Monthly Wholesale Trade Survey – Estimation Procedures

Estimates of monthly sales and end-of-month inventories are derived from data collected in the MWTS. Each month, firms in the MWTS sample are asked to report their sales and inventory data for the month just ending. Monthly totals are computed as the sum of weighted data (reported and imputed) for all selected sampling units that meet the sample canvass and tabulation criteria given below. The weight for a given sampling unit is the reciprocal of its probability of selection into the MWTS sample. The monthly totals are then benchmarked to the latest totals from the AWTS. Period-to-period (e.g., month-to-month) change estimates are computed using the benchmarked monthly totals. See Explanation of Revisions for a description of the benchmarking procedures.

To be eligible for the sample canvass and tabulation, an EIN selected in the noncertainty sampling operations must meet both of the following requirements:

- It must be on the latest available IRS mailing list for FICA taxpayers from the previous quarter.
- It must have been selected from the Business Register in either the initial sampling or during the quarterly birth-selection procedure.

Monthly total estimates for broad industry groups (e.g., 2-, 3-, and 4-digit NAICS levels) are computed by summing the benchmarked monthly totals for the appropriate detailed industries comprising the broader industry group.

Variances are estimated using the method of random groups.

Reliability of the Estimates

Estimates in published tables are based on data from the Monthly Wholesale Trade Survey, Annual Wholesale Trade Survey, and administrative records. To maintain confidentiality, no estimates are published that would disclose the operations of an individual firm. The total error of a published estimate may be considered to be comprised of sampling error and nonsampling error. Individuals who use Monthly Wholesale Trade Survey estimates to create new estimates should cite the Census Bureau as the source of only the original estimates.

The published estimates may differ from the actual, but unknown, population values. For a particular estimate, statisticians define this difference as the total error of the estimate. When describing the accuracy of survey results, it is convenient to discuss total error as the sum of sampling error and nonsampling error. Sampling error is the error arising from the use of a sample, rather than a census, to estimate population values. Nonsampling error encompasses all other factors that contribute to the total error of a sample survey estimate. The sampling error of an estimate can usually be estimated from the sample; whereas, the nonsampling error of an estimate is difficult to measure and can rarely be estimated. Consequently, the actual error in an estimate exceeds the error that can be estimated. Further descriptions of sampling error and nonsampling error are provided upon request. Data users should take into account the estimates of sampling error and the potential effects of nonsampling error when using the published estimates.

Sampling Error

Because the estimates are based on a sample, exact agreement with results that would be obtained from a complete enumeration of firms on the sampling frame using the same enumeration procedures is not expected. However, because each firm on 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.

The particular sample used in this survey is one of a large number of samples of the same size that could have been selected using the same design. If all possible samples had been surveyed under the same conditions, an estimate of a population parameter of interest could have been obtained from each sample. For the parameter of interest, estimates derived from the different samples would, in general, differ from each other. Common measures of the variability among these estimates are the sampling variance, the standard error, and the coefficient of variation (CV). 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. Note that measures of sampling variability, such as the standard error and CV, are estimated from the sample and are also subject to sampling variability. (Technically, we should refer to the estimated standard error or the estimated CV of an estimator. However, for the sake of brevity we have omitted this detail.) It is 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 published estimates incorporate this information into their analyses, as sampling error could affect the conclusions drawn from these estimates.

The estimate from a particular sample and its associated standard error can be used to construct a confidence interval. A confidence interval is a range about a given estimator that has a specified probability of containing the average of the estimates for the parameter derived from all possible samples of the same size and design. Associated with each interval is a percentage of confidence, which is interpreted as follows. If, for each possible sample, an estimate of a population parameter and its approximate standard error were obtained, then:

- For approximately 90 percent of the possible samples, the interval from 1.65 standard errors below to 1.65 standard errors above the estimate would include the average of the estimates derived from all possible samples of the same size and design.<
- For approximately 95 percent of the possible samples, the interval from 1.96 standard errors below to 1.96 standard errors above the estimate would include the average of the estimates derived from all possible samples of the same size and design.

To illustrate the computation of a confidence interval for an estimate of total sales, assume that an estimate of total sales is $10,750 million and the CV for this estimate is 1.8 percent, or 0.018. First obtain the standard error of the estimate by multiplying the total sales estimate by its CV. For this example, multiply $10,750 million by 0.018. This yields a standard error of $193.5 million. The upper and lower bounds of the 90-percent confidence interval are computed as $10,750 million plus or minus 1.65 times $193.5 million. Consequently, the 90-percent confidence interval is $10,431 million to $11,069 million. If corresponding confidence intervals were constructed for all possible samples of the same size and design, approximately 9 out of 10 (90 percent) of these intervals would contain the average of the estimates derived from all possible samples.

Nonsampling Errors

Nonsampling error encompasses all other factors, other than sampling error, that contribute to the total error of a sample survey estimate and may also occur in censuses. It is often helpful to think of nonsampling error as arising from deficiencies or mistakes in the survey process. Nonsampling errors are difficult to measure and can be attributed to many sources: the inclusion of erroneous units in the survey (overcoverage), the exclusion of eligible units from the survey (undercoverage), nonresponse, misreporting, mistakes in recording and coding responses, misinterpretation of questions, and other errors of collection, response, coverage, or processing. Although nonsampling error is not measured directly, the Census Bureau employs quality control procedures throughout the process to minimize this type of error.