Census; and published trade, business, and media reports.

3. Methods to Maximize Response

One month after the initial mail out, the National Processing Center in Jeffersonville, Indiana, mails a follow-up form to respondents who have failed to return their completed form. Also, the mailing name and address list is updated for any revision prior to each follow-up mailing. After a period of four weeks, a third request will be mailed to the remaining non-respondents. Because of the smaller amount of outstanding cases and because of timing, this request is sent via certified mail.

Following this, a telephone follow-up for all firms that have not yet responded by a certain date is conducted and processing continues in Jeffersonville, Indiana. The analytical staff in Washington follows up with firms who return an incomplete report form or report questionable data.

A laser printer facsimile connected to an "800" telephone line permits facsimile reporting to our collection facility on a 24-hour basis. Census Taker allows respondents to report at anytime as well. The Census Bureau also has an "800" toll free telephone number in both Washington, D.C. and Jeffersonville, Indiana, to permit respondents to call in data or ask questions without extra expense.

4. Testing of Procedures

Based on prior knowledge of other annual surveys, we have working plans to test data collection, editing, and tabulation operations.

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

Direct questions regarding the sample design and statistical methodology used for this survey to Ruth Detlefsen, Assistant Division Chief for Research and Methodology, Service Sector Statistics Division, U.S. Census Bureau, (301) 763-5171.

Direct questions regarding the planning and implementation for this survey to John Miller, Chief of the Current Wholesale Branch, Service Sector Statistics Division, U.S. Census Bureau, (301) 763-2758.

List of Attachments

- 1 - BEA Support Letter
- 2 - Copy of Report Forms
- 3 - Sample Cover Letter