# Quarterly Financial Report (QFR)

## How the QFR Survey Data are Collected 06/08/2020

**Sampling frame.** The QFR target population consists of all corporations engaged primarily in manufacturing with total assets of $5 million and over, and all corporations engaged primarily in mining; wholesale trade; retail trade; information; or professional and technical services (except legal services) industries with total assets of $50 million and over. The QFR sampling frame is developed from a file received annually from the IRS. This file contains select information for corporations who elect to file Form 1120 with the IRS. Once this file is received, the QFR examines the information and assigns an initial NAICS industry code to each corporation. Then this file is stratified based on the NAICS code, size of total assets and size of receipts.

Corporations whose operations are within scope of the QFR and have total assets of $250 million and over may be included in the sample with certainty and canvassed every quarter. In addition, receipts cut-off values are determined on an industry basis. If a corporation has receipts at or above the receipts cut-off value for their industry, that corporation is included in the sample with certainty and is canvassed every quarter. When a new sample is selected, all new certainty cases are canvassed starting in the fourth quarter. Note that all certainty cases from the prior years' samples will remain active in the QFR survey if they still meet the in-scope criteria.

**Sample selection.** Simple random samples are selected from the eligible noncertainty units in the remaining industry-by-size strata. The sample in each cell is divided systematically into four panels that are introduced over the next year. Each noncertainty panel is in the survey for eight successive quarters. Each quarter, one noncertainty panel is rotated out and a new panel is rotated into the sample. This means that the noncertainty sample for adjacent quarters is seven-eighths identical, and is one half identical for quarters ending one year apart. This panel rotation scheme also means that panels from up to three different sample frames could be active in the QFR survey. Panels from the most recent sample are introduced into the QFR survey starting in the fourth quarter, with the remaining three panels being introduced, one at a time, into the sample in each succeeding quarter (i.e., quarters one, two, and three). Then this process starts over again with the new sample selected in the following year.

See *Table N – Composition of the Sample*, in the Summary Data Tables section of the current QFR publication (available at http://www.census.gov/econ/qfr/index.html) for the number of active corporations in the sample by NAICS sector and asset size for the current quarter.

As a result of the Paperwork Reduction Act of 1995, QFR sample units are subject to time-in / time-out constraints. If a sampled company has less than $50 million in total assets and has been in the survey for eight quarters, that company is not eligible for selection again for the next ten years. If a company has total assets between $50 million and $250 million and has been in the survey for eight quarters, it is not eligible for selection again for the next two years. Because of the time-in / time-out constraints, it is necessary to evaluate the frame to ensure there are enough eligible units for the four panels of the current sample to be selected, and that there will remain enough units on the frame that will be eligible for selection in subsequent years. If there are too few units, the optimal sample size for the current year is reduced to allow enough units for future years' selections. This adjustment to the optimal sample size results in increased variance for these strata.

**Sample maintenance.** The file received from the IRS to create the QFR frame does not contain all corporate tax returns. This is due to several factors, including but not limited to, corporations filing late tax returns, mergers and acquisitions, spinoffs, corporate restructuring, etc. In an attempt to keep the coverage of certainty corporations as complete as possible, the QFR staff reviews current corporate news releases and public records to identify any potential additions and changes to the QFR target population. Corporations thus identified and thought to meet the QFR in-scope criteria are contacted to verify their in-scope status. Once verified, they will be included in the survey with certainty for the remainder of the year.

**Data Quality Indicators**

The QFR produces two indicators of data quality: the Unit Response Rate (URR) and the Total Quantity Response Rate (TQRR). The URR is defined as the percentage of sampled units (unweighted) that responded to the current

quarter's survey. In determining the URR, a unit is classified as a respondent if it provided adequate information for key QFR data items. The TQRR is defined as the proportion of the weighted estimated total of a key QFR data item that is either reported by a survey unit or acquired from some other information source (such as a publically available quarterly or annual report) and that is determined to be equivalent in quality to reported data. The current quarter data quality indicators are available for each sector at https://www.census.gov/econ/qfr/mmws/current/qfr_mg.pdf and https://www.census.gov/econ/qfr/retail/current/qfr_rt.pdf. Historical data quality indicators spanning 20 quarters are also available at https://www.census.gov/econ/qfr/documents/hist_mmw_resp_rates.pdf and https://www.census.gov/econ/qfr/documents/hist_ret_resp_rates.pdf. Because the average URR is less than 80% for all sectors, QFR periodically evaluates the estimates for nonresponse bias.

**Nonresponse**. Beginning with the fourth quarter 2010, the QFR implemented an improved methodology to account for companies that are active in the sample but who do not respond to the questionnaire. Prior to fourth quarter of 2010, nonrespondent noncertainty sample companies were accounted for by adjusting the weight of the respondent companies. At the same time, certainty companies were imputed. The improved methodology imputes data for all nonresponding companies (units) using statistical procedures that utilize previously reported data (if available) and data from current respondents of similar asset size and industry classification. Imputation also balances the accounting form items so that detail items sum to total items. The improved imputation methodology, on average, produces more accurate data estimates and better reflects current financial standing

Each item on the form is treated independently with regard to imputation method. Prior studies have determined the priority order of imputation methods for each item. The primary imputation method utilized by the QFR is Ratio of Identicals (AUXRAT), which is carried out via the formula:

$$Y_{jt} = \frac{\sum_i W_{it} \times X_{it}}{\sum_i W_{it\wedge} \times X_{it\wedge}} \times X_{jt\wedge}$$

where:

$Y_{jt}$ = Value imputed for non-responding company j at time t (Current Quarter)

$W_{it}$ = Weight for responding company i at time t (Current Quarter)

$X_{it}$ = Value for responding company i at time t (Current Quarter)

$W_{it\wedge}$ = Weight for responding company i at time t^ (1st or 2nd Prior Quarter)

$X_{it\wedge}$ = Value for responding company i at time t^ (1st or 2nd Prior Quarter)

$X_{jt\wedge}$ = Prior value for current quarter non-responding company j at time t^ (1st or 2nd Prior Quarter)

Other imputation methods include Weighted Means, Mode, Value, Simple Regression, and Sum.

Most of the imputation methods require the use of an imputation base. Each company is assigned to an imputation cell based upon its asset size class and industry. The imputation base is restricted such that only respondent cases that pass the edit and are balanced are included. Outlier detection methods, including Hidiroglou-Berthelot and Asymmetric Fences, are utilized to remove influential cases from the base.

Note: Imputation for nonrespondent cases is performed for current quarter processing only. If an originally imputed case reports data for the current quarter after the initial publication, that case's reported data will be included in

revisions to that quarter's data in the next scheduled publication; however, the original imputed data for the remaining nonresponding cases for that quarter will not be reimputed.

**Estimation.** The QFR uses a variable weight estimator. This estimator takes into account the panel rotation scheme described above as well as the sample maintenance techniques employed to keep the industry classification and company structure for active sample cases as up-to-date as possible. Data for those cases that are active in the sample but do not respond to the questionnaire are imputed (see Nonresponse paragraph above).

The estimates are calculated as follows:

$$\hat{X}_{kit} = \left[\frac{\frac{(4-b_{kit})}{q_{kit}}\,\hat{N}_{kit}^{(-2)} + \frac{4}{q_{kit}}\,\hat{N}_{kit}^{(-1)} + \frac{b_{kit}}{q_{kit}}\,\hat{N}_{kit}^{(0)}}{n_{kit}^{(-2)} + n_{kit}^{(-1)} + n_{kit}^{(0)}}\right] * \left[\sum_{h\in ki}\sum_{j\in hki} I_{jhkit}x_{jhkit}\right]$$

$$= \left[\frac{\hat{N}_{kit}}{n_{kit}}\right] * \left[x_{kit}\right]$$

$$= \left[\hat{W}_{kit}\right] * x_{kit}$$

where:

$\hat{N}_{kit}^{(0)}$, $\hat{N}_{kit}^{(-1)}$, and $\hat{N}_{kit}^{(-2)}$, are the estimated population sizes in enumerated industry *k* and asset class *i* at time *t* for the sample from the current year frame (0), sample from prior year's frame (-1), and sample from prior prior year's frame (-2);

$n_{kit}^{(0)}$, $n_{kit}^{(-1)}$, and $n_{kit}^{(-2)}$ are the number of sampled cases in currently-active panels in enumerated industry *k* and asset class *i* at time *t* from the (up to) three eligible sample frame years;

$q_{kit}$ is the total number of active panels in enumerated industry *k* and asset class *i* at time *t* (usually 8);

$b_{kit}$ is the number of active panels in the sample from the current year frame in enumerated industry *k* and asset class *i* at time *t* (usually 1 in fourth quarter, 2 in first quarter, 3 in second quarter, and 4 in third quarter);

$I_{jhkit}$ is an indicator variable (either 1 or 0) indicating that company *j* was sampled in sampling industry *h* and enumerated in industry *k* and asset class *i* at time *t*;

$x_{jhkit}$ is the reported or imputed data value for company *j* which was sampled in sampling industry *h* and enumerated in industry *k* and asset class *i* at time *t*;

$\hat{N}_{kit}$ is the estimated population size in enumerated industry *k* and asset class *i* at time *t*;

$n_{kit}$ is the number of sampled cases in currently-active panels in enumerated industry *k* and asset class *i* at time *t*;

$x_{kit}$ is the sum of reported and imputed data for cases in enumerated industry *k* and asset class *i* at time *t*; and

$\hat{W}_{kit}$ is the variable weight for enumerated industry *k* and asset class *i* at time *t*.

**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 making sure the results are still useful.

The Census Bureau has reviewed the data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied. (Approval ID: CBDRB-FY20-270).

**Sampling error.** The sample used in this survey is one of many possible samples that could have been selected using the sampling methodology described earlier. Each of these possible samples would likely yield different results. These samples give rise to a distribution of estimates for the unknown population value. A statistical measure of the variability among these estimates is the standard error (SE), which can be estimated from any one sample. The standard error is defined as the square root of the variance. The relative standard error (RSE) of an estimate is the standard error of the estimate divided by the estimate, and is usually expressed as a percent. It provides a measure of the variation of the data relative to the estimate being made. Note that measures of sampling variability, such as the standard error and relative standard error, are estimated from the sample and are also subject to sampling variability. *Table O – Sampling Variability*, in the Summary Data Tables section of the current QFR publication, gives the estimates of the relative standard error in percent for the estimates of net sales, depreciation, inventories, and total assets, and the estimates of standard error for income before income taxes.

Beginning with the third quarter 2006 publication, the methodology for calculating the sample variance changed. Prior methodology used a design-based approach, whereas the new methodology uses a model-based approach. This new approach provides a better estimate of sample variance in strata with few sample cases, and also reflects some of the variability due to our methods used to account for nonresponding companies. The relative standard errors presented in *Table O – Sampling Variability*, mentioned above, can be used to derive the standard error of the estimate. The standard error can be used to derive interval estimates with prescribed levels of confidence that the interval includes the average results of all samples:

- Intervals defined by one standard error above and below the sample estimate will contain the true value approximately 68 percent of the time.
- Intervals defined by 1.6 standard errors above and below the sample estimate will contain the true value approximately 90 percent of the time.
- Intervals defined by two standard errors above and below the sample estimate will contain the true value approximately 95 percent of the time.

To illustrate the computations involved in the above confidence statements as related to dollar volume net sales estimates, assume that an estimate of net sales for all manufacturing published in *Table 1.0 – Income Statement For Corporations in the NAICS Manufacturing Sector*, in the Financial Data Tables section of the current QFR publication, is $525,000 million for a particular quarter and that the relative standard error for this estimate, as given in *Table O – Sampling Variability*, in the Summary Data Tables section, is 0.2 percent. Then the SE is calculated as:

$$SE\left(\hat{X}_{kit}\right) = \left[\frac{RSE\left(\hat{X}_{kit}\right)}{100}\right] * \hat{X}_{kit}$$

$$= \left[\frac{.2}{100}\right] * \$525{,}000\ million$$

$$= \$1{,}050\ million$$

Therefore, a 90-percent confidence interval is:

$525,000 million ± 1.6 ($1,050 million)
= $525,000 ± $1,680 million,
or $523,320 million to $526,680 million.

If corresponding confidence intervals were constructed for all possible samples of the same size and design, approximately 90 percent of the intervals would contain the figure obtained from a complete enumeration.

**Nonsampling error.** The relative standard error estimates sampling variation, but does not measure all errors in the data. Nonsampling errors can be attributed to many possible sources:

- Coverage error, such as failure to accurately represent all population units in the sample, or the inability to obtain information about all sample cases;
- Response errors, possibly due to definitional difficulties or misreporting;
- Mistakes in recording or coding the data obtained;
- Other errors of coverage, collection, nonresponse, response, processing, or imputing for missing or inconsistent data.

These nonsampling errors also occur in complete censuses. Although no direct measures of these errors have been obtained, precautionary steps were taken in all phases of the collection, processing, and tabulation of the data to minimize their influence.

**Data Quality Indicators.** The QFR produces two indicators of data quality: the Unit Response Rate (URR) and the Total Quantity Response Rate (TQRR). The URR is defined as the percentage of sampled units (unweighted) that responded to the current quarter's survey. In determining the URR, a unit is classified as a respondent if it provided adequate information for key QFR data items. The TQRR is defined as the weighted estimated total of a key QFR data item that is either reported by a survey unit or acquired from some other information source (such as a publically available quarterly or annual report) and that is determined to be equivalent in quality to reported data. The TQRR is calculated for the QFR data items net sales, depreciation, income before income taxes, and total assets. The imputation rate, or the percentage of the weighted estimated total of a key QFR data item that is imputed for nonresponding units, is calculated as 100% minus TQRR. The current quarter Unit Response Rates and Total Quantity Response Rates are published in the QFR press releases available at http://www.census.gov/econ/qfr.

**Adjustment for seasonal variation**. The QFR seasonally adjusts net sales and net income after taxes for all manufacturing, all durable manufacturing, and all nondurable manufacturing; and net sales for all retail trade, using the X-13 ARIMA-SEATS[1] program.

Seasonal adjustment is the process of estimating and removing the seasonal elements in a data series. This adjustment eliminates the seasonal trend present in the series and reveals non-seasonal features and trends. The precision of the adjustment is dependent upon economic conditions and various other elements that introduce significant changes to seasonal patterns. A more stable climate will result in a more precise adjustment, while volatility can reduce precision, especially in the short-term.

Seasonally adjusted data are published in the QFR press releases available at http://www.census.gov/econ/qfr. The releases provide seasonally adjusted estimates for the current quarter, all previously published quarters in the current calendar year, and for eleven prior calendar years. All prior seasonally adjusted data are updated each quarter based on the most recent data and seasonal factors.

The seasonal adjustment models are periodically reviewed and updated to account for recent changes in seasonal trends. For more information about seasonal adjustment as it related to QFR, contact the Financial Reports Branch at 301-763-6600 or csd.qfr@census.gov.

[1] X-13ARIMA-SEATS is an updated version of the Census Bureau's X-12-ARIMA software. More information about X-13ARIMA-SEATS is available at http://www.census.gov/srd/www/x13as/.