UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

May 21, 2012

To: William Davie, Jr.
Assistant Division Chief, Service Sector Statistics Division
Associate Director for Economic Programs

Sunitha A. Sebastian
Statistical Methods Branch
Service Sector Statistics Division

From: Joanna Fane Lineback
Chief, Response and Measurement Staff
Office of Statistical Methods and Research for Economic Programs

Eric B. Fink
Response and Measurement Staff
Office of Statistical Methods and Research for Economic Programs

Subject: Advance Monthly Retail Trade Survey Nonresponse Bias Analysis Report

Attached is the Advance Monthly Retail Trade Survey Nonresponse Bias Analysis Report. Portions of this report were redacted, because some data are sensitive under U.S. Codes Title 13 and Title 26. Individuals with Special Sworn Status may request a full copy of this report. If you have questions or need additional information, please contact Joanna Fane Lineback by phone at 301-763-7599 or by email at joanna.fane.lineback@census.gov.

Attachments

cc:
B. Davie (SSSD)
D. Weidenhamer
S. Sebastian
X. Liu (OSMREP)
E. Fink
J. Lineback

Advance Monthly Retail Trade Survey Nonresponse Bias Analysis Report

1. Introduction

This paper presents the results of a nonresponse bias analysis for the Advance Monthly Retail Trade Survey (MARTS) conducted by staff in the Office of Statistical Methods and Research for Economic Programs (OSMREP). This study was performed at the request of the Service Sector Statistics Division (SSSD). A similar study was recently conducted for MARTS parent survey, the Monthly Retail Trade Survey (MRTS).

The purpose of this analysis was to look for the potential for nonresponse bias in MARTS estimates of sales and month-to-month change in sales. There were two major study components:

- Response rate analysis
- MRTS nonresponse bias analysis

This paper is organized as follows. Section 2 provides relevant background information on MARTS. Section 3 discusses the link relative estimator for estimates of total sales and relative change in sales. Section 4 discusses MARTS response rates. Section 5 presents results from the MRTS nonresponse bias analysis. Section 6 discusses limitations of the analysis. Finally, Section 7 presents conclusions and recommendations.

2. Background

MARTS is conducted to produce early national estimates of total and month-to-month relative change in sales for retail and food service establishments located in the United States. A retail establishment is one that sells merchandise to the general public (final consumers). The estimates from MARTS are released approximately nine business days after the end of the reference month and revised one month later using estimates from MRTS. Estimates are summarized by industry classification based on the North American Industry Classification System (NAICS).

The MARTS sample includes approximately 5,000 units [companies and Employer Identification Numbers (EINs)] selected from the MRTS sample of about 11,600 units. The MARTS sample is stratified by industry and sub-stratified by monthly sales as measured in MRTS. There are 33 primary strata defined by industry and three to twelve size substrata for each primary stratum. Sample sizes are calculated to meet reliability constraints on estimated monthly sales totals for specified industries. Sample selection is conducted independently within each size stratum using a systematic probability-proportional-to-size procedure, using MRTS sampling weights. Sampling weights range from 1 to 1,000. Of the 5,000 units selected for sample, approximately 1,800 are "certainty" cases, which have a sampling weight of one. The sample is not updated (e.g., with births). Instead, the MARTS sample is redesigned and reselected every 2½ to 3 years.

Data are collected by mail, facsimile, or telephone. Collection units may be companies, parts of companies (as defined by EINs or divisions of diversified companies), or single unit establishments (also defined by EINs). Data are requested for activity taking place during the calendar month. Follow-up is conducted by telephone beginning the third business day after the reference month. Data collection is completed by the sixth business day following the end of the reference month. Response to the survey is voluntary. A link relative estimator is used for estimates of sales and change in sales. The link relative estimator is defined in Section 3 below. Estimates are adjusted for seasonal variation and holiday and trading day differences, and there are no nonresponse adjustments. Nonemployer businesses data are obtained from administrative sources and included in the estimates.

As mentioned above, MARTS key estimates are total monthly sales and month-to-month relative change in sales. Analysis in this report focuses largely on sales with observations about the statistical properties of the estimator of relative change in sales. Section 3 addresses the potential for nonresponse bias in relative change in sales estimates. MARTS response rate analysis and MRTS nonresponse bias analysis in Sections 4 and 5, respectively, examine the potential for nonresponse bias in total sales estimates.

3. Link Relative Estimator

MARTS sales estimates are computed using a type of ratio estimator known as the link relative estimator. Within kind-of-business (KB) category, the link relative estimator for total current month sales is defined as

$$\hat{Y}_{MARTS} = \left(\frac{\sum w_i y_i}{\sum w_i x_i} \right) \times \hat{X}_{MRTS},$$

where

w_i is the sampling weight for unit *i* that reported in both the current and prior months,

y_i is current month reported sales for unit *i* that reported in both the current and prior months,

x_i is prior month reported sales for unit *i* that reported in both the current and prior months, and

$\hat{X}_{MRTS}$ is the prior month MRTS estimate of total sales.

MARTS month-to-month relative change in sales is defined as

$$Rel\left(\Delta \hat{Y}_{MARTS}\right) = \frac{\left[\left(\frac{\sum w_i y_i}{\sum w_i x_i} \right) \times \hat{X}_{MRTS} \right] - \hat{X}_{MRTS}}{\hat{X}_{MRTS}},$$

which through derivation becomes

$$Rel\left(\Delta \hat{Y}_{MARTS}\right) = \frac{\sum w_i y_i}{\sum w_i x_i} - 1.$$

Madow and Madow (1978) note in their paper on link relative estimators that these estimators by definition may result in biased estimates of change, as samples change over time through activities such as mergers, splits, deaths, births, purchases, and sales. Unless the biases are cumulatively small, it is important to add supplementary sample or use an adjustment procedure.

Madow and Madow (1978) show that if $\hat{X}_{MRTS}$ (the benchmark) is biased, then $\hat{Y}_{MARTS}$ is biased and $Rel\left(\Delta \hat{Y}_{MARTS}\right)$ is possibly biased. Lineback and Fink (2011) show that there is evidence of nonresponse bias in MRTS sales estimates. It follows that there is evidence of nonresponse bias in MARTS total sales estimates and the potential for nonresponse bias in MARTS relative change in sales estimates. Madow and Madow (1978) also show that if $\hat{X}_{MRTS}$ is unbiased then $Rel\left(\Delta \hat{Y}_{MARTS}\right)$ is also unbiased. Therefore, correcting for bias in MRTS estimates of total sales is *key* for producing unbiased MARTS estimates of relative change in sales.

4. Response Rate Analysis

Response rates are often used as indicators of data "quality," even though it is well known that response rates alone often tell us little about quality. Response rate analysis, however, is important for this type of study, because nonresponse bias is a function of response rate, response propensity, and response prediction (Little and Andridge, 2011). If we somehow managed to obtain 100% response, then nonresponse bias is no longer a concern.

The unit response rate (URR) and the Dollar Volume Response Rate (DVRR), which is an item level response rate used by the MARTS program, were examined. We analyzed these rates at the program level and for key subgroups. Our analysis was somewhat limited, with data available for only three statistical periods, August 2010-October 2010, at the time this study was conducted. Only a few months of data were available, because MARTS was recently migrated to the Standard Economic Processing System (StEPS).

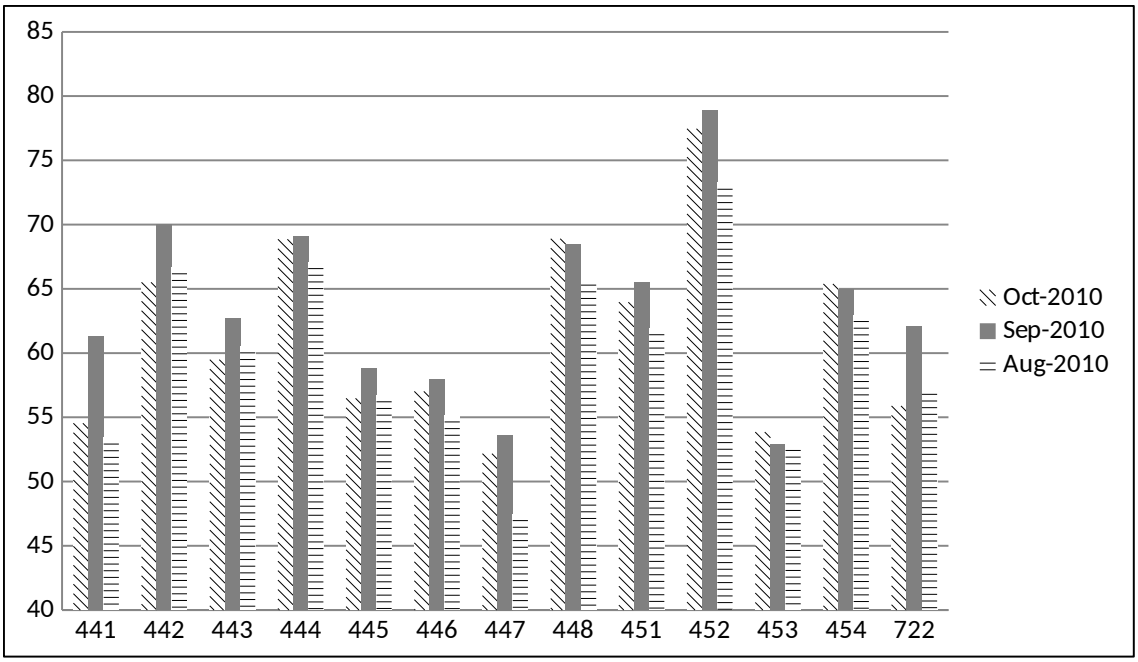
4.1 URR Analysis

The URR for MARTS is based on sample units that responded to MARTS in the current month. The average program level URR across the three months was 58%. As seen in Figure 1, the range of URRs by 3-digit NAICS, or subsector, was 51.4% to 70.7%.

For certainty units only, the average URR was 61.9%. The range across subsectors was 44.6% to 76.4%. For noncertainty units, the average URR was 58.5%, with a range of 44.3% to 74.4% across subsectors. A complete breakdown of URRs by subgroup is provided in Attachment 1.

In an analysis of response rates for similar programs, such as MRTS (Lineback and Fink, 2011) and the Quarterly Services Survey (Rosenthal and Davie, 2008), certainty and noncertainty response rates were often markedly different, with certainty units having much higher response rates than noncertainty units. However, it should be noted that many of the units that were selected with certainty in MRTS were noncertainty selections in MARTS, something that likely drove up noncertainty URRs. This does not, however, rule out the potential for nonresponse bias in estimates of sales.

Figure 1: MARTS URR by 3-Digit NAICS for 3 Statistical Periods



4.2 Dollar Volume Response Rate Analysis

A total quantity response rate, the standard economic area item response rate, is not calculated for MARTS estimates of sales because it is not an appropriate measure for the link relative estimator. Instead, an alternative DVRR is calculated.

Within KB category, the DVRR is given by

$$DVRR = \frac{\sum w_i y_i}{\hat{Y}_{MARTS}},$$

which through derivation becomes

$$DVRR = \frac{\sum w_i x_i}{\hat{X}_{MRTS}},$$

with the terms defined previously in Section 3.

All published monthly sales estimates that do not have an annual constraint at the time of benchmarking are derived by multiplying the Horvitz-Thompson monthly sales estimate by the ratio of the benchmarked to input monthly sales estimate for December of the year of the latest annual constraint. This ratio, which is called a carry-forward factor (CFF), remains the same and is used to derive published monthly sales estimates until the next benchmarking operation. Within KB category, this ratio is defined as

$$DVRR_{CFF} = \frac{\sum w_i x_i}{\hat{X}_{MRTS}} \times \frac{Benchmarked\,Series\,Value}{Original\,Series\,Value}.$$

Table 1 gives the DVRR and $DVRR_{CFF}$ rates over three months.

Table 1: Average DVRR and $DVRR_{CFF}$

Statistical Period	Average DVRR	Average $DVRR_{CFF}$
August 2010	70.77	69.38
September 2010	73.13	71.69
October 2010	71.84	70.35

Using MRTS estimates as "gold standards," discrepancies between MARTS reported sales and MRTS estimated sales are cause for concern. The potential for nonresponse bias exists, as MARTS prior month estimate averages 72% of MRTS prior month estimate over the three month period August 2010 to October 2010.

5. MRTS Nonresponse Bias Analysis

Staff in the Office of Statistical Methods and Research for Economic Programs (OSMREP) recently completed a nonresponse bias study for MRTS (Lineback and Fink, 2011). The MRTS study also serves as a component of the MARTS nonresponse bias analysis, because MARTS sample is a subsample of MRTS sample, and because MRTS sales estimates are used in the MARTS link relative estimation procedure. The MRTS analysis included the following components:

- Response rate analysis,
- A comparison of MRTS inventories estimates to ARTS inventories estimates, and
- A comparison of respondents to nonrespondents on substantive frame data.

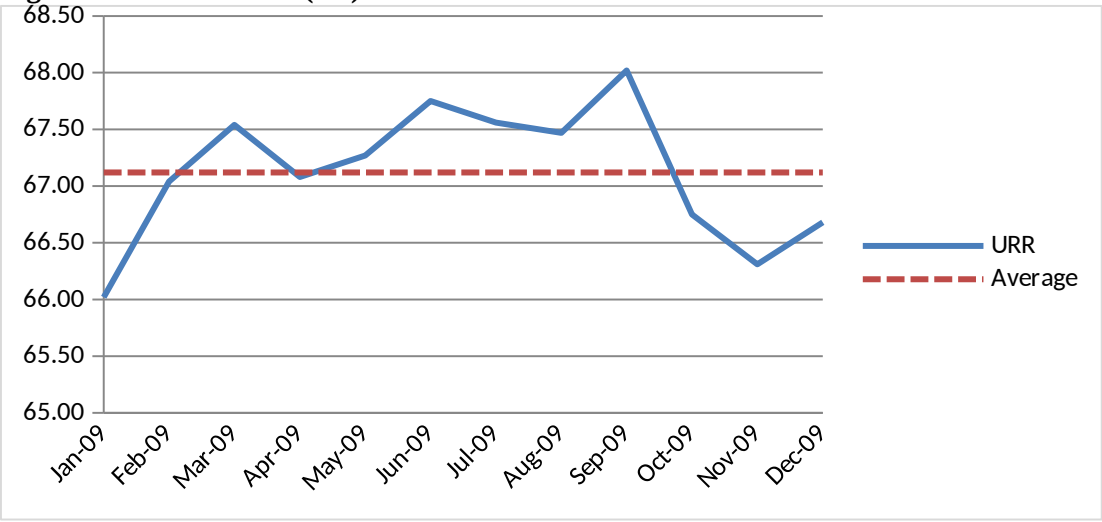
Key results from the MRTS report are discussed in this paper.

5.1 Response Rate Analysis

Response rate analysis was conducted using MRTS data from the 12-month period January 2009 to December 2009. We examined URRs and TQRRs for sales over time, at the program level, and for the key subgroups certainty status and 3-digit NAICS subsector.

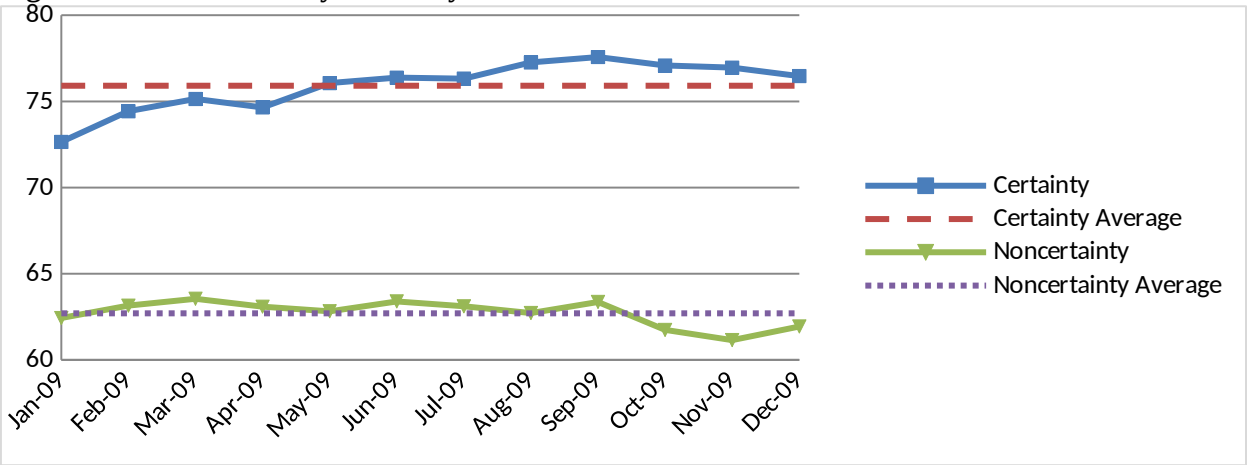
Figure 2 gives the URR over the 12 study months. The average program level URR over the period was 67.1%.

Figure 2: MRTS URRs (All) for 2009



The URR for the certainty component was on average much higher than its noncertainty counterpart, the average was 75.9%, while the average noncertainty rate was 62.7% (see Figure 3 below). For a detailed table of URRs, see Attachment 2.

Figure 3: MRTS URRs by Certainty Status for 2009



The program level average TQRR for sales was 78.3%. The TQRR for certainty units was consistently high, averaging 94.4%. However, the TQRR for the noncertainty component was relatively much lower, averaging 58.3% (see Figure 4). Similarly, TQRRs for certainty units by subsector were high, approximately 85%, in all but one sector, while the noncertainty TQRR ranged from 30.5% to 62.8%. For additional information, see Attachment 3.

Figure 4: MRTS Sales TQRRs by Certainty Status for 2009

Figure 4 Redacted

Given these results, we cannot rule out the potential for nonresponse bias in estimates of sales. Particularly concerning were the low TQRRs among noncertainty units, which in some subgroups was as low as 30.5%.

5.2 Comparison of ARTS and MRTS Estimates

To further evaluate the potential for nonresponse bias in MRTS estimates, we compared 2008 ARTS year-end inventories estimates prior to benchmarking with December 2008 MRTS year-end inventories estimates prior to benchmarking.[1] We consider ARTS estimates the "gold standard" as ARTS is a mandatory survey with higher response rates than MRTS. Since inventories and sales are highly correlated (0.9 unweighted and 0.7 weighted), evidence of nonresponse bias in inventories estimates is also evidence of bias in sales estimates.

For our analysis, we used the following two-sample *t*-test:

$$t^{¿} = \frac{\hat{x}_A - \hat{x}_M}{\sqrt{\left(\hat{s}_A\right)^2 + \left(\hat{s}_M\right)^2 - 2\rho_{A,M}\left(\hat{s}_A\right)\left(\hat{s}_M\right)}}$$

[1] ARTS benchmarks to the Economic Census and MRTS benchmarks to ARTS; therefore, it would be too confounding to use benchmarked estimates.

where
$\hat{x}_A$ is the ARTS estimated value of the variable of interest,
$\hat{x}_M$ is the MRTS estimated value of the variable of interest,
$\hat{s}_A$ is the ARTS estimated standard error of the variable of interest,
$\hat{s}_M$ is the MRTS estimated standard error of the variable of interest,
$\rho_{A,M}$ is the correlation coefficient between the ARTS and MRTS variable of interest, and
$H_0: \hat{x}_A = \hat{x}_M$ vs. $H_1: \hat{x}_A \neq \hat{x}_M$.

Since both ARTS and MRTS use 16 random groups (RGs) in their variance estimation, under $H_0, t^{¿}$ t_{15}. As we do not have a precise measure of $\rho_{A,M}$, a sensitivity analysis was conducted where the value for $\rho_{A,M}$ was changed in 0.1 increments, from zero to one. Negative values were not used as we were reasonably sure, *a priori*, that there was a positive correlation between ARTS and MRTS.

It can be intuited from the $t^{¿}$ formula that as $\rho_{A,M}$ increases, and holding all else constant, the value of the denominator decreases, yielding a larger $t^{¿}$ value. This results in an increased probability of rejecting the null hypothesis that there is no difference in the ARTS and MRTS estimates. This effect is seen in Table 2, which gives the resulting *p*-values from the two-sample *t*-test described above. Those values highlighted in dark grey were significant at the 0.10 α-level, while those highlighted in light grey were significant at the 0.05 α-level. We found in 97 of 120 cells that the null hypothesis that the totals equal was rejected.

Table 2: *p*-values Based on the Difference of ARTS and MRTS Inventories Estimates by 3-digit NAICS (Values in dark grey are significant at the α = 0.10 level, in light grey at the α = 0.05 level.)

NAICS/ $\rho_{A,M}$	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
441	0.310	0.288	0.264	0.238	0.210	0.179	0.145	0.108	0.070	0.033	0.007
442	0.029	0.023	0.017	0.012	0.007	0.004	0.002	0.001	0.000	0.000	0.000
443	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
444	0.029	0.023	0.018	0.013	0.009	0.005	0.003	0.001	0.000	0.000	0.000
445	0.018	0.013	0.010	0.006	0.004	0.002	0.001	0.000	0.000	0.000	0.000
446	0.006	0.005	0.003	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000
447	0.418	0.394	0.368	0.337	0.302	0.262	0.214	0.159	0.094	0.029	0.000
448	0.021	0.016	0.011	0.008	0.005	0.002	0.001	0.000	0.000	0.000	0.000
451	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
452	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
453	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
454	0.619	0.600	0.579	0.554	0.524	0.487	0.440	0.377	0.288	0.154	0.001

In order to put Table 2 in context, Table 3 shows the size of the bias, with, again, the underlying assumption that ARTS inventories estimates are the "gold standard." The subsectors with the smallest number of differences across the columns also have the smallest difference in bias as a percentage of their MRTS estimates.

Table 3: Bias in MRTS Inventories Estimates by Subsector

3-Digit NAICS	MRTS Inventories ($)	ARTS Inventories ($)	Bias	Relative Bias
441			3,327,966,447	2.25
442			(1,009,326,623)	(6.04)
443			(1,823,466,746)	(12.04)
444	*Estimates Redacted*	*Estimates Redacted*	(3,597,080,216)	(6.98)
445			1,245,167,246	3.22
446			(2,214,386,300)	(6.28)
447			231,513,568	2.92
448			3,421,921,621	8.24
451			2,273,663,329	12.15
452			2,275,561,307	3.30
453			2,540,795,680	15.62
454			(320,610,417)	(1.44)

We do not attribute these differences to sampling error, since MRTS is a subset of ARTS. Instead, we attribute the large number of differences to some systematic nonsampling error that we have not yet identified. Hence, we cannot rule out error due to nonresponse bias.

5.3 Comparison of Respondents to Nonrespondents on Substantive Frame Data

MRTS handles unit and item nonresponse through imputation. Imputation cells are generally defined by 6-digit NAICS industry cross-classified by monthly sales quartiles. Imputation cells can contain both certainty and noncertainty units. A ratio-of-identicals model is used for imputation.

Within a given imputation cell the model is

$$y_i = x_i \times \frac{\sum_{j=1}^{M} w_j y_j k_j}{\sum_{j=1}^{M} w_j x_j k_j},$$

where
y_i = Current month sales (inventories) for sample unit i,
x_i = Prior month sales (inventories) for sample unit i,
y_j = Current month sales (inventories) for reporting unit j,
x_j = Prior month sales (inventories) for reporting unit j,
w_j = Current month weight for reporting unit j,
k_j = Current month kind-of-business (KB) factor for reporting unit j, and
M = Total number of respondent units in the current statistical period.

The KB factor represents the proportion of business a firm conducts within a given industry. Respondents are in the imputation cell for the NAICS industry that represents at least 80% of the type of business the company performs. Approximately 28% of the total retail and food services sales estimate and 32% of the total retail inventories estimate is based on imputed data.

The underlying assumption is that the ratio of the current-to-prior month sales is missing at random (MAR) with "ignorable" missingness within imputation cells. Evidence that respondent and nonrespondent values differ within cells is evidence of nonresponse bias. The only frame data available for analysis, however, was annualized sales, which we suspect is highly related to the current to prior month sales ratio. Therefore, we tested the hypothesis that there was no difference in mean sales for respondents and nonrespondents within imputation cell.

The test statistic in each imputation cell *h* was computed as

$$t_h^{\iota} = \left(\hat{y}_{R,h} - \hat{y}_{NR,h}\right) \sqrt{\hat{\hat{v}}\left(\hat{y}_{R,h}\right) + \hat{\hat{v}}\left(\hat{y}_{NR,h}\right) - \sqrt{\hat{\hat{v}}\left(\hat{y}_{R,h}\right)\hat{\hat{v}}\left(\hat{y}_{R,h}\right)}\,\hat{\hat{\rho}}_{R.NR,h}}$$

where
$\hat{y}_{R,h}$ is the Hàjak estimator of the respondent mean in imputation cell *h*,
$\hat{y}_{NR,h}$ is the Hàjak estimator of the nonrespondent mean in imputation cell *h*,
$\hat{\hat{v}}\left(\hat{y}_{R,h}\right)$ is the 12-month mean RG variance estimate of the respondent mean in imputation cell h,
$\hat{\hat{v}}\left(\hat{y}_{NR,h}\right)$ is the 12-month mean RG variance estimate of the nonrespondent mean in imputation cell *h*,
$\hat{\hat{\rho}}_{R.NR,h} C\hat{o}v\left(\hat{y}_{R,h}, \hat{y}_{NR,h}\right)$ is the 12-month mean RG lag 1 autocorrelation estimate in imputation cell *h*.

Since MRTS uses 16 RGs, under H_0, $t_h^* \sim t(15)$. To reduce the variability of the variance estimates, we used averaged standard errors and lag 1 autocorrelations in our analyses; autocorrelations for all other lags were not significantly different from zero. Note that because MRTS uses the RG variance estimator, the analysis was limited to imputation cells that contained both certainty and noncertainty units. The RG method breaks down when sample weights are all one. Therefore, all such imputation cells were excluded from analysis.

Table 4 contains the results of the *t*-tests. The null hypothesis was not accepted on average 15.5% of the time. Within NAICS sectors such as 441 and 448 the same imputation cells were consistently flagged. Within other NAICS sectors, such as 444, 452, and 722, the rejection cells were more variable over time.

Based on these results, we cannot deduce from our results that the data are MAR with ignorable missingness within imputation cells. Moreover, these systematic differences are evidence of nonresponse bias, as we use respondent data to impute nonrespondent data.

Table 4: Summary of *t*-test Results for the Differences in Respondent and Nonrespondent Means Within Imputation Cell

NAICS		Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09	Jul-09	Aug-09	Sep-09	Oct-09	Nov-09	Dec-09
441	**Total Cells**	20	20	19	20	20	20	20	20	20	20	20	20
	Diff. Dists	7	7	7	5	6	6	3	6	7	6	6	6
	Percentage	36.8%	36.8%	36.8%	25.0%	30.0%	30.0%	15.0%	30.0%	35.0%	30.0%	30.0%	30.0%
442 & 443	**Total Cells**	26	26	26	26	26	26	26	25	25	25	25	25
	Diff. Dists	5	5	4	4	5	6	7	5	3	5	5	4
	Percentage	19.2%	19.2%	15.4%	15.4%	19.2%	23.1%	26.9%	20.0%	12.0%	20.0%	20.0%	18.2%
444	**Total Cells**	14	14	13	13	13	14	14	14	14	13	14	14
	Diff. Dists	2	1	1	0	0	1	1	1	0	0	0	0
	Percentage	14.3%	7.1%	7.7%	0.0%	0.0%	7.1%	7.1%	7.1%	0.0%	0.0%	0.0%	0.0%
445	**Total Cells**	22	22	22	22	22	22	22	22	22	22	22	21
	Diff. Dists	6	4	2	3	3	3	2	2	4	1	5	4
	Percentage	27.3%	18.2%	9.1%	13.6%	13.6%	13.6%	9.1%	9.1%	18.2%	4.5%	22.7%	19.0%
446	**Total Cells**	9	9	9	9	8	9	8	8	9	8	9	9
	Diff. Dists	0	1	1	1	0	1	0	0	1	0	1	1
	Percentage	0.0%	11.1%	11.1%	11.1%	0.0%	11.1%	0.0%	0.0%	11.1%	0.0%	11.1%	11.1%

447	**Total Cells**	4	4	4	4	4	4	4	4	4	4	4	4
	Diff. Dists	2	3	2	3	3	3	3	1	2	2	2	2
	Percentage	50.0%	75.0%	50.0%	75.0%	75.0%	75.0%	75.0%	25.0%	50.0%	50.0%	50.0%	50.0%
448	**Total Cells**	21	20	20	20	20	20	20	20	20	20	21	19
	Diff. Dists	3	3	4	3	3	3	3	3	2	3	3	2
	Percentage	14.3%	15.0%	20.0%	15.0%	15.0%	15.0%	15.0%	15.0%	10.0%	15.0%	14.3%	10.5%
451	**Total Cells**	15	14	14	14	14	15	15	15	15	15	15	15
	Diff. Dists	1	2	3	2	2	1	1	2	1	2	2	3
	Percentage	6.7%	14.3%	21.4%	14.3%	14.3%	6.7%	6.7%	13.3%	6.7%	13.3%	13.3%	20.0%
452	**Total Cells**	3	3	3	3	3	3	3	3	3	3	3	3
	Diff. Dists	0	0	0	1	0	0	0	0	0	0	0	0
	Percentage	0.0%	0.0%	0.0%	33.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
453	**Total Cells**	20	20	20	20	20	20	20	20	20	20	20	20
	Diff. Dists	0	0	1	0	1	0	0	0	0	0	0	0
	Percentage	0.0%	0.0%	5.0%	0.0%	5.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
454	**Total Cells**	16	16	15	15	15	15	16	15	15	15	16	15
	Diff. Dists	5	5	4	2	4	6	4	5	5	5	4	4
	Percentage	31.3%	31.3%	26.7%	13.3%	26.7%	40.0%	25.0%	33.3%	33.3%	33.3%	25.0%	20.0%
722	**Total Cells**	23	23	23	23	23	23	23	23	23	23	23	23
	Diff. Dists	3	1	2	2	2	3	1	3	3	4	3	2
	Percentage	13.0%	4.3%	8.7%	8.7%	8.7%	13.0%	4.3%	13.0%	13.0%	17.4%	13.0%	8.7%

6. Limitations

One of the challenges facing those of us tasked with conducting post-production analyses of data quality is that surveys are often ill-designed to conduct such tasks. We do not truly know what the missing data look like, and we often have little measureable information about the nature of the nonrespondents. Appropriate planning at the survey design stage is critically important for understanding if and how nonrespondents differ from respondents. In this section we discuss some of the limitations to our analysis.

To begin, the analysis was limited by the amount of available data. There were MRTS and MARTS survey data, but no auxiliary data which could have allowed us to assess trends in nonresponse. Ideally, there are frame data or other covariates available for analysis for studies such as this one. Some of the analysis we had planned, we were unable to complete because of incomplete auxiliary information. In one case, we attempted to examine nonresponse by type (e.g., refusal, noncontact, etc.), but we learned that such information is not consistently recorded for MARTS, although this is something the program plans to keep track of in the future.

We also make the observation that ARTS, MRTS, and MARTS are linked in ways that are not clearly documented and understood. An estimate of MARTS sales, the best we can tell, is derived from at least five data sources: MARTS, MRTS, ARTS, the Economic Census, and administrative data, all of which have different error properties. Similarly, extensive benchmarking goes on for these surveys. ARTS benchmarks to the Economic Census, MRTS benchmarks to ARTS, and MARTS, through the use of the link relative estimator, benchmarks to MRTS. It is not clear if we know about the error properties of estimates after benchmarking. Moreover, we often take at face value that the Economic Census is the "gold standard" for current business surveys, when in fact, in our estimation, little is known about this survey's data quality (e.g., the lack of response metrics and the extensive use of administrative data filling).

Additionally, we learned mid-way through our analysis that MARTS reported sales data are also treated as reported data for MRTS. How this might have affected the analysis is not entirely clear. We also learned that a reporting unit is assigned to an imputation cell based on the predominant KB factor, which could possibly skew sales estimates.

7. Conclusions and Recommendations

The purpose of this study was to identify areas of potential nonresponse bias and make recommendations related to mitigating sources of error in MARTS sales and change in sales estimates. Low MARTS response rates and results from Madow and Madow (1978), coupled with evidence of nonresponse bias in MRTS sales and inventories estimates, led us to conclude that there is evidence of nonresponse bias in MARTS estimates of sales and month-to-month relative change in sales. For sales, there is some evidence that the bias is relatively large.

Based on our results from the MARTS and MRTS analyses, we make the following recommendations:

1. Conduct an imputation/weighting study for MRTS
2. Assess reasons for nonsampling error in MRTS estimates
3. Target MRTS and MARTS nonresponse follow-up to subgroups with low response rates
4. Update MARTS with supplemental sample
5. Plan a nonresponse bias study for MARTS at the survey design stage, or prior to data collection
6. Study the relationship between MARTS, MRTS and ARTS estimates

The recommendations are detailed below, but can be summed up as follows: take steps to improve MRTS estimates of total sales and maintain MARTS estimates through the use of supplemental sample.

In a comparison of MRTS respondents and nonrespondents using frame data (annual sales), we found systematic differences within imputation cells. These systematic differences are evidence of nonresponse bias, as we use respondent data to impute nonrespondent data. Given MARTS use of the link relative estimator, this is also evidence of nonresponse bias in MARTS estimates of sales. Improving MRTS estimates could also improve MARTS estimates. Therefore, we recommend conducting an imputation study aimed at determining how to best mitigate nonresponse bias in MRTS sales estimates.

In a comparison of ARTS and MRTS inventories estimates by subsector, in 97 of 120 cells the null hypothesis that total inventories is statistically the same was rejected (see Table 2). We do not attribute these results to sampling error, because MRTS shares its survey units with ARTS. We do, however, attribute the large number of differences to some systematic nonsampling error, which does not rule out the possibility of nonresponse bias. To address the evidence of unspecified nonsampling error, we recommend research into why the differences exist. Given that inventories and sales are highly related, improving MRTS inventories estimates could also improve MRTS and, ultimately, MARTS sales estimates.

From our response rate analysis, low MARTS and MRTS response rates were an indication of nonresponse bias, particularly at subgroup levels. There was evidence of low response rates by certainty status and by subsector. Targeted nonresponse follow-up will help bring up response and possibly reduce nonresponse bias.

Madow and Madow (1978) suggest that activities such as mergers, splits, deaths, births, purchases, and sales are indicative of bias. They recommend adding supplemental sample or doing nonresponse adjustments to correct for the bias. Changes to the sample may mean that variance constraints are no longer being met. We reiterate their suggestion, given that the sample currently is not redesigned or updated for 2 ½ to 3 year intervals.

Given our challenges conducting a post-production analysis of MARTS nonrepsonse bias, we suggest planning a nonresponse bias study for MARTS prior to data collection. One such study might entail subsampling nonrespondents. This is probably the best but most expensive option for learning about the differences between respondents and nonrespondents. If resources are an issue, we recommend a study,

when enough MARTS data are available in StEPS, comparing MARTS change in sales to MRTS change in sales for respondents and nonrespondents.

Finally, we suggest a study into the effects of using information from multiple data sources on estimates of sales and change in sales. The retail sector surveys are highly dependent on one another (e.g., benchmarking), and it is difficult to disentangle these surveys for analytic purposes. We recommend considering less dependency in survey estimators.

8. References

Andridge, R. R. and Little, R. J. A. 2008. "Proxy Pattern-Mixture Analysis for Survey Nonresponse." Joint Statistical Meetings: Section on Survey Research Methods.

Huang, E. T. 1984. An Imputation Study for the Monthly Retail Trade Survey. U.S. Census Bureau Statistical Research Division Report Series, CENSUS/SRD/RR-84-13.

Huang, E. T. 1986. Report on the Imputation Research for the Monthly Retail Trade Survey. U.S. Census Bureau Statistical Research Division Report Series, CENSUS/SRD/RR-86-09.

Lineback, J. F. and Fink, E. B. 2011. "Monthly Retail Trade Survey Nonresponse Bias Analysis Report." Internal Census Bureau Memorandum.

Madow, L. H. and Madow, W. G. 1978. "On Link Relative Estimators." Joint Statistical Meetings: Section on Survey Research Methods.

Rosenthal, M. and Davie, W. 2008. "Nonresponse Bias Analysis for the Quarterly Services Survey." Internal Census Bureau Memorandum.

Thompson, K. J., and Oliver, B. 2010. "Monthly Retail Trade Survey Nonresponse Bias Analysis Report." Internal Census Bureau Memorandum.

Attachment 1

MARTS URRs

	Sector	Average	Aug 2010	Sep 2010	Oct 2010
PROGRAM	All	58.0	53.7	61.5	58.9
	441	64.9	66.7	70.0	58.1
	442	62.3	60.5	62.7	63.7
	443	65.1	67.1	69.1	59.3
	444	60.9	56.7	58.8	67.2
	445	56.2	55.1	57.9	55.6
	446	51.4	47.1	53.2	53.9
	447	60.7	65.6	68.3	48.3
	448	64.8	61.9	65.5	66.9
	451	70.7	73.2	78.9	60.1
	452	60.3	52.9	52.9	75.0
	453	60.0	63.0	64.9	52.0
	454	61.3	57.5	62.0	64.4
	722	57.2	53.7	61.5	56.5
	All	61.9	60.9	63.8	61.1
	441	50.2	47.1	53.3	50.2

CERTAINTY	442	70.1	71.8	73.1	65.4
	443	61.0	59.2	61.2	62.5
	444	56.5	56.1	55.2	58.2
	445	61.5	61.3	63.3	59.9
	446	53.2	50.0	56.5	53.2
	447	62.5	58.2	64.8	64.5
	448	75.6	74.7	75.9	76.3
	451	64.7	62.0	67.6	64.4
	452	76.4	72.4	79.3	77.6
	453	44.6	46.7	45.6	41.4
	454	62.3	62.2	62.2	62.5
	722	69.9	70.8	74.3	64.6
NON-CERTAINTY	All	58.9	57.2	61.4	57.9
	441	62.3	58.2	67.2	61.5
	442	64.8	63.4	68.0	63.0
	443	60.9	61.1	63.3	58.2
	444	74.4	74.3	78.2	70.6
	445	55.0	54.4	56.6	54.1
	446	57.4	59.0	59.0	54.2
	447	44.3	41.7	47.6	43.6
	448	60.1	57.7	61.7	61.0
	451	61.5	61.9	64.3	58.3
	452	73.5	76.9	76.9	66.7
	453	56.1	56.0	56.5	55.8
	454	65.7	63.7	67.6	65.7
	722	54.2	51.8	56.8	54.1

Attachment 2

MRTS URRs

	Sector	Average	Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09	Jul-09	Aug-09	Sep-09	Oct-09	Nov-09	Dec-09
PROGRAM	All	67.1	66.0	67.0	67.5	67.1	67.3	67.8	67.6	67.5	68.0	66.8	66.3	66.7
	441	69.6	69.3	69.6	70.6	68.6	68.7	69.6	69.1	69.3	70.7	69.4	70.2	69.9
	442	61.7	61.8	62.2	61.9	62.0	61.1	61.0	61.0	62.9	63.0	61.3	60.5	61.2
	443	65.4	67.1	67.1	66.6	67.8	66.2	66.4	65.7	62.6	64.2	64.6	64.2	62.6
	444	75.4	74.7	73.8	75.7	76.7	76.2	76.6	76.8	75.9	75.4	75.0	73.8	74.6
	445	68.5	65.1	68.6	68.8	68.4	69.4	70.1	69.7	68.9	69.7	67.9	68.2	67.5
	446	61.4	56.4	60.3	61.2	57.9	60.7	61.3	60.3	63.9	65.8	63.7	62.9	62.9
	447	66.5	67.7	66.8	67.8	66.9	66.3	66.3	66.4	67.0	66.2	66.1	65.3	65.5
	448	65.5	64.6	65.9	66.9	65.5	65.9	66.6	66.2	66.0	64.8	64.4	63.6	65.3
	451	67.9	67.8	67.5	67.8	67.1	69.9	69.7	68.8	68.0	68.9	65.9	66.4	67.1
	452	62.5	63.6	62.7	63.3	61.0	60.6	61.1	63.8	64.6	63.0	62.9	61.2	62.5
	453	60.6	59.2	60.9	61.4	61.0	61.2	60.2	61.4	60.5	61.7	60.6	60.1	59.0
	454	68.2	66.7	68.0	68.0	67.9	68.0	68.9	68.0	69.5	69.3	68.4	66.9	68.2
	722	68.7	66.4	67.6	67.6	69.1	69.3	70.4	70.2	69.4	70.5	68.2	67.0	69.2

CERTAINTY	All	75.9	72.7	74.4	75.1	74.7	76.1	76.4	76.3	77.3	77.6	77.1	77.0	76.5
	441	75.9	73.2	75.0	75.5	73.5	74.4	76.6	75.5	77.1	77.8	77.5	77.5	77.5
	442	73.2	69.9	71.1	71.7	72.0	72.7	71.5	71.6	76.6	77.2	74.3	75.0	74.9
	443	78.2	78.2	77.8	77.6	78.2	78.4	79.2	80.5	76.2	78.5	79.7	78.5	76.0
	444	81.9	78.9	79.5	81.7	82.4	82.8	83.3	83.3	82.4	82.4	83.1	81.3	81.3
	445	77.3	71.3	76.9	76.5	76.0	78.0	78.8	78.9	78.7	80.0	77.5	77.6	77.6
	446	71.0	61.1	66.6	71.3	65.7	69.4	69.4	69.4	72.7	77.9	73.4	79.0	76.6
	447	77.0	75.7	75.7	78.5	76.9	79.5	75.7	75.9	76.7	77.1	77.9	77.9	77.1
	448	79.0	76.3	76.5	77.5	77.1	78.6	80.1	80.8	84.2	79.2	79.1	78.4	80.6
	451	79.4	78.6	78.1	77.9	76.4	81.0	80.2	79.9	79.0	80.6	77.7	81.6	81.4
	452	64.9	65.8	64.6	65.3	63.0	62.8	62.4	67.2	67.7	65.5	65.3	65.2	64.3
	453	67.1	63.8	66.4	66.9	66.9	68.6	67.7	67.1	66.7	69.6	69.8	67.3	64.0
	454	74.9	72.5	74.0	74.7	75.0	74.3	75.5	74.2	77.3	75.4	75.7	75.1	74.6
	722	81.6	74.6	76.8	77.2	79.8	83.6	83.2	82.6	82.6	84.3	85.2	84.8	84.2
NONCERTAINTY	All	62.7	62.4	63.2	63.6	63.1	62.8	63.4	63.1	62.7	63.4	61.7	61.2	61.9
	441	66.6	67.4	66.9	68.3	66.2	66.0	66.4	66.1	65.8	67.5	65.7	66.8	66.4
	442	57.8	58.8	58.9	58.4	58.4	57.2	57.4	57.4	58.6	58.5	57.1	55.9	56.9
	443	56.8	59.2	59.7	59.0	60.5	57.9	57.7	55.7	53.8	54.9	54.8	54.8	53.9
	444	73.2	73.2	71.8	73.6	74.8	73.9	74.3	74.5	73.7	73.0	72.2	71.3	72.4
	445	63.9	61.6	64.1	64.6	64.2	64.9	65.5	64.9	63.9	64.4	62.9	63.4	62.3
	446	56.6	53.8	56.9	55.7	53.7	56.1	57.1	55.4	59.5	60.0	59.1	55.1	56.4
	447	61.7	63.7	62.5	62.7	62.1	60.3	62.0	62.0	62.7	61.4	60.8	59.7	60.4
	448	58.4	57.9	60.1	61.0	59.1	59.2	59.5	58.6	57.0	57.7	57.2	56.3	57.6
	451	61.5	61.3	61.2	62.0	61.7	63.6	63.6	62.5	62.2	62.5	59.5	58.3	59.4
	452	51.1	52.6	53.9	53.9	51.3	50.0	55.0	47.5	50.0	51.2	51.3	42.5	53.9
	453	58.1	57.1	58.7	59.2	58.7	58.3	57.2	59.1	58.2	58.7	57.2	57.4	57.2
	454	61.8	60.8	62.1	61.4	60.9	61.9	62.7	62.0	62.3	63.8	61.8	59.5	62.3
	722	64.6	63.4	64.4	64.3	65.4	64.6	66.2	66.2	65.3	66.3	63.0	61.6	64.6

Attachment 3

MRTS Sales TQRRs*

	Sector	Average	Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09	Jul-09	Aug-09	Sep-09	Oct-09	Nov-09	Dec-09
PROGRAM	All	70	70	70	70	70	70	70	70	70	70	70	70	70
	441	60	60	60	60	60	60	60	60	60	60	60	60	60
	442	60	60	60	60	60	60	60	60	60	60	60	60	60
	443	70	70	70	70	70	70	70	70	70	70	70	70	70
	444	70	70	70	70	70	70	70	70	70	70	70	70	70
	445	80	80	80	80	80	80	80	80	80	80	80	80	80
	446	80	80	80	80	80	80	80	80	80	80	80	80	80
	447	70	70	70	70	70	70	70	70	70	60	70	70	70
	448	80	80	70	80	80	80	80	80	80	80	80	80	80
	451	70	70	70	70	70	70	70	70	70	70	70	70	70
	452	90	90	90	90	90	90	90	90	90	90	90	90	90
	453	60	60	60	60	60	60	60	60	60	60	60	60	60
	454	70	70	70	70	70	70	70	70	70	70	70	70	70
	722	60	70	70	70	70	70	60	60	60	60	60	60	60
CE	All	90	90	90	90	90	90	90	90	90	90	90	90	90
	441	80	80	80	80	80	80	80	80	80	80	80	80	80

RTAINTY	442	80	80	80	80	80	80	80	80	80	80	80	80	80
	443	90	90	90	90	90	90	90	90	90	90	90	90	90
	444	90	90	90	90	90	90	90	90	90	90	90	90	90
	445	90	90	90	90	90	90	90	90	90	90	90	90	90
	446	90	90	90	90	90	90	90	90	90	90	90	90	90
	447	90	90	90	90	90	90	90	90	90	70	80	90	90
	448	90	90	80	90	90	90	90	90	90	90	90	90	90
	451	90	90	90	90	90	90	90	90	90	90	90	90	90
	452	90	90	90	90	90	90	90	90	90	90	90	90	90
	453	90	90	90	90	90	90	80	80	90	90	90	80	80
	454	90	90	90	80	80	90	90	90	90	90	80	90	90
	722	90	90	90	90	90	90	90	90	90	90	90	90	90
NONCERTAINTY	All	50	50	50	50	50	50	50	50	50	50	50	50	50
	441	60	60	60	60	60	60	60	60	60	60	60	60	60
	442	50	50	50	50	50	50	50	50	50	50	50	50	50
	443	30	30	30	30	30	30	30	30	30	30	20	20	20
	444	50	50	50	50	50	50	50	50	50	50	50	50	50
	445	60	60	60	60	60	60	60	60	60	60	60	60	50
	446	50	40	50	40	40	40	50	50	50	50	50	50	40
	447	50	50	50	50	50	50	50	50	50	50	50	50	50
	448	50	60	50	50	50	50	50	50	50	60	50	50	50
	451	40	40	40	40	50	40	50	50	50	50	40	40	40
	452	50	50	50	50	50	50	50	50	50	50	50	50	50
	453	50	50	50	50	50	50	50	50	50	50	40	50	50
	454	40	50	40	40	40	40	40	40	40	40	40	40	40
	722	60	60	60	60	60	60	60	60	50	60	50	50	50

*Information redacted through the use of a rounding algorithm.