Conducting Nonresponse Bias Analyses for Two Business Surveys at the U.S. Census Bureau: Methods and (Some) Results

Katherine J. Thompson[1]
[1]U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233.

Abstract

Business surveys in the U.S. Census Bureau's Economic Directorate compute two "flavors" of response rates: unit response rates, defined as the ratio of the total unweighted number of "responding" units to the total number of sampled units eligible for tabulation; and total quantity response rates, which measure the weighted proportion of a key estimates reported by responding units and from equivalent quality sources. I present the computation methods for both sets of performance measures along with a discussion of how these measures were used to target areas of potential non-response bias in two ongoing programs: the Quarterly Services Survey and the Annual Capital Expenditures Survey. The response rate analyses led to subsequent analyses using auxiliary data available for all sampled units to confirm initial findings. In this paper, I focus primarily on the analysis methods, illustrating each method with selected examples from each case study.

Key Words: Nonresponse bias analysis, business surveys

1. Introduction

This paper presents the methods of assessing the presence and influence of unit nonresponse bias on survey estimates used by two economic programs conducted by the U.S. Census Bureau: the Quarterly Services Survey (QSS) and the Annual Capital Expenditures Survey (ACES). These two surveys represent a cross-section of business surveys conducted by the U.S Census Bureau. Both use stratified samples, but the QSS obtains repeated measurements from sampled units each quarter over a five-year period, whereas the ACES selects a new independent sample each year. The QSS frame variable used to compute unit measure-of-size is highly correlated with the program's key characteristic, whereas the corresponding ACES primary frame variable may or may not be. Lastly, the QSS uses ratio imputation to account for unit nonresponse, while the ACES uses weight adjustment. Section 2 provides background about the two case studies.

In Section 3, I present the standard methodology implemented in the U.S. Census Bureau's Economic Directorate to compute response rates, followed by a discussion of how these measures were used to target areas of potential nonresponse bias in each program. When discussing response rates, I inject a statistical process control analysis framework in the presentations, deviating from the original analyses which focused strictly on determination of consistent "problem areas" in subpopulations. The response rate analyses led to subsequent nonresponse

---

This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

bias analyses described in Section 4 that used auxiliary frame data available for all sampled units to investigate – and in some cases confirm – the initial findings. Each analysis is accompanied by illustrative examples from the referenced case study. For the complete analyses, see Rosenthal and Davie (2008) and Smith (2009).

My presented approaches could be viewed as structured "detective work." Groves and Brick (2005) and Tucker et al (2007) provided a variety of tools for conducting nonresponse bias studies, with the first reference targeted at demographic surveys, and the second at establishment or business surveys. Our analysis used (modified versions) of a subset of these tools. Each study had four goals: (1) Identify subpopulations with low unit or weighted item response rates as areas of potential nonresponse bias; (2) Examine the ignorable response propensity assumptions in the imputation/weighting cells by comparing statistics available on both respondents and non-respondents; (3) Determine areas where nonresponse bias issues could be addressed by alternative imputation or weighting adjustments, including assessment of the underlying prediction models when possible; and (4) Identify areas where nonresponse bias issues could only be addressed by improvements to data collection. Although this paper concentrates on analysis of response rates and frame data comparisons (as suggested by the 2006 Federal Register Notice), both programs' full studies included other analysis techniques such as comparison to benchmark estimates and analysis of late reporters.

2. Case Studies

2.1. Quarterly Services Survey (QSS)

The nonresponse analysis presented for the Quarterly Services Survey (QSS) examined data collected from the first quarter of 2004 through the fourth quarter of 2005. The QSS provides estimates of total and change in quarterly receipts (published about 75 days after the end of the reference quarter) and early estimates of calendar year receipts for selected service sectors. Standard errors are computed using the method of random groups, with noncertainty units systematically assigned to 16 random groups and certainty units included in all random group replicates.

The QSS covers the following North American Industry Classification System (NAICS) sectors: Information; Professional, Scientific, and Technical Services; Administrative and Support and Waste Management and Remediation Services; and Hospitals and Nursing and Residential Care Facilities. Sampling units for the QSS are groups of establishments under common ownership – generally companies or administratively convenient parts of companies, including Employer Identification Numbers (EINs). The QSS sample comprises approximately 6,000 units and is subsampled from the Services Annual Survey (SAS). Calendar year receipts estimates from the QSS are revised about nine months after their release when receipts estimates from the SAS are published.

A new QSS sample is selected every five years. During the five-year cycle, sample maintenance activities are performed each quarter. During this process, out-of-business units are identified and removed from mailing; and newly formed businesses are identified, subjected to a two-phase sampling process, and selected units are added to the sample. These procedures are designed to alleviate undercoverage. Sample units are interviewed each quarter. Thus, QSS estimates are repeated measures estimates.

The measure of size variable on the QSS frame is highly correlated with the survey's key item (receipts). For most units in the QSS sampling frame, the measure of size is the value of the receipts reported in the most recent Economic Census. In the rare case where a unit was not included in the most recent Economic Census, an adjusted administrative data value is substituted. Henceforth, I refer to the auxiliary frame data variable for QSS as census-equivalent receipts.

QSS uses ratio imputation to account for unit nonresponse. The preferred imputation method is referred to in-house as the "ratio-of-identicals" model. With this imputation model, imputed values in the current time period *t* are obtained as $y'_{ijt} = \hat{\beta}_j y_{ij,t-1}$, where $y_{ij,t-1}$ is value[1] of item *y* from unit *i* in imputation cell *j* at time *t-1* and $\hat{\beta}_j = \frac{\sum_{i \in j} w_i y_{ijt} I_i J_i}{\sum_{i \in j} w_i y_{ij,t-1} I_i J_i}$, (w_i is the sampling weight for unit *i*, I_i is an indicator variable of response from unit *i* in both quarters, and J_i is an indicator variable of eligibility of data item y_{ij} in both quarters). For QSS, if y_{ijt} or $y_{ij,t-1}$ is identified as an outlier in an independent review procedure, the unit's data are excluded from the imputation base for $\hat{\beta}_j$. Imputation cells are defined by six-digit NAICS code cross-classified by tax-status unless the imputation cell contains fewer than ten respondents. In this case, the imputation cell is collapsed to the three-digit NAICS code cross-classified by tax status for all six-digit NAICS contained within the three-digit NAICS code. Within each NAICS by tax-status cell, separate imputation cells are created for large companies (mainly consisting of large businesses selected with certainty) and EINs (primarily consists of small and medium-sized businesses selected with a weight greater than one). Further details about QSS are available at http://www.census.gov/indicator/qss/qsstechdoc.pdf.

2.2. Annual Capital Expenditures Survey (ACES)

The nonresponse analysis for the Annual Capital Expenditures Survey (ACES) used data collected from the 2002 through 2006 survey years. The ACES is an annual survey of companies that collects data about the nature and level of capital expenditures by non-farm businesses operating within the United States. Respondents report capital expenditures for the calendar year in all subsidiaries and divisions for all operations within the United States. ACES respondents report total capital expenditures, broken down by type (expenditures on Structures and expenditures on Equipment). The ACES key estimates are totals and year-to-year change. Standard errors are computed with the delete-a-group jackknife, with noncertainty units systematically assigned to 15 random groups and certainty units included in all random group replicates. The ACES covers the following NAICS sectors: Forestry, Fishing, and Agricultural Services; Mining; Utilities; Construction; Manufacturing; Wholesale Trade; Retail Trade; Transportation and Warehousing; Information; Finance and Insurance; Real Estate and Rental Leasing; Professional, Scientific, and Technical Services; Management of Companies and Enterprises; Administrative and Support and Waste Management and Remediation Services; Educational Services; Health Care and Social Assistance; Arts, Entertainment, and Recreation; Accommodation and Food Services; and Other Services (except Public Administration).

The ACES universe contains two sub-populations: employer companies and non-employer companies. Different forms are mailed to sample units depending on whether they are employer (ACE-1) companies or non-employer (ACE-2) companies. New ACE-1 and ACE-2 samples are

[1] The most frequently employed model uses data collected from the prior periods. However, the same imputation model can be applied to "future quarter" data when current period and prior period data are collected on the same form or from matched (end-of-year) SAS data in the first quarter of the business year.

selected each year, so that ACES estimates are based on independent samples. The ACE-1 sample comprises approximately seventy-five percent of the total ACES sample.

The ACE-1 frame is developed from administrative payroll data. This auxiliary variable is not necessarily highly correlated with capital expenditures. The ACE-1 survey strata are defined by five company size class categories – each based on payroll – within industry: one certainty stratum per industry, and four noncertainty strata. The majority of the capital expenditures estimate in a given industry is usually obtained from the certainty and large noncertainty strata; reported zero values for capital expenditures are quite frequent with units from other strata. Thus, the auxiliary data available for the ACE-1 sample is an inconsistent predictor of capital expenditures.

There are no corresponding auxiliary/frame variables for the ACE-2 component. The ACE-2 sampling frame is comprised of businesses without paid employees or payroll, sole proprietors, and companies for which no administrative data have been received. From the ACE-2 frame, four sub-strata are formed based on legal form of organization and available administrative data. A simple random sample is selected within each sub-stratum independently, across all industry categories within each stratum.

ACES uses weight adjustment to account for unit nonresponse, using "adjustment-to-sample" procedures (Kalton and Flores-Cervantes, 2003). To do this, sampling weights for unit i (computed as the inverse probability of selection) are multiplied by a weighting-cell specific adjustment factor that is based on data known for both respondents and nonrespondents. For ACES, the weighting cells are the design strata, denoted by an h subscript. In the event of complete nonresponse in a certainty or large company noncertainty stratum within an industry, the two adjacent cells are collapsed.

The ACE-1 component employs a **ratio adjustment** procedure to account for unit nonresponse, using administrative payroll values obtained from the sampling frame. The ACE-1 weighting adjustment factor in weighting cell h is computed as $f_h = \dfrac{\sum_{i \in h} w_{hi} x_{hi} J_{hi}}{\sum_{i \in h} w_{hi} x_{hi} J_{hi} I_{hi}} = \dfrac{\hat{X}_h}{\hat{X}_{hr}}$, where x_{hi} is administrative payroll obtained from the sampling frame, w_{hi} is the sampling weight, J_{hi} is a sample inclusion indicator variable ($\equiv$ 1 for all sampled units), and I_{hi} is a response status indicator variable.

The ACE-2 component employs a **count adjustment** procedure to account for unit nonresponse, with $x_{hi} \equiv 1$ for all sampled units, so that $f_h = \dfrac{\sum_{i \in h} w_{hi} J_{hi}}{\sum_{i \in h} w_{hi} J_{hi} I_{hi}} = \dfrac{\hat{N}_h}{\hat{N}_{hr}} = \dfrac{n_h}{r_h}$. More details concerning the ACES survey design, methodology, and data limitations are available online at www.census.gov/econ/aces/.

3. Analysis of Response Rates

Business surveys in the Economic Directorate of the Census Bureau compute two "flavors" of response rates: unit response rates, defined as the rate of the total unweighted number of "responding" units to the total number of sampled units eligible for tabulation (see Section 3.1);

and total quantity response rates (TQRR), which are the weighted proportion of key estimates reported by responding units and from equivalent quality sources (see Section 3.2).

Economic data generally have very different characteristics from their household counterparts. First, business populations are highly skewed, with a large proportion of the estimated totals originating from a small set of cases. Consequently, the majority of economic programs administered at the U.S. Census Bureau utilize stratified designs that include these "large" cases with certainty, that may sample "medium sized" cases with high sampling rates, and that sample the remaining cases with very low sampling rates (usually less than 0.01). As a result, sampled cases with large design weights often contribute very little to the overall tabulated totals. To avoid over-representation of such small cases in computation, **unit response rates** are computed without using sampling weights. Note that a missing response from a certainty unit will induce nonresponse bias in the estimates, although the degree and influence of this bias on total survey estimates needs to be evaluated on a case-by-case basis.

Another characteristic of economic data is the availability of "equivalent quality" auxiliary data. The Census Bureau conducts an Economic Census every five years and maintains an "up-to-date" business register of administrative data. Frame variables may be timely and highly correlated with survey characteristics of interest. Moreover, in contrast to household surveys, in some cases it is possible to obtain a valid value of characteristic from an alternative source: for example, a published company report might contain quarterly sales figures that could be effectively substituted for the missing response data.

Lastly, an economic program may need to distinguish between the "reporting unit" and the "tabulation unit." The reporting unit is the sampled unit, assigned to at most one industry on the sampling frame. The same company may operate in several different industries. To deal with this, the data received from the reporting unit may be split into "tabulation units." In other cases, a program may consolidate establishment or plant level data to the company level to create a tabulation unit. Thus, for economic surveys, unit response rates are based on the disposition of the reporting unit and item response rates (TQRR) are based on the disposition of the tabulation units.

For economic data, the skewed populations make the unit response rate an inconsistent measure of data quality. In economic surveys, interpretable measures of data quality include a measure of size (e.g., payroll, capital expenditures) to account for the unit's relative importance in the estimates (Tucker et al, 2007), as with the TQRR measures.

Response rates are performance measures, compared against benchmarked targets. As such, for regularly recurring surveys, they should be evaluated and monitored using statistical process control methods. In defining process control measures, Montgomery (1991) categorizes two types of processes: (1) controllable process, where "assignable causes" affect operation(s); and (2) uncontrollable or random process, where inherent or natural variation affects operation. The unit response rate is a measure of combined controllable and uncontrollable processes. The program managers have control over the contact process, so that as the number of respondents increase, the response rate increases (controllable process). The Census Bureau Standards define a respondent as an eligible reporting unit for which an attempt was made to collect data, the unit belongs to the target population, and *the unit provided sufficient data to be classified as a response.* To satisfy the latter requirement, each program determines which collected data items are **required** in advance of data collection: response status for each unit is determined after all data processing – including analyst review and editing – is completed. As the data quality restrictions on required data items for the program increase, the greater the likelihood of the response rate decreasing because it is more difficult for reporting units returning the questionnaire to qualify as a respondent. Thus, the additional restrictions on the *quality* of the data received add a random (uncontrollable) element. Consequently, it is possible to offer specific protocols designed to improve the amount of contact with nonrespondents that do **not** improve the unit response rates.

In-house procedures for analyst review and follow-up of survey data are designed to improve the quality of the **estimates**. Analysts strive to reduce imputation rates for all key items. This is usually best accomplished by unit non-response follow-up of large cases (expected to contribute substantially to the estimate), followed by intensive analyst research for "large impute" cases comprised of more phone calls (targeted questions) and searches for auxiliary data sources (e.g., financial reports) to replace imputed values with equivalent data. Thus, the TQRR measures controllable processes, so that increasing an item's TQRR generally leads to improved data quality for the estimate. For programs that publish information on one or two characteristics, the TQRRs for each item are clearly superior performance measures over unit response rates. However, if the survey publishes several key characteristics, it may be unwise to measure performance by setting target TQRR values for all characteristics, since meeting all target values may be difficult or even impossible depending on the number of collected items and the processing cycle time allotment.

The sections below discuss computation and analysis of unit response rate and TQRR respectively for the two case studies. In each section, the discussion of the measures includes time-series plots. Such plots are useful for a variety of purposes. First, they provide some visual evidence of the stability of the process. A variable process that has an unacceptably low average response rate would <u>require</u> a detailed analysis of the data collection methodology and the sample design to determine the cause. A heretofore stable process that now has an unexpected lower or higher than-expected response rate could be indicative of a one-time or rare event that does not require future changes or could also require direct intervention to change the level of the measure over time. A key premise of statistical process control is that some amount of variation in any process is expected and legitimate. The concern arises when the process is "out of control," indicated when the performance measure falls outside of predetermined upper or lower control limits, when the performance measure is consistently above or below the median or average value, or when the performance measure exhibits an unusual or nonrandom pattern in the data. For example, in light of the OMB Standards, an "in-control" response rate process would be above 80-percent on the average.

When these nonresponse bias analyses were conducted, standard formulae for the response rate measures were being developed for the Economic programs. Now that baseline values have been established, we can begin to develop viable process control limits for each program.

3.1. Unit Response Rates

Unit response rates are computed as [R/(E+U)] * 100, where

- **R** is the number of responding **reporting** units;
- **E** is the number of **reporting** units eligible for data collection in the collection period (including chronic refusals); and
- **U** is the number of **reporting** units for which eligibility for data collection could not be determined.

This is a conservative calculation, since an unknown percentage of the "U" cases will be out-of-scope for the program. The unit response rate denominator excludes the cases from which there was no attempt to collect data because of planned "imputation" from auxiliary data. Unit response rates are **performance measures** and are reported without standard errors. These performance measures are reported to OMB’s Performance Assessment Review Tool (PART) for selected programs.

Figures 1 and 2 present time-series plots of unit response rates for QSS and ACES, respectively. Each rate is marked with a diamond. The average rate is indicated by the center asymptote.

For QSS, the program-level unit response rates during the studied quarters consistently hover around 70-percent, with little deviation. This is a stable process that whose mean value falls short of the targeted 80-percent goal required by the OMB Guidelines. The consistent 70-percent average response rate pattern was repeated in three of the four sectors; in one sector, the unit response rates fluctuate around 80-percent. Since these QSS rates demonstrate a stable process, direct interventions (changes) in collection procedures – such as a revised collection instrument (questionnaire) or a revised respondent contact protocol **--** would be required to change the overall rate.

For ACES, after 2002, the unit response rates are generally around 75-percent. Here, the atypically high response rate from 2002 provides some evidence that further scrutiny is needed: the process appears to stabilize afterwards. Subject-matter experts surmised that this could be a “Census-year effect” of higher reporting with increased contact and follow-up due to the Economic Census. This pattern – high unit response rate in 2002, consistently lower response rates in subsequent years -- is repeated across-the-board in all the ACES sectors. Starting with the 2003 data collection, there was a process intervention: the introduction of the Information and Communication Technology (ICT) Survey, a program that uses the same sample employer units and is conducted by the **same analysts** as the ACES.

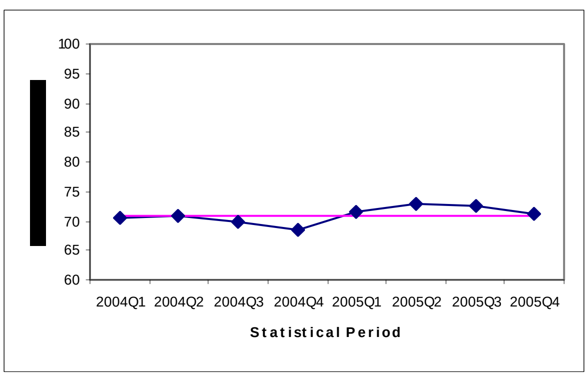

Figure 1: QSS Unit Response Rates

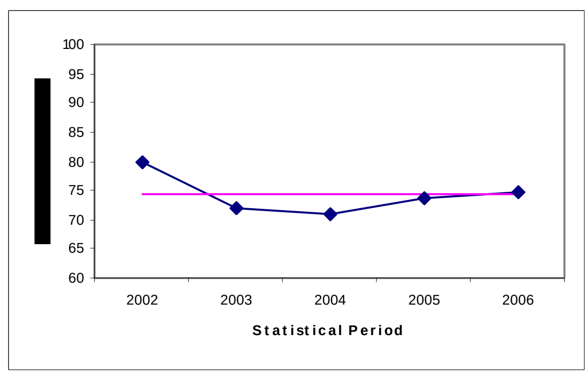

Figure 2: ACES Unit Response Rates

3.2. Total Quantity Response Rates (TQRR)

Total quantity response rates (TQRR) measure the weighted proportion of key estimates reported by responding units and from equivalent quality sources. Because these measures are derived from program estimates, they are computed from **tabulation** units. The TQRR for item *t* is computed as $[(R^t + A^t)/T]*100$, where

- $\mathbf{R^t}$ is the weighted estimate obtained by summing reported data for item *t*;
- $\mathbf{A^t}$ is the weighted estimate obtained by summing equivalent source data (auxiliary data from the same unit) for item *t*;

- **T** is the estimated weighted total of the variable over the entire population represented by the frame; it includes all data used to develop the publication estimate, including imputed data and (non-mailed) auxiliary data imputation cases.

Both numerator and denominator cases are weighted by "unbiased" weights, which include subsampling and outlier adjustment factors. In addition, denominator weights may include unit nonresponse adjustment factors. TQRRs are both performance measures and analytical statistics. When reported with survey estimates, standard errors of TQRRs are not included. However, complex survey design features must be incorporated into their analysis, especially when making statements of contrast or discussing averaged rates.

The TQRR analyses for both QSS and ACES are fairly straightforward, since both programs publish one key characteristic. We examined these rates at the program level, the publication sector level, and by certainty/noncertainty classification. Statistical analyses of rates over time account for the stratified sample designs in test statistics as well as the repeated measures collection for QSS.

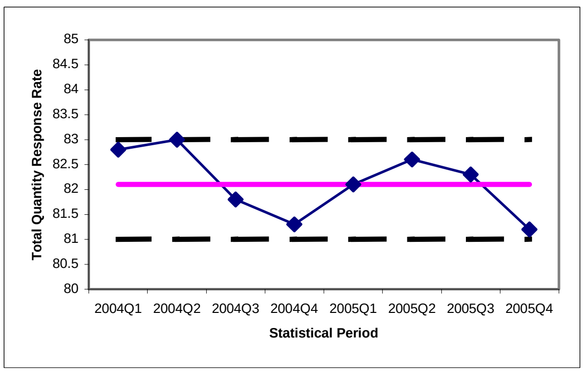

Figure 3: QSS TQRR (Receipts)

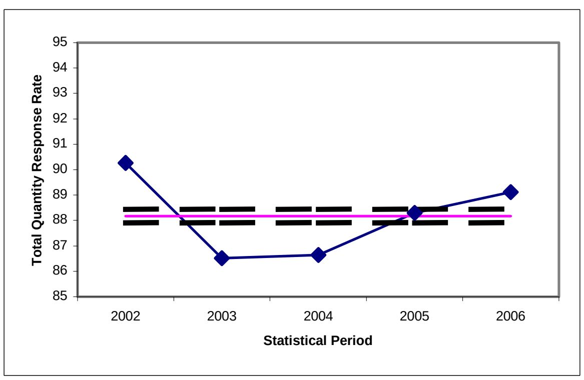

Figure 4: ACES TQRR (Capital Expenditures)

Figures 3 and 4 present Shewart process control charts of TQRR for QSS total receipts and for ACES total capital expenditures, respectively. For QSS, the upper and lower control limits were determined by conducting general linear hypothesis tests on the average TQRR and obtaining the largest and smallest point estimate (m) for which we failed to reject the null hypothesis that the 8-quarter averaged TQRR = m. For ACES, the upper and lower control limits are 95-percent confidence limits using averaged rates and standard errors from the five statistical periods. For presentation clarity, the scale of the two graphs' y-axes are different.

For the QSS, the average TQRR is 82.1 percent, and the potential rate range is (81 - 83). As with the corresponding process control chart for unit response rate, this chart provides evidence of a stable in control process over the eight studied quarters. This stable pattern is repeated in all of the sector subpopulations. Recall that Figure 3 presents the TQRR for the entire QSS program and demonstrates an overall attainment of the OMB limit. However, in two of the four sector subpopulations, the average TQRR is consistently lower than the preferred target value of 80-percent.

For ACES, the average TQRR is 88.2 percent. In 2003 and 2004, the ACES TQRRs fall well below the 88-percent lower control limits, but are consistently above the 80-percent mandate. In subsequent years, the ACES TQRRs are within the control limits or exceed them. This pattern was repeated across-the-board in **all** publication industries. Except for one sector, the TQRR was generally above the 80-percent target. We believe that the initial drop was caused by two intervention factors: the introduction of the ICT survey in 2003 (mentioned in section 3.1.) and the collection of more detailed capital expenditures data in 2003 for the ACE-1 component.

Besides examining TQRRs at the program and publication sector level, we analyzed these measures by certainty/noncertainty status. Recall that missing data from certainty cases are always indicative of nonresponse bias, since it is difficult, perhaps impossible, to develop a valid prediction model for certainty cases since they are by definition unique in the population. In contrast, low unit response rates or TQRR for noncertainty cases may not be indicative of nonresponse bias if we can find sufficient evidence that the response mechanism is ignorable and the prediction model used to justify the adjustment method is appropriate.

Both programs exhibit quite high TQRRs in their **certainty** components: above 85 percent for all QSS sectors and above 95 percent for ACE-1. This result is not unexpected. To reduce imputation rates for key items, analysts typically focus follow-up and review efforts on larger units. For the ACES, the high TQRRs for the certainty component of the sample are somewhat encouraging in terms of nonresponse bias concerns, although one should not completely dismiss the potential for nonresponse bias from this component. For the QSS, further investigation into nonresponse bias was warranted (see Section 4.1).

In both case studies, the average TQRRs in the **noncertainty** components were generally "close to" 70-percent. For QSS, at the program level, the average noncertainty TQRR range was (70-74), with two of the sector rates ranges including 70-percent, one range above 80-percent, and one rate consistently below 70-percent. For ACES, the TQRR ranges are much narrower: the ACE-1 noncertainty TQRR range was approximately 78 (actual range of 77.5 – 77.8) and the ACE-2 noncertainty TQRR range was approximately 61 (actual range of 60.6 – 61.6). Clearly, the very low range of potential rates in the one QSS sector and in the ACE-2 survey component merited further investigations. These low TQRRs may or may not be indicative of nonresponse bias, depending on the response mechanism and the validity of the prediction model. In Section 4, I outline the methods used to assess the response mechanism in noncertainty imputation cells.

4. Targeted Nonresponse Bias Analyses

In Section 3, I demonstrated how both case studies used response rates computed from different subpopulations to identify potential areas of nonresponse bias. In particular, the low TQRR for key items from the noncertainty component of each program was a "red light." The next phase of the nonresponse bias analysis undertaken by each program was to develop a research plan that investigated the targeted areas "uncovered" by the response rate analyses.

Low response rates are <u>not</u> necessarily indicative of nonresponse bias. It is well known that nonresponse bias is a function of the difference between the respondent and nonrespondent mean values and the non-response rate. If the respondent and nonrespondent mean values are not (statistically) different – that is, if the respondents comprise a random sample – then the nonresponse bias in the estimate can be mitigated by weighting or imputation. In this fortunate circumstance, the assumed response mechanism is "ignorable," meaning that the nonresponse bias in the estimates is "*ignored*" (i.e., negligible) after this corrective model is applied to the data. An ignorable response mechanism assumes that the probability of response is assumed to be related to a covariate, not the characteristic of interest (see Chapter 8 of Lohr (1999) or Chapter 1 of Little and Rubin (2002)). The adjustments used by the QSS and ACES assume a "missing-at-random" (MAR) response model, where the probabilities of response differ by weighting/imputation cell but are uniform within cell. Systematic differences between the respondent and nonrespondent mean values within a weighting/imputation cell are indicative of a non-ignorable response mechanism (i.e., the probability of response depends on the studied variable). Although some research has been done in the area of complex modeling to alleviate the nonresponse bias under an non-ignorable response mechanism, the easiest way to mitigate the nonresponse bias in these cases is to "get the data:"

Along with the appropriateness of the response model, the employed prediction model should be considered in nonresponse bias analyses. Under a MAR response mechanism, if the covariate used to develop the weighting/imputation cells is highly correlated with the probability of response and the units within the weighting/imputation cells have the same cell mean, then nonresponse bias and the nonresponse variance components are minimized by weighting the respondent data with the inverse response rate or imputing missing observations with the cell mean value (the "quasi-randomization" estimator uses the design-weighted response rates; Vartivarian and Little (2002) recommend using unweighted response rates). The ACE-2 weighting procedure assumes this response and prediction model. However, if an "explanatory" covariate x exists for characteristic y such that $y_{hi} = \beta_h(x_{hi}) + \varepsilon_{hi}$, $\varepsilon_{hi} \sim (0, x_{hi}\sigma_i^2)$ for each weighting/imputation cell h, then a better prediction can be obtained by using the ACE-1 ratio adjustment weighting procedure or the QSS ratio imputation procedures. When the model holds in all weighting/estimation cells, then this adjustment procedure should decrease the variance over the inverse response rate procedure described above. Moreover, if the covariate is also associated with response propensity, then both estimation bias and total variance are minimized.

A missing response from a **certainty** unit induces nonresponse bias in the estimates by definition. Thus, the response mechanism for a certainty unit should be considered non-ignorable, and any subsequent weighting or imputation adjustments will not correct the nonresponse bias. In the subsections below, I outline the approaches used for each case study to use auxiliary data from their sampling frame data to assess response models and prediction assumptions for the programs' noncertainty components.

4.1. Targeted Analysis of the QSS Using Frame Data

The QSS uses ratio imputation to adjust for unit nonresponse. The underlying prediction models assume that the previous reported value from unit i is a good predictor of the current value. Given the frequency of the QSS data collection, this is not an unreasonable assumption. The QSS imputation cells are defined by three or six-digit NAICS industry cross classified by tax-status and sample unit definition (company or EIN). The majority of large companies in the QSS sample are certainty units, although there are some cases of large companies being selected with probability less than one. The potential for commingling certainty and noncertainty cases in the imputation parameters is a limitation of the QSS imputation procedure in terms of nonresponse bias reduction.

Our targeted analysis for QSS focused on two issues:

- The degree of nonresponse bias induced in estimates of totals obtained from using only responding **certainty** cases in computations;
- The assumption of a uniform response mechanism in each of the other imputation cells (noncertainty cases).

We conducted these analyses using the census-equivalent receipts (See Section 2.2.).

To examine the first issue, we estimated the relative bias in each certainty-case imputation cell obtained by using respondent average census-equivalent receipts obtained from respondents instead of the population average as $b_h = (N_h - R_h)(\overline{Y}_{Rh} - \overline{Y}_{NR,h})$, where N_h is the number of certainty units in imputation cell *h*, R_h is the number of responding certainty units in imputation cell *h*, $\overline{Y}_{R,h}$ is the mean value of census-equivalent receipts for QSS-responding certainty units in imputation cell *h*, and $\overline{Y}_{NR,h}$ is the mean value of census-equivalent receipts obtained for QSS-nonresponding certainty units in imputation cell *h*. This is a very indirect assessment of the degree of nonresponse bias, since the QSS ratio imputation model consists of cases that responded in concurrent quarters and QSS imputation cells may contain both certainty and noncertainty units. In general, we found that the respondent based cell means were (algebraically) larger than the nonrespondent based cell means. This provided more evidence of a non-ignorable response mechanism for certainty cases.

To examine the second issue, we performed two-sample *t*-tests of equivalence of the average (mean) census-equivalent receipts obtained from respondents to the corresponding value obtained from nonrespondents for each imputation cell. These comparisons examined ignorable response mechanism assumptions, hoping to find evidence that the noncertainty respondents comprise a random sample (of all sampled units) in each imputation cell (i.e., MAR assumption). The test statistic within each weighting cell *h* was computed as $t_h^* = (\hat{y}_{Rh} - \hat{y}_{NR,h}) \big/ \sqrt{\hat{v}(\hat{y}_{R,h}) + \hat{v}(\hat{y}_{NR,h}) - 2C\hat{o}v(\hat{y}_{Rh}, \hat{y}_{NR,h})}$, where $\hat{y}_{Rh}$ and $\hat{y}_{NR,h}$ are Hajek estimators[2] of the respondent and nonrespondent imputation cell *h* means respectively, $\hat{v}(\hat{y}_{Rh})$ and $\hat{v}(\hat{y}_{NR,h})$ are the random group variance estimates of those statistics, and $C\hat{o}v(\hat{y}_{Rh}, \hat{y}_{NR,h})$ is the random group covariance estimate. Since QSS uses 16 random groups, under H_0, $t_h^* \sim t(15)$.

We evaluated the t-test results in two different ways. In each sector, we tested whether the number of cells with significantly different respondent and nonrespondent means was larger than expected due to random variability using binomial tests. Due to sample size limitations of the binomial test, we could not perform this analysis in two of the four sectors that contained five imputation cells apiece. In one of the two remaining sectors, we found evidence of a systematic difference in respondent and nonrespondent imputation cell means[3]. We also looked at the hypothesis test results in the individual imputation cells over time. This examination uncovered three imputation cells that had fairly consistent differences between mean respondent and nonrespondent values over the studied eight quarters. These particular subpopulations would be good candidates for the other analyses suggested in Groves and Brick (2005) such as subsampling nonrespondents to determine why the systematic difference is occurring in order to determine appropriate corrective measures.

---

[2] $\hat{y}_{Rh} = \dfrac{\sum_{i \in h} w_{hi} y_{hi} I_{hi} J_{hi}}{\sum_{i \in h} w_{hi} I_{hi} J_{hi}} = \dfrac{\hat{Y}_{Rh}}{\hat{N}_{Rh}}$ and $\hat{y}_{NRh} = \dfrac{\sum_{i \in h} w_{hi} y_{hi} (1 - I_{hi}) J_{hi}}{\sum_{i \in h} w_{hi} (1 - I_{hi}) J_{hi}} = \dfrac{\hat{Y}_{NR,h}}{\hat{N}_{NR,h}}$, where I_{hi} and J_{hi} are the indicator variables defined in Section 2.2.

[3] The aggregated comparisons test the null hypothesis that the respondent and nonrespondent means are equally likely to be significantly different or the same against the alternative that it is more likely that the means are different using a binomial (sign) test at the 10% significance level.

4.2. Targeted Analysis of the ACE-1 Using Frame Data

The ACES uses weight adjustment to account for unit nonresponse. Unfortunately, there are no auxiliary frame variables that can be used to assess nonresponse bias in the ACE-2 component. Instead, our analysis concentrated on the ACE-1 component, noting that both the unit response rates and TQRRs for the ACE-2 component were consistently low enough in all industries and survey years to warrant concern. The ACE-1 analysis presented two different challenges from the QSS. First, the ACE-1 frame is stratified by size of company payroll, which may or may not be correlated with a company's capital expenditures value. Second, the delete-a-group jackknife variance estimator used for ACE-1 estimates includes the fpc correction, so that sampling variances for the certainty component are not available. Our targeted analysis for ACE-1 focused on three issues:

- The assumption that payroll is a good predictor of the probability of unit response;
- The assumption that payroll is a good predictor of reported capital expenditures;
- The assumption that the respondents in a weighting cell comprise a random sample.

To examine the first issue, we fit no-intercept logistic regression models in each noncertainty size strata within industry using the SAS® SURVEYLOGISTIC procedure. For these models, the independent variable is payroll and the dependent variable is the response indicator (I_{hi}). The certainty strata were excluded from this analysis because of the fpc-adjustments to the sampling variances. In fact, an analysis of the response mechanism for the certainty strata is not necessary, since the analysts always perform extensive follow-up of all nonresponding certainty cases. Except for the smallest size-class-strata, these results provided evidence of a relationship between the amount of company payroll and the probability of responding in the majority of industries/strata.

To examine the second issue, we fit no-intercept linear regression models in each noncertainty size strata within industry, again excluding the certainty strata using the SAS® SURVEYREG procedure. For these models, the independent variable is payroll and the dependent variable is total capital expenditures; see Section 4 for the explicit model. Again, except for the smallest size class strata within industry, these results provided evidence that the amount of company payroll could be used to predict capital expenditures. This analysis served two purposes. First, it provided evidence of a valid prediction model for three of the four size-class-within-industry strata, reinforcing the validity of the ratio model used for weight adjustments. In fact, for this (large) subset of noncertainty strata, the ACE-1 adjustment is in the desirable situation of using an auxiliary variable that is both related to response propensity and to characteristic outcome.

A secondary goal of the analysis was to determine whether payroll could be used as a proxy variable for capital expenditures in two-sample t-tests comparing mean payroll value for respondents to nonrespondent values to assess the random (ignorable) response mechanism assumption (c.f., the QSS analysis described in Section 4). Using the results from the linear regression analysis, we conducted these two sample t-tests in all noncertainty weighting cells <u>except</u> for the smallest size-class-strata cells within industry. Very few means were significantly different, and the overall number of cells with significantly different means in each size-class-strata was well below the expected 10-percent. Moreover, the significant cells differed by year, so there was no consistent evidence of a subpopulation with a nonignorable response mechanism in these size strata.

From these analyses, we concluded that the ratio weight adjustment methodology used for the three largest size class strata in the ACE-1 design demonstrates the "ideal" properties where the auxiliary variable is related to both the response propensity and the outcome variable (Little and Vartivarian, 2005). We found the opposite in the smallest size class strata in the ACE-1 design. Consequently, we recommended a change to the count adjustment weighting procedure used by ACE-2 in these (small company) strata.

Both the ACE-2 subpopulations and the smallest size strata ACE-1 subpopulations are characterized by very high reported-zero rates. We suspect that unit nonrespondents' would report zero values as well, so that missing data in these populations may not greatly affect the level of the **estimates.** As with the identified QSS subpopulations, interviewing a subsample of nonrespondents in these populations to confirm this hypothesis could do quite a bit to determine the extent of the nonresponse bias in the ACE-2 component. Specifically, if the nonresponse in these populations is generally related to the zero-level of the characteristic, then one could conclude that nonresponse bias is not a major concern.

5. Conclusion

Most of the time, the nature of an evaluation study leads to the conclusion that more research is needed. Nonresponse bias analyses are slightly different. The outcome of such analysis should be ***action items***, i.e., changes in survey methodologies and procedures that lead to improved data quality that can be immediately addressed. The two case studies discussed in the paper determined some "pockets" of nonresponse bias that could be addressed by changing the survey methodology by developing a few alternative imputation cells in selected cases (QSS) or implementing alternative weighting adjustments (ACES). They also pinpointed subpopulations where the likelihood of response appears to be related to the collected characteristic -- cases where the only possible correction to bias was to obtain the data from the nonrespondents. Here, the logical next step is to conduct cognitive analysis in these subpopulations to develop effective recontact approaches or to figure out a way to better collect the data during the survey processing cycle.

In practice, the completion of a nonresponse bias analysis study should be a beginning. Monitoring response and assessing the potential for nonresponse bias is not a static analysis: it truly falls under the domain of statistical process control. Although a nonresponse bias analysis can provide valuable insight into the quality of a survey's estimates, it is limited. A total quality management approach would continue to monitor the response rate process measures on an ongoing basis to ensure that once acceptable rates are achieved, the process remains in control. Such an approach should be untaken jointly by survey methodologists and subject-matter expert program managers.

Acknowledgements

I thank Carol Caldwell, Broderick Oliver, Rita Petroni, and Samson Adeshiyan for their careful review of earlier versions of this manuscript. Thanks are also due to Miriam Rosenthal, William C. Davie, Jr., and Justin Z. Smith for their research and their generous agreement to collaborate on these projects.

References

Federal Register Notice (2006). OMB Standards and Guidelines for Statistical Surveys.

Groves, R.M. and Brick, J. M. (2005). Practical Tools for Nonresponse Bias Studies JPSM Course Notes.

Kalton, G. and Flores-Cervantes, I. (2003). Weighting Methods. *JOS*, **19**, pp. 81-97.

Oh, H.L. and Scheuren, F.J. (1983). Weighting Adjustment Of Unit Nonresponse. *Incomplete Data in Sample Surveys*, **20**, New York: Academic Press, pp. 143-184.

Lohr, S.L. (1999), Sampling: Design and Analysis, Pacific Grove: Duxbury Press.

Little, R.J.A. and Rubin, D.R. (2002). Statistical Analysis With Missing Data. New York: John Wiley.

Little, R.J.A. and Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means? *Survey Methodology*, **31**, pp.161-168.

Montgomery, D.C. (1991). Introduction to Statistical Quality Control. New York: John Wiley.

Rosenthal, M. and Davie, J. (2008). Nonresponse Bias Analysis for the Quarterly Services Survey. Internal unpublished U.S. Census Bureau memorandum, available upon request.

Smith, J.Z. (2009). Nonresponse Bias Study for the Annual Capital Expenditures Survey. Internal unpublished U.S. Census Bureau memorandum, available upon request.

Tucker, C., Dixon, J., and Cantor, D. (2007). Measuring the Effects of Unit Nonresponse in Establishment Surveys. Introductory Overview Lecture, *Third International Conference on Establishment Surveys*, American Statistical Association.

Vartivarian, S. and Little, R.J. (2002). On The Formation Of Weighting Adjustment Cells For Unit Non-Response. *Proceedings of the Section of Survey Research Methods,* American Statistical Association, 3553-3558.