Challenges in the Treatment of Nonresponse for Selected Business Surveys

Katherine Jenny Thompson and Katrina T. Washington[1]
U.S. Census Bureau
4600 Silver Hill Road, Washington DC 20233
Katherine.J.Thompson@census.gov / Katrina.T.Washington@census.gov

Abstract
Probability sample selection procedures gift methodologists with quite a bit of control ***before*** data collection; the "optimal" design for a given frame ensures that the selected sample is representative. This situation can change after data collection. Not all sample units respond and those that do will not always provide data on every questioned characteristic, which can lead to biased estimates of totals. The degree of bias is determined by several factors, including the representativeness of the respondent set, the magnitude of the aggregated missing data values, and the effects of improper adjustment procedures.

Keywords: weight adjustment, ratio imputation, prediction model, propensity model

1. Introduction

Probability sample selection procedures gift methodologists with quite a bit of control ***before*** data collection. After verifying that the sampling frame is a fair representation of the population, the methodologist can determine an "optimal" design and can ensure that the selected units are "representative" of the population. This control can evaporate when the survey is conducted and forms are returned. Not all sample units respond, and those that do will not always provide data on every questioned characteristic. Both of these types of nonresponse will lead to biased estimates of totals if the respondent-based sample estimates are not adjusted. The degree of bias is a function of several factors, including the representativeness of the respondent set, the magnitude of the aggregated missing data values, and the effects of "improper adjustment" procedures on the respondent data. For surveys that collect totals and associated additive detailed data items, the effect of nonresponse bias can be particularly evident with the details, which may not be available from all respondents. Consequently, analysis of the respondent data may not be sufficient to determine the presence of nonresponse bias.

In this paper, we illustrate the challenges of developing appropriate nonresponse adjustment treatments via an analysis of the imputation procedures used by two sections of the Service Annual Survey (SAS). The analysis is a continuation of research presented in Thompson and Washington (2012), which focused on validation of the SAS programs' treatment for unit nonresponse. The SAS programs impute complete records for unit nonresponse using a hierarchy that imputes items in a pre-specified sequence determined by the expected reliability of available imputation models; item nonresponse is treated similarly. Each item has its own imputation model hierarchy that maximizes the use of logical edits and direct substitution before attempting model imputation. This approach

[1] 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.

allows for maximum flexibility in modeling and preserves the expected cell totals, but does not preserve multivariate relationships between items and creates variance estimation challenges. As mentioned above, there are generally an abundance of reliable auxiliary data to formulate the imputation models for revenue and expenses, but there is often little auxiliary information available for the detail items. Our analysis focuses entirely on the ratio model imputation used by these programs, examining data from the 2005 through 2010 collection years. In this paper, we extend our analysis to include item nonresponse, using the same data sets.

Section 2 provides background on the SAS, including a description of the sample design, the data collected, and the imputation procedures used. Section 3 examines the unit level response mechanism, evaluating the current set of adjustment cells and discussing weighting alternatives. In Section 4, we evaluate the underlying regression prediction models used for ratio imputation by item. We conclude in Section 5 with a summary of the challenges encountered and suggestions for future research.

All of the results presented in this paper are summary statistics that describe the **collective** set of annual results. To perform our evaluation, we repeat the same analysis independently in each collection year. To summarize the unit response rates (URR) results presented in Section 3, we average the six individual imputation cell estimates (one per collection year) to obtain a single summary statistic. For the individual item measures presented in Section 4, we average the imputation cell rates by item across time and then compute an overall average by item across imputation cells. To summarize the logistic regression analysis and chi-squared test results presented in Section 3 and the regression analysis presented in Section 4, we perform a longitudinal analysis, concluding a significant effect iff the null hypothesis was rejected in at least four of the six studied collection periods. In general, the URRs do not vary greatly by statistical period, so little information is lost by presenting the summary measures. However, the item level rates can vary quite a bit across imputation cells, and this variation is presented as error bars in the applicable figures. For disclosure avoidance reasons, the complete set of individual (annual) results are not available upon demand.

2. The Service Annual Survey (SAS)

The Service Annual Survey (SAS) is a mandatory survey of approximately 70,000 employer businesses having one or more establishments located in the U.S. that provide services to individuals, businesses, and governments. The SAS surveys companies in North American Industry Classification Series (NAICS) sectors 22, 48-49, 51, 52, 53, 54, 56, 61, 62, 71, and 81. The SAS collects aggregate and detailed revenues and expenses, e-commerce, exports and inventories data from a sample of business firms with paid employees. For processing purposes, the SAS is divided into five sections, each covering one or more NAICS service sectors. Our analysis is restricted to the SAS sections covering the transportation and health industries (SAS-T and SAS-H, respectively), which are NAICS sectors 48-49 and 62. Information on the SAS design and methodology is available at http://www.census.gov/services/sas/about_the_surveys.html.

The SAS uses a stratified random sample design that includes certainty and noncertainty strata assigned for each industry. Companies are stratified by their major kind of business (determined by the industry containing the largest portion of total receipts for the company), then are further sub-stratified by estimated annual receipts or revenue. All companies with total receipts above applicable size cutoffs for each kind of business are included in the survey as part of the certainty stratum and are asked to report for all their service industry locations. For companies with receipts below the applicable size cutoff, the Employer Identification Numbers (EINs) of these companies are then stratified by major kind of business and sub-stratified by total annual receipts or revenue. Within each noncertainty size stratum, a simple random sample of (EINs) is selected without replacement. Thus, the sampling units are either companies or EINs. Each sampling unit represents one or more establishments/locations owned or controlled by the same firm. The initial sample is updated quarterly to reflect births and deaths, adding new employer businesses identified in the Business and Professional Classification Survey and dropping firms and EINs that are no longer active.

The SAS collects total revenue, total and detailed expenses, and e-commerce for all industries, both taxable and tax-exempt; sources of revenue and expenses by type, as well as export and inventory data for selected industries; operating expenses for tax-exempt firms; and other selected industry-specific items. The key items collected by SAS are total revenue and total expenses. For both revenue and expenses, there are many detail revenue and expense items collected that sum up to their respective totals. Collected detail revenue items vary across industries within NAICS sectors. Expense detail items are primarily the same for all sectors, with an occasional additional expense detail or two collected for select industries. Table One describes the item relationships in SAS-H and SAS-T. Notice that the two totals items (1800 and 1900) are requested from all sampled units, and that the collected detail items vary by sector and subsector.

Table One: Item Relationships in SAS-H and SAS-T

Total	Details	For Sector/subsectors:
1900 (Total Expenses)	1821[1], 1822, 1823, 1824, 1825, 1826, 1827, 1828, 1829, 1830, 1831, 1832, 1899	49, 62 (2008 onward)
	1821, 1822, 1823, 1824, 1825, 1826, 1827, 1828, 5097, 1829, 5098, 1830, 5099, 1831, 1832, 1899	48
1800 (Total Revenue)	1741, 1742, 1798	624
	4001A, 4002A, 4003A, 4004A, 4005A, 4006A, 4007A, 4009A, 4061A, 4062A, 4063A, 4064A	6215
	4001A, 4002A, 4003A, 4004A, 4005A, 4006A, 4008A, 4071A, 4072A, 1741, 1742, 1809	Tax-exempt 623
	4001A, 4002A, 4003A, 4004A, 4005A, 4006A, 4008A, 4071A00, 4072A, 1809	Taxable 623
	4001A, 4002A, 4003A, 4004A, 4005A, 4006A, 4007A, 4008A,4009A	621, excluding selected industries
	4001A, 4002A, 4003A, 4004A, 4005A, 4006A, 4007A, 4008A,1741 (selected industries), 1742 (selected industries), 1809	Tax-exempt 622 and selected industries
	5061, 1799	484
5090 (Truck Inventory)	5088, 5089	484
5093 (Truck-Tractor Inventory)	5091, 5092	484
5096 (Trailer Inventory)	5094, 5095	484

[1]Item 1821 is total payroll, which is a detail item that has available administrative auxiliary data.

Data collection and nonresponse adjustment for the total items are much less problematic than for the detail items. Companies are usually able to proportion out their “bottom line” revenue or expenses in a number of ways. However, methods of bookkeeping, limited staffing, company structure, and smaller company size may make it difficult for some respondents to calculate or even estimate values for a number of requested detailed revenue or expense items. For example, companies may do all their accounting by region, as opposed to by types of industries in which they do business. Similarly, a company’s line item in its bookkeeping may hold their expenses for all computer needs -- both hardware and software -- together. However, the SAS collects hardware and software expenses separately.

Imputation methodology is used to account for both unit and item nonresponse in the SAS. These models use auxiliary survey and administrative records data as input. Such data are sometimes available for the total items, but survey data for the detailed receipts and expenses are often very sparse and there is no available alternative administrative data. SAS-H and SAS-T utilize two ratio imputation models, both presented in Section 4.

For SAS, the imputation cells are six-digit industry (NAICS) code cross-classified by tax-exempt status. Unlike the sampling strata definitions, the imputation cells do not account for unit size, and imputation parameters use certainty and (weighted) noncertainty units within the same cell. The imputation base for the ratio imputation parameters is restricted to complete **respondent** data, subject to outlier detection and treatment.

In the two studied data sets, the NAICS code, tax-exempt status (all units in SAS-T are in taxable industries), certainty status, weights, and sampling stratum are available for all sampled units, and frame measure of size (MOS) is available for most units.

3. Assessing the Fitness of the Imputation Cells

As mentioned in Section 2 above, instead of using adjustment cell weighting to adjust for unit nonresponse, SAS imputes a complete record. Each item has its own imputation model hierarchy that maximizes the use of logical edits and direct substitution before attempting model imputation. The objective of the unit imputation is to reduce or eliminate the bias in all total estimates due to unit nonresponse. For this to happen – regardless of whether we employ imputation or adjustment is weighting– we need to answer the following questions:

1. Do the existing imputation cells satisfy an ignorable response mechanism assumption?
    - Are categorical variables used to form adjustment cells predictive of unit nonresponse?
    - Do response propensities differ between cells?
    - Are other variables missing from adjustment cells?
2. Within imputation cell, are the respondents a random sample?

The research objective of the first question can be restated as determining how best to partition the sample into the response homogeneity groups (RHGs) described in Särndal, Swensson, and Wretman (1992). The RHG model assumes the following conditions:

i. The probability of response (φ_{pi}) for unit i in cell p is the same for all sampled units i ($i = 1, 2, \ldots, n_p$) in cell p ($\varphi_{pi} = \varphi_p$); and
ii. The probability of response in cell p differs from that of cell p' for all $p \neq p'$

These two properties are sufficient for assuming a missing-at-random (MAR) response mechanism, i.e., that the probability of response depends on auxiliary variable(s) used to form adjustment cells and is not directly related to characteristic(s) of interest (Assumption M in Shao and Thompson, 2009). This is formally stated as $P(M_i | Y,X) = P(M_i | Y_O, X)$ where M_i is the missingness indicator, Y represents the characteristic of interest (subject to nonresponse), X represents the auxiliary variables used for the adjustment cells, and Y_O are the observed values of Y. Note that these conditions are not necessary for the more general covariate-dependent response mechanism, which allows the probability of response to depend on the auxiliary variable(s), not characteristic(s) of interest, i.e. $P(M_i |Y,X) = P(M_i |X)$. Under the covariate-dependent response mechanism, it is sufficient to prove that the probability of response differs by unit and is predicted by the level of auxiliary variable(s) within the imputation cells (Assumption P in Shao and Thompson, 2009); this is usually validated via logistic regression analysis.

It can be shown that under the MAR response model, an “inverse response rate” adjustment to the design weights produces an “unbiased” total from the respondent data. The inverse response rates can be computed using design weights (Kalton and Flores-Cervantes, 2003), which is mathematically equivalent to mean imputation using the Hajak estimator. Alternatively, the adjustment can be performed without design weights, as recommended in Little and Vartivarian (2005), which is mathematically equivalent to using an unweighted cell mean for imputation. Under the covariate-dependent response mechanism, the ratio imputation procedure described in Section 4 can yield a best linear unbiased estimator under specified error model assumptions.

Thompson and Washington (2012) present a logistic regression analysis of the SAS-T and SAS-H imputation cells. Our logistic regression analysis utilized two types of models: a simple model that used only the existing imputation cells as instrumental variables (Särndal and Lundström, 2005); and a nested model that included the continuous measure-of-size variable as a covariate nested within the imputation cells. We fit these models using the SAS SURVEYLOGISTIC procedure, which accounts for the complex survey design but excludes certainty cases. For both SAS-H and SAS-T, the logistic regression analysis provides evidence that industry/tax-exempt status category used to form adjustment cells is not strongly related to response propensity, but the unit size (MOS) nested within industry/tax-exempt status category is. Moreover, an analysis of the cell response propensities demonstrated that the larger cases (i.e., the cases with smaller sampling weights) were more likely to respond.

Certainly, the logistic regression analysis demonstrated that the existing sets of categorical variables used to form imputation cells – industry and tax status – were insufficient for forming RHGs. Indeed, we recommended developing adjustment cells that accounted for industry, tax-exempt status, and unit size to remediate the limitations of the current adjustment cells.

Ideally, the SAS sampling strata would serve as adjustment cells. In this case, not only would conditions (i) and (ii) hold, but the expected total item means should differ by adjustment cell and are expected to be the same within adjustment cell, thus improving the imputation models’ predictive properties. Unfortunately, the sample sizes in many of the sampling strata are prohibitively small (often less than five units) because of the highly stratified design and the limited number of large companies and large tax-entities in the sampling universe.

Rather than arbitrarily develop stratum-collapsing procedures, we created ad hoc size categories within the existing imputation cells for the noncertainty units; these same size-categories are used throughout the remainder of the paper. The size category cut-offs were the 33^{rd} and 66^{th} percentiles, respectively, of the imputation cell’s distribution of sampling weights. We chose these percentile values to ensure that each adjustment cell would contain at least five sampled units; we did not attempt to perform any optimality analysis. An alternative approach would be to develop cell partitions using estimated response propensities as described in Eltinge and Yansaneh (1997), assuming that the individual unit response propensities did not differ dramatically by collection year.

To assess whether the respondents comprised a representative subsample within the current imputation cells, Thompson and Washington (2012) conducted a contingency table analysis, constructing 3 *x* 2 contingency tables shown in Figure 1 using the cell size cut-offs described above for rows and response status as columns.

	Respondent	Nonrespondent	
$0 <= w_j < P_{33}$	n_{11}	n_{21}	$n_{1\bullet}$
$P_{33} <= w_j < P_{66}$	n_{12}	n_{22}	$n_{2\bullet}$
$P_{66} <= w_j$	n_{13}	n_{23}	$n_{3\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet\bullet}$

Figure 1: Contingency Table for Tests of Independence

The null hypothesis of interest is that response status is independent of unit size. Failing to reject the null hypothesis provides evidence of a random subsample within imputation cell. For this analysis, we used the SAS SURVEYFREQ procedure to perform the chi-squared tests for independence, again excluding certainty units.

For SAS-H, 11 of the 74 imputation cells were excluded from the 3*x*2 contingency cell analysis due to small numbers of respondents in some cells; 44 of the 63 remaining cells had “good fits,” providing evidence that the currently used adjustment cells could be “improved” by adding within-cell size categories. For SAS-T, only four of the 12 industry imputation cells had good fits, providing evidence that subdividing industry by size does not result in more representative samples. This result is not unexpected, as the sampling weights for noncertainty SAS-T units are not very variable, especially when compared to SAS-H.

Due to the finite population sampling correction, the logistic regression and contingency table analyses exclude certainty cases (the largest units in sample). We have considerable anecdotal evidence that – at a minimum – the imputation cells should include certainty status under an assumed MAR response mechanism. In business surveys, 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 quality data. In short, the certainty cases are more likely to respond if only because their response is strongly elicited.

To examine whether the response propensities differ for certainty and noncertainty units within the same imputation cell, we compare their unit response rates (URR), following the recommendations of Särndal and Lundström (2005). Figure 2 presents average URR for each SAS-H imputation cell, with blue squares presenting the certainty-unit URR, and the red squares presenting the noncertainty-unit URR within the same imputation cells. Figure 3 presents the corresponding measures for SAS-T.

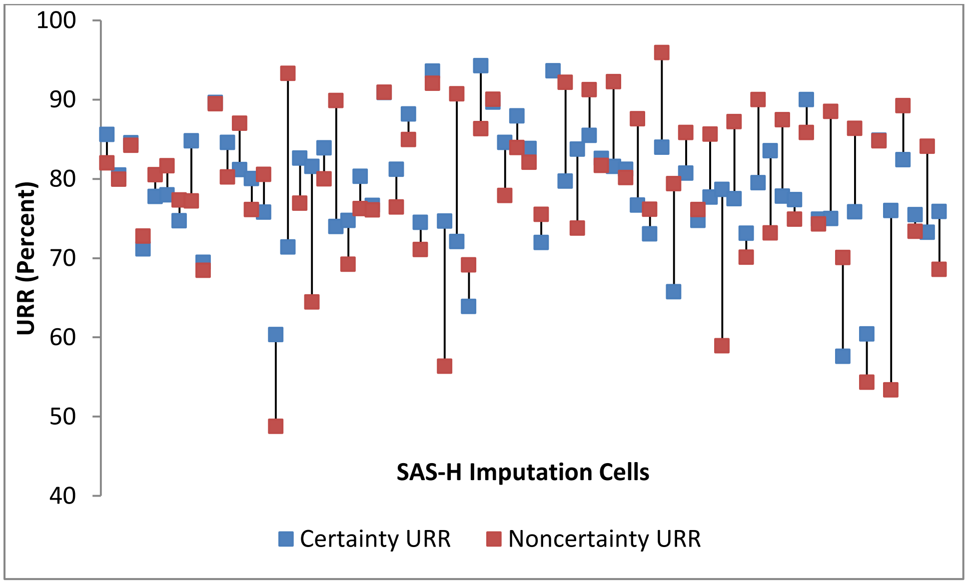

Figure 2: URR for Certainty and Noncertainty Cases within Imputation Cell (SAS-H**)**
Rates Averaged Over Study Period (2005 – 2010)

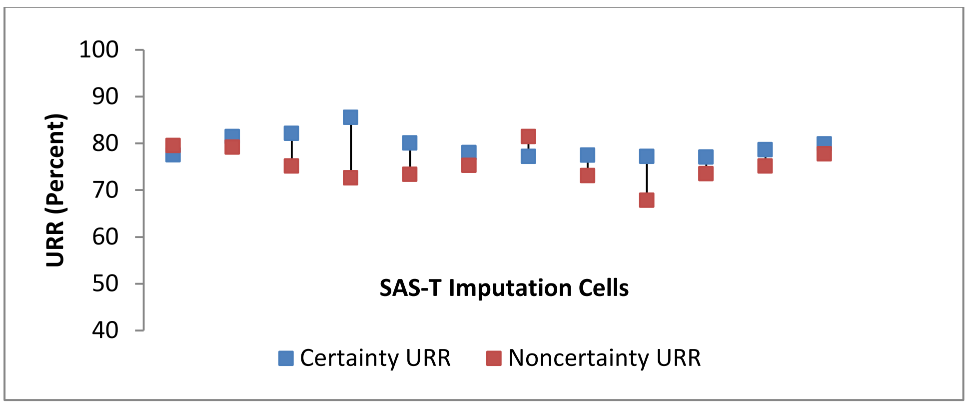

Figure 3: URR for Certainty and Noncertainty Cases within Imputation Cell (SAS-T)
Rates Averaged Over Study Period (2005 – 2010)

Figures 2 and 3 demonstrate that URR's are generally different within same cell for certainty and noncertainty cases. In 39 of the 70 SAS-H imputation cells[2], the average certainty-unit URRs are larger than the corresponding measures for the noncertainty units; in 10 of the 12 SAS-T imputation cells, the average certainty-unit URRs are larger than the corresponding measures for the noncertainty units. This provides more evidence that response propensity differs by unit size within imputation cell.

These analyses highlight issues with unit nonresponse in business data and challenges in remediating these issues. First, the unit response rate (URR) – even when fairly high – is not necessarily a good measure of representativeness of the sample (Peytcheva and Groves, 2009): in our examples, most of the URRs are at acceptable level, but the other analyses show that the larger units respond at a higher rate than the smaller units. By

[2] Four are excluded because they did not contain sufficient certainty or noncertainty units in all studied six years for computation.

partitioning the existing imputation cells by size categories, we can reduce some of the nonresponse bias. However, there are insufficient numbers of sampled units in the sampling strata to use them as adjustment cells, and the small number of sampled "large" units makes it challenging to subdivide the existing cells.

4. Evaluation of the Prediction (Imputation) Models

Ratio estimation improves estimate precision if and only if the auxiliary (independent) variable is highly positively correlated with the (imputed) dependent characteristic. SAS-H and SAS-T utilize two ratio imputation models, known in-house as "ratio-of-identicals" imputation.

Trend $y_{ti}^{p} = \beta^{p} y_{t-1,i}^{p} + \varepsilon_{ti}^{p},\ \varepsilon_{ti}^{p} \sim (0, y_{t-1,i}^{p}(\sigma^{p})^{2})$

Auxiliary $y_{ti}^{p} = \beta^{p} x_{ti}^{p} + \varepsilon_{ti}^{p},\ \varepsilon_{ti}^{p} \sim (0, x_{ti}^{p}(\sigma^{p})^{2})$

where p indexes the imputation cell (adjustment cell), t indexes the statistical period, i indexes the sampled unit, y is the characteristic being imputed, and x is the strictly positive auxiliary variable. The trend model is used exclusively for the totals items; auxiliary imputation is available for all items. Matthews (2011) provides specific imputation model information for SAS-H; Nelson (2011) provides the corresponding information for the SAS-T.

Under the auxiliary model, the B.L.U.E. for β^p is given by $\hat{\beta}^{p} = \sum_{i}^{n_p} w_i y_i^p I_i^p J_i^p / \sum_{i}^{n_p} w_i x_i^p I_i^p J_i^p$, where w_i is the design weight, I_i^p is a sampling indicator variable, and J_i^p is a response indicator variable. If this prediction model is not valid or if the strength of association between x and y is weak, then the bias induced by the ratio estimation increases the estimate's MSE, even though the use of ratio estimation reduces the variance component. This situation often occurs with the ancillary survey values, i.e. the detail items.

In Thompson and Washington (2012), we assessed the goodness-of-fit of each imputation model (H_0: $\boldsymbol{\beta} = \mathbf{0}$) with our survey data using the SAS SURVEYREG procedure within the currently used (industry code by tax-exempt status) imputation cells, again excluding certainty cases.[3] Figures 4 and 5 plot the average R^2 value for each collected item using the ratio-of-identicals model (items are labeled on the horizontal axis); a blue diamond indicating a consistently significant model, and a red square indicating the reverse.

Figures 4 and 5 demonstrated highly predictive models (R^2 greater than 75%) for the total items (1800 and 1900), and generally poor predictive properties for the detail items. So, model-imputation for totals is appropriate, but rarely used due to the availability of alternative data sources such as administrative or historic data, and model-imputation for details is not necessarily appropriate, but is generally employed.

[3] In contrast to the logistic regression analysis discussed above, the regression parameter tests require an additional distributional assumption of normally distributed error terms.

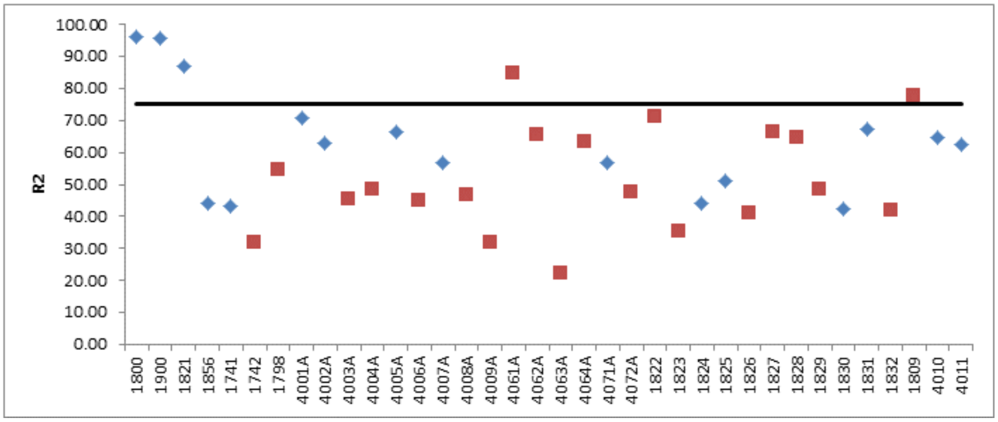

Figure 4: Average R^2 for SAS-T Ratio Imputation Models

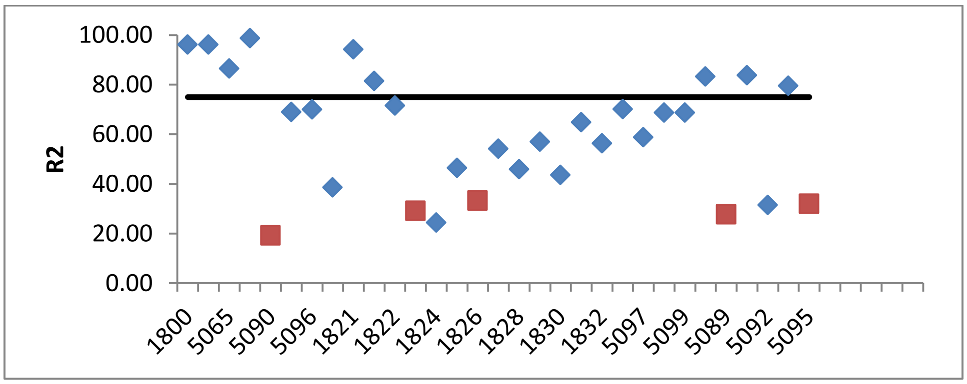

Figure 5: Average R^2 for SAS-T Imputation Models

Of course, the effectiveness of a ratio imputation model for correcting nonresponse bias is highly dependent on the data used to estimate the parameters. For SAS, the respondent units provide valid values for either revenue **or** expenses. "Reasonable" URR values – say greater than 70% -- therefore do not necessarily imply that the respondents provided data for other items. Figure 6 presents average **unweighted item response rates** computed within imputation cell for certainty and noncertainty cases from **respondent** units for the SAS-H items. The blue bars in the charts represent these double-average item response rates for certainty units, and the red bars present the corresponding noncertainty unit statistics. Error bars indicate the minimum and maximum item response rates, respectively[4]. Figure 7 presents the corresponding information for SAS-T.

The item response rates presented in the graphs were affected by an atypical data collection year and are consequently low (the year varied by program). For the totals, using medians instead of mean would have provided a more accurate picture of rates over time. However, for detail items, the differences obtained using medians instead of means were trivial. Consequently, we present averages for purely aesthetic reasons, namely putting all item response rates on essentially the same scale. The patterns displayed by these figures are consistent, regardless of averaging method, with each rate within the same imputation cell varying by certainty/noncertainty status, often by a large amount.

[4] Some detail items are omitted from the SAS-T and SAS-H graphs because they were not collected in all six studied years.

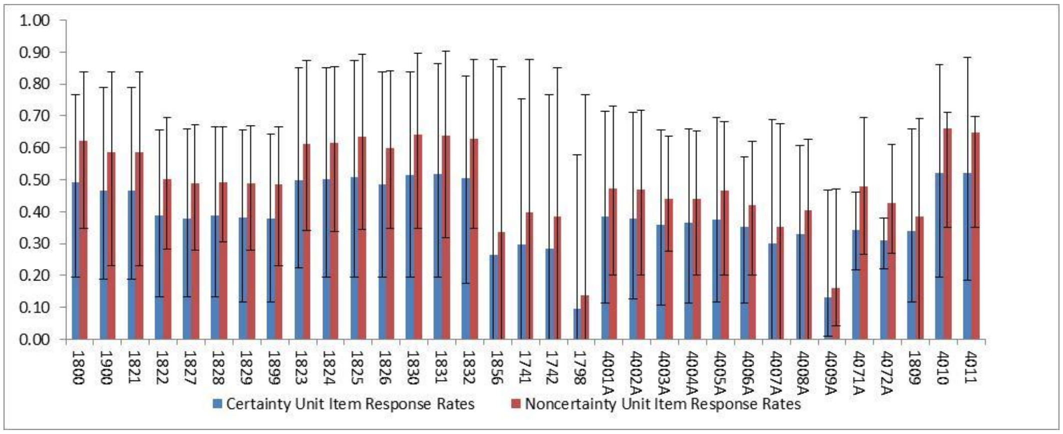

Figure 6: Item Response Rates by Certainty and Noncertainty Status for SAS-H

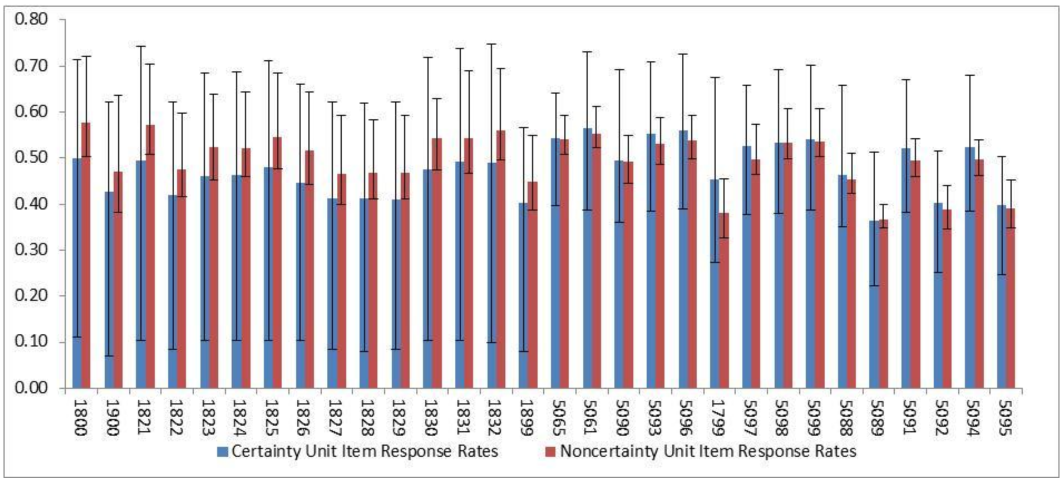

Figure 7: Item Response Rates by Certainty and Noncertainty Status for SAS-T

In Section 3, we examined the unit response mechanism for SAS-H and SAS-T, and concluded that response propensity was a function of industry/tax-exempt status and unit size. Consequently, we expect that the adjusted estimates are subject to nonresponse bias. That said, the degree of nonresponse bias is a function of the magnitude of the aggregated missing data values, and the effects of "improper adjustment" procedures on the respondent data. Table Two presents the average percentage provided by certainty units of the tabulated estimate computed from reported data by item over all imputation cells.

Table Two: Average Percentage Provided by Certainty Units of Tabulated Totals (Reported Data)

SAS-H						SAS-T					
Item	Percent	Item	Percent	Item	Percent	Item	Percent	Item	Percent	Item	Percent
1800	0.46	1828	0.41	4006A	0.39	1800	0.53	1829	0.47	5091	0.38
1900	0.46	1829	0.44	4007A	0.43	1900	0.53	1830	0.4	5092	0.38
1741	0.38	1830	0.37	4008A	0.47	1821	0.51	1831	0.52	5094	0.43
1742	0.45	1831	0.48	4009A	0.47	1822	0.58	1832	0.46	5095	0.43
1798	0.48	1832	0.40	4010	0.42	1823	0.47	1899	0.52	5065	0.43
1809	0.59	1856	0.48	4011	0.49	1824	0.37	5097	0.55	5061	0.46
1821	0.45	1899	0.47	4061A	0.65	1825	0.44	5098	0.35	5090	0.16
1822	0.48	4001A	0.46	4062A	0.74	1826	0.48	5099	0.36	5093	0.39
1823	0.43	4002A	0.43	4063A	0.18	1827	0.44	5088	0.14	5096	0.44
1824	0.37	4003A	0.40	4064A	0.46	1828	0.46	5089	0.23	1799	0.42
1825	0.42	4004A	0.40	4071A	0.31						
1826	0.39	4005A	0.51	4072A	0.25						
1827	0.42										

Recall that the imputation parameters use **all** respondent cases in the imputation cell. However, a non-trivial percentage of the tabulated weighted reported data used to create the parameters originates from the certainty units. From a nonresponse bias correction perspective, as long as the imputation parameters are approximately the same for each unit size category within an imputation cell, then the "dominance" of the certainty data does not affect the correction. The bar chart presented in Figure 8 illustrates this "ideal situation." For each SAS-H imputation cell, we stack the median[5] ratio of identical imputation parameters from the current adjustment cells (red) along with the corresponding median ratios computed within the same cells for only certainty units (green), large noncertainty units (blue), medium size noncertainty units (indigo), and small noncertainty units (purple).

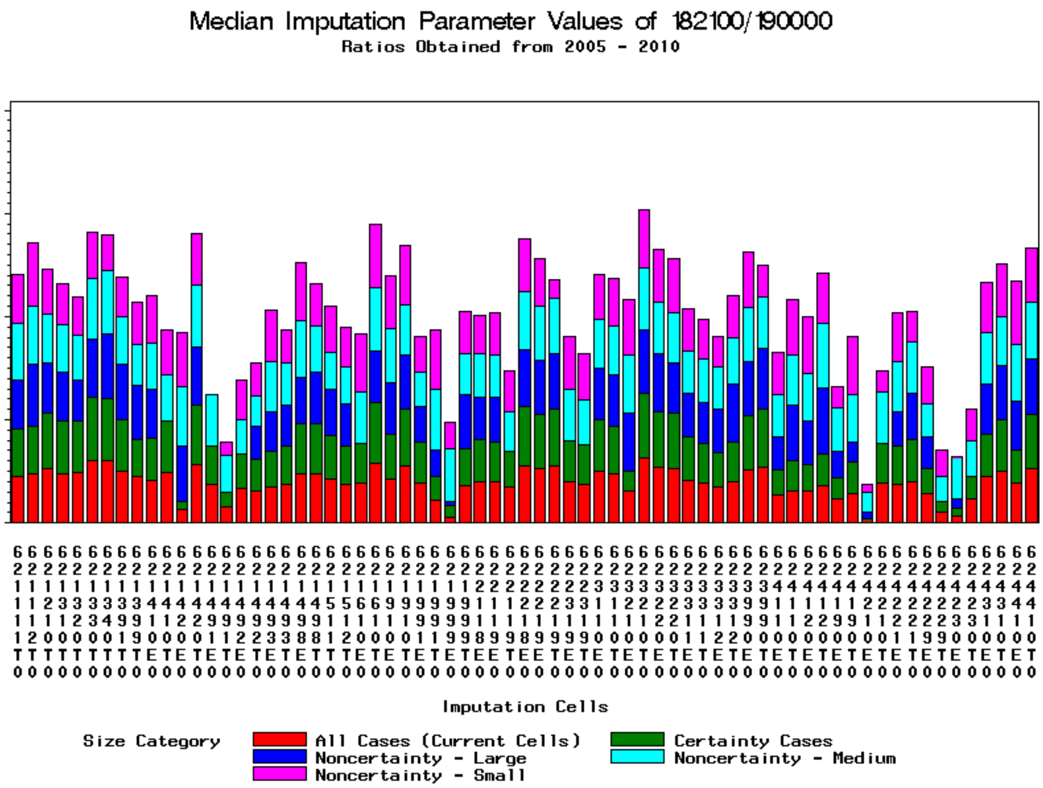

Figure 8: Ratio Imputation Parameters for SAS-H (Total Payroll/Total Expenses)

In this example, all of the imputation parameters are approximately the same. However, this is a ratio of two items that are generally well-reported in SAS-H. When we examine the plot for the same ratio parameters from SAS-T, the situation is quite different, as shown in Figure 9. Here, the imputation parameters computed from the certainty cases have almost exactly the same median ratio value as the parameters computed from the complete data in the imputation cell, and the imputation parameters for the other three size category cells (for noncertainty units) are each quite different. In this case, the ratio imputation model that SAS-T uses causes all imputed units to resemble the certainty units, even though it appears that is not supported by the data.

[5] For SAS-H, several detail items were collected in a subset of the considered six years. Average ratios could be highly affected by one atypical observation, whereas medians are not.

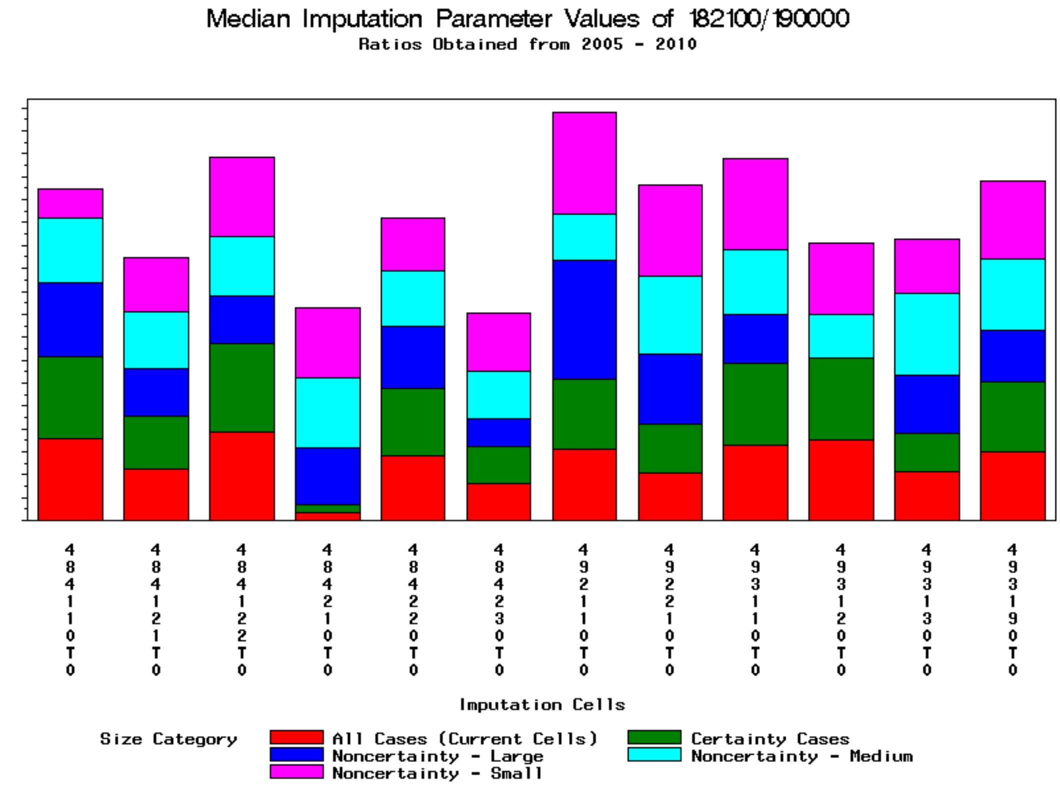

Figure 9: Ratio Imputation Parameters for SAS-H (Total Payroll/Total Expenses)

Similar plots are available for all ratio comparison upon demand.The vast majority of these analyses demonstrated similar patterns as that displayed in Figure 9, with the differences between size category within cell parameters being particularly prominent for the detail item ratios.

For SAS, imputation is performed independently in each adjustment cell. Consequently, the improper adjustment bias is aggregated, and it is impossible to determine what the cumulative effects of the bias are (if it exists). However, there is a data quality cost. Because all imputed items maintain the certainty-unit ratios, the imputed individual micro-data are not realistic, and all multivariate item relationships are lost. Furthermore, there is little evidence to validate the ratio models used for the detail items.

Although it would be unwise to recommend procedural changes for the entire SAS based on an analysis of two of five sections of the survey, there are many indications that existing adjustment procedures could be improved. At a minimum, the imputation cells should be refined to incorporate unit size. After completing item imputation, adjustment cell weighting can be used to adjust estimates for unit nonresponse, thus avoiding the issue of poor predictive imputation models for detail items. The determination of which weighting adjustment to use – weighted or unweighted or even a ratio– is a topic for future research, as is the method to optimally define imputation cells.

For missing or invalid data item imputation of detail items, weighted mean or mean imputation models could be substituted for the ratio-of-identicals imputation models for the detail item. This approach would have the advantage of **not** relying on poor predictive models but would have the disadvantage of increasing the estimated variance. A more substantive concern is the validity of the usage of imputation cell means for imputation. The imputation cells are not sampling strata, so it is a stretch to assume that the imputation cell respondents share the same mean. Moreover, the MAR response mechanism that validates the usage of the cell mean requires a random subsample within imputation cell. For the detail items especially, we have no evidence that the responding units provided representative values. Perhaps these items are missing because they do not exist for some sampled units – i.e., are legitimate zeros – and using the cell mean for imputation could yield overestimates. Alternatively, one could attempt donor hot deck imputation for the detail items as suggested in Bechtel et al (2011). More complex models that take missingness patterns in the covariates into account such the maximum likelihood and Bayesian inference methods described in Little (1992) could also be investigated.

5. Discussion

This analysis highlights several of the major challenges that business surveys encounter in addressing unit nonresponse. Respondents often do not comprise a random subsample, as larger units are more likely to provide data than smaller units. This phenomenon is an artifact of several factors, including the perceived benefits of the survey by the business community and the existing analyst nonresponse follow-up procedure, which focuses on obtaining the most accurate estimated totals.

Developing a set of adjustment cells that satisfy the most common ignorable response mechanism conditions and contain sufficient respondents is equally challenging, as there are considerably fewer "large" units in the population than small units. Finally, there are data challenges, as several of the detail items that the survey would like to collect may not be available from the majority of the sampled units. Again, the respondent sample size issues for the detail items are compounded by collecting different sets of detail items by industry or sector.

For treating unit response, there are benefits of using adjustment cell weighting instead of imputation, especially if the reweighting is performed **after** maximizing the use of administrative data substitution or logical substitutions. For SAS-H and SAS-T, the item-level response rates for many of the detail items are quite low and further tend to be reported by the larger units. Consequently, the reweighted detail items will still suffer from some model bias, and the extent of this bias cannot be determined without additional data collection/nonresponse follow-up.

We can easily improve existing adjustment techniques by refining the adjustment cells to account for missing covariates simply by subdividing the cells into certainty and noncertainty components. It might be possible to determine whether the respondent sample is "representative" based on auxiliary measure-of-size (frame) variables within the imputation cells or using other covariates to construct R-indicators or partial R-indicators (Schouten et al, 2009; Schouten et al, 2011) or to assess the "balance" of the sample or the response set (Särndal, C.E., 2011) – for the totals items. However, especially with low item response, we have no way of validating the appropriate adjustment procedure. Simply put, we need data.

Obtaining these data requires modifying the current data collection strategy described in Section 3. There are several excellent references on the use of adaptive or responsive designs to reduce the incidence of nonresponse bias by monitoring data collection and adapting procedures on a flow basis, utilizing different nonresponse follow-up strategies depending on response propensity (Groves and Heering, 2009; Laflamme et al, 2008). As a start, this approach could be extended to nonresponse follow-up for business surveys, building on our simplified propensity modeling with random subsampling for nonresponse follow-up within imputation cell, targeted phone follow-up of small cases, and modifications (if needed) of data collection procedures. This adaptive strategy could provide the information needed to learn about the missing data characteristics and would yield more statistically defensible bias-amelioration procedures.

Acknowledgements

The authors thank William Davie Jr., Donna Hambric, Xijian Liu, and Daniel Whitehead for their useful comments on earlier versions of this manuscript.

References

Bechtel,L, Gonzalez, Y., Nelson,and Gibson, R. (2011). Assessing Several Hot Deck Imputation Methods UsingSimulated Data from Several Economic Programs. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Eltinge, J.L. and Yansaneh, I.S. (1997). Diagnostics for Formation of Nonresponse Adjustment Cells, With An Application to Income Nonresponse in the U.S. Consumer Expenditure Survey. *Survey Methodology*,**23**, pp. 33-40.

Groves, R. and Heeringa, S. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. *Journal of the Royal Statistical Society, Series A*, **169**, pp. 439-457.

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

Laflamme, F., Maydan, M., and Miller, A. (2008). Using Paradata to Actively Manage Data Collection Survey Process. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Little, R.J.A. (1992). Regression With Missing X's: A Review. *Journal of the American Statistical Association*, **87**, No. 420, pp. 1227-1237.

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

Matthews, B. (2011). StEPS Imputation Specifications for the Health Portion of the 2010 Service Annual Survey (SAS-H). U.S. Census Bureau internal memorandum, EDMS #7526, available upon request.

Nelson, M. (2011). StEPS Imputation Specifications for the Transportation Portion of the 2010 Service Annual Survey (SAS-T). U.S. Census Bureau internal memorandum, EDMS #38537, available upon request.

Peytcheva, E. and Groves, R. (2009). Using Variation in Response Rates of Demographic Subgroups and Evidence of Nonresponse Bias in Survey Estimates. *JOS*, **25**, pp. 193-201.

Särndal, C.E. (2011). The 2010 Morris Hansen Lecture: Dealing with Survey Nonresponse in Data Collection, in Estimation. *JOS*, **27**, pp. 1-21.

Särndal, C.E., Swenssen, B., and Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer-Verlag.

Särndal, C.E. and Lundström, S. (2005). Estimation in Surveys with Nonresponse. New York: John Wiley & Sons, Inc.

Schouten, B., Cobben, F. and Bethlehem, J. (2009). Indicators For The Representativeness Of Survey Response. *Survey Methodology*, **35**, pp. 101-113.
Schouten, B., Shlomo, N., and Skinner, C. (2011). Indicators for Monitoring and Improving Representativeness of Response. *JOS*, **27**, pp. 231-253.
Shao, J. and Thompson, K.J. (2009). Variance Estimation In The Presence of Nonrespondents and Certainty Strata. *Survey Methodology*, **35**, pp. 215-225.
Thompson, K.J. and Washington, K. (2012). A Response Propensity Based Evaluation of the Treatment of Unit Nonresponse for Selected Business Surveys. *Proceedings of the Federal Committee on Statistical Methodology Research Conference*, www.fcsm.gov.