A Model-Based Approach to Assessing Nonresponse Bias for the Monthly Wholesale Trade Survey

Joanna Fane Lineback and Eric B. Fink[1]
U.S. Census Bureau, Washington, D.C. 20233

Abstract
This paper addresses the topic of nonresponse bias in business surveys. Some common techniques for investigating nonresponse bias post-data collection have been established from a design-based perspective. Many of these techniques, however, cannot account for small nonresponse adjustment cells which are common for business surveys. For the Monthly Wholesale Trade Survey nonresponse bias analysis, we tried alternative, model-based approaches. To evaluate the potential for nonresponse bias, we examined propensity and prediction response models using frame data compiled from auxiliary data sources, including the Economic Census and the American Community Survey. Additionally, we modeled response propensity and prediction given the current adjustment cells variables. To minimize variance and bias, the variables used to define nonresponse adjustment cells should be highly predictive of key survey estimates and the likelihood of responding to the survey. Evidence otherwise is evidence of nonresponse bias. Our findings are discussed.

Key words and phrases: model-based, design-based, auxiliary data, response propensity, response prediction, nonresponse bias

1. Introduction

With the rising costs of maintaining surveys' response rates, nonresponse bias analysis has become the focus of many studies. The Economic Programs Directorate at the U.S. Census Bureau began formally conducting nonresponse bias studies in 2008. Many of the diagnostics used in the first studies were loosely based on sets of tools presented by Groves and Brick (2005) and at the Federal Committee on Statistical Methodology Workshop on 'How to Do Nonresponse Bias Analysis in Household and Establishment Surveys' held in June 2009. Economic Directorate staff members have used many of these tools. For example, see Lineback and Thompson (2010). Many of the tools, however, do not account for small nonresponse adjustment cells which are common for business surveys. This has led us to look for more robust techniques.

For the Monthly Wholesale Trade Survey (MWTS) nonresponse bias analysis, we investigate model-based approaches. We use a technique proposed by Andrige and Little (2011) to assess the potential for nonresponse bias in estimates of sales, relative change in sales, inventories, and relative change in inventories. We use a technique proposed by Little and Vartivarian (2005) to evaluate the effectiveness of MWTS nonresponse adjustments at mitigating potential nonresponse bias. For the analysis, we use relevant frame data including 2007 Economic Census employment, annual payroll, sales, and operating expenses data and American Community Survey data.

---

[1] Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

The remainder of this paper is organized as follows. Section 2 provides background information. Section 3 describes the analysis techniques. Section 4 gives the results. Finally, Section 5 provides a discussion of the results and future research.

2. Background

The MWTS provides estimates of end-of-month inventories, monthly sales, and month-to-month relative change in sales and inventories for employer businesses classified as merchant wholesalers, excluding manufacturers' sales branches, as defined by the North American Industry Classification System (NAICS). The month-to-month relative change in inventories is the key statistic produced from this survey.

The motivation for conducting a nonresponse bias study for the MWTS is response rates that have consistently failed to meet target levels. In the remainder of this section, we discuss the survey design aspects critical for conducting a meaningful post-data collection analysis of nonresponse bias: sample design, data collection, nonresponse adjustment, and estimation.

2.1 Sample Design

The MWTS sample is a subsample of units selected for the Annual Wholesale Trade Survey. The sampling frame for these surveys is based on data extracted from the Business Register for all employer establishments located in the United States and operating in merchant wholesale. Here, an establishment is defined as a single physical location where business transactions take place and for which payroll and employment records are kept. The extracted data items include primary identifiers, sales, payroll, employment, business name and address, NAICS code, and other business data.

The sampling frame contains two types of sampling units: Employer Identification Numbers (EINs) and companies. Both represent clusters of one or more establishments owned or controlled by the same firm. The frame is stratified by industry group, which is based on the detail required for publication. A sampling unit is assigned to the industry group in which it has the largest proportion of sales relative to the sampling unit's measure of size (MOS). To calculate the MOS, establishments' estimated annual sales are aggregated to both the company and EIN levels. Sample sizes are based on sales and inventories coefficient of variation constraints. Sampling units are further stratified within industry group by MOS. Sampling units expected to have a large effect on estimates' precision are selected with certainty (i.e., with a sampling weight of one). All firms not selected with certainty are sampled on an EIN basis. For a firm with more than one EIN, each EIN is treated as a separate non-certainty sampling unit. Within each non-certainty stratum, a simple random sample of EINs is selected without replacement. The final MWTS sample includes approximately 1,700 companies and 2,700 EINs.

The MWTS sample is redesigned and reselected at approximately 5-year intervals. It is also updated quarterly to reflect births (i.e., new EINs that appear on the Business Register) and other changes to the business universe.

2.2 Data Collection

Each month, firms in the MWTS sample are asked to report their sales and inventories data for the month just ending. Data are collected by mail, facsimile, Internet, and telephone. Between 50 and 60% of the responses are received via the Internet.

For large, diverse companies, reporting units are created to facilitate the collection and tabulation of data by industry. For a large company that is not able to provide data by industry, the data are split into tabulation units (i.e., units used for estimation) based on the company's percentage of activity in each industry as calculated from frame data. This is referred to as the kind-of-business factor. For sampling units that have activity in only one industry, the reporting unit and tabulation unit are usually equivalent to the sampling unit.

2.3 Nonresponse Adjustments

The MWTS uses imputation to account for unit and item nonresponse and edit failures. Approximately 28% of the total wholesale sales estimate and 32% of the total wholesale inventories estimate are based on imputed data.

The ratio model $y_i = Rx_i$ is used for imputation. Within a given imputation cell,

$$y_i = \left(\frac{\sum_{j=1}^{M} w_j y_j k_j}{\sum_{j=1}^{M} w_j x_j k_j}\right) x_i,$$

where

M = Total number of reporting units with reported data in both the current and prior months,

y_i = Imputed current month sales (inventories) for reporting unit i,

x_i = Prior month sales (inventories) for reporting 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 final weight for reporting unit j, and

k_j = Current month kind-of-business factor for reporting unit j.

Imputation cells are created when a new sample is introduced and are used for the lifetime of the sample. Imputation cells are generally defined by industry cross classified by size (sales or inventories) quartiles. These size quartiles are based on reported data from the same month in a prior year. Two or four size cells are created based on the expected number of reporting units that will be tabulated in an industry. For an industry with few expected reporting units, the cell remains undivided.

Each month, a reporting unit is assigned to an imputation cell based on its prior month sales (inventories are used when imputing inventories). Reporting units that have current and prior month data that are reported, passed edits, and are greater than zero are used in the imputation base. For a reporting unit whose data have to be split for tabulation, the unit must have greater than 80% of its total sales (inventories) in the industry of the imputation cell to which it contributes. EINs added to the sample during the quarterly birth process are not eligible for inclusion in an imputation cell the first month they are in sample.

An imputation cell ratio calculated with too few reporting units, or that falls outside pre-determined limits, uses a collapsed ratio for imputation. The collapsed ratio is calculated using data for all reporting units in the industry that meet the criteria listed above without distinguishing units by size. Note that the collapsed imputation cell ratio is also subject to a minimum number of units and ratio limits. There is no collapsing across industries.

2.4 Estimation

Preliminary MWTS estimates are released approximately 40 days after the reference month, and revised estimates are released approximately 70 days after the reference month. The MWTS produces Horvitz-Thompson estimates of sales and inventories which are then benchmarked to Annual Wholesale Trade Survey and Economic Census totals. The estimates are also seasonally adjusted. Variance estimates are calculated using the method of random groups.

For more information, refer to
census.gov/wholesale/www/how_surveys_are_collected/monthly_methodology.html.

3. Analysis

We analyze sales, inventories, relative change in sales, and relative change in inventories estimates using the two model-based approaches discussed below. In general, a model-based approach assumes an underlying model generated the population of interest (Sarndal, Swensson, and Wretman 2003). Such an approach is dependent on the availability of good covariates.

3.1 Examining the Potential for Nonresponse Bias in Reported Data

Andridge and Little (2011) treat nonresponse bias as a function of three key components: response rates, response propensity, and response prediction. By themselves, response rates tell us little about the potential for nonresponse bias. When used with information about the likelihood of response and the ability to predict response, we paint a truer picture of the effect of survey nonresponse.

The authors assume a variable Y and a set of covariates, Z, then they find the proxy variable, X, that is most highly correlated with Y (i.e., the response prediction component). X is a strong proxy for Y if the correlation, ρ, is high (e.g., 0.8). Conversely, if ρ is low (e.g., 0.2), the relationship is weak.

Given the proxy X, the extent to which the population estimate differs from the respondent-only estimate provides information about the missing data mechanism (i.e., the propensity component). They consider the difference $\bar{x} - \bar{x}_R$ where R represents respondent-only values. The larger the difference, the more evidence the data are not missing completely at random.

The authors analyze the prediction and propensity components together, proposing a sensitivity analysis to explore the deviation of an adjusted estimator from missing at random. If X is a strong proxy and the difference $\bar{x} - \bar{x}_R$ is small, then there is little evidence of nonresponse bias in Y. Or, if X is a strong proxy but the difference is large, then there is evidence of nonresponse bias in Y, *but* information is available to correct for the bias. Finally, if X is a weak proxy, whether the difference is small or large, then there is little to say about the potential for nonresponse bias in Y given the possibility of unobserved covariates.

For our purposes, we do not consider a sensitivity analysis for an adjusted estimator. Our analysis will only tell us if there is evidence of nonresponse bias in unadjusted estimates. In the following section, we explain a separate approach for evaluating the nonresponse mitigation technique.

3.2 Examining the Nonresponse Mitigation Technique

Little and Vartivarian (2005) describe adjusting for nonresponse as a bias-variance trade-off. They note that it is important that nonresponse adjustment cells are formed with response propensity and prediction in mind, and they provide the following table to illustrate this point.

Table 1. The Effect of Adjustments on Bias and Variance By Strength of Association of Adjustment Cell Variables With Nonresponse and Outcome

		Association with Outcome	
		Low	High
Association with Nonresponse	Low	Bias – Variance –	Bias – Variance ↓
	High	Bias – Variance ↑	Bias ↓ Variance ↓

From the table, we see that it is not enough that adjustment cell variables are highly predictive of response. In fact, this could increase the overall variance without correcting for bias. The best case scenario is when the variables used to form adjustment cells are associated with the likelihood of responding and the outcome. In this case, we are correcting for both variance and bias.

While the authors' frame the problem in terms of weighting adjustments, we use them in context of imputation adjustment cells and, therefore, are not accounting for any imputation model misspecification within cells.

4. Results

In this section, we present the results of our analysis.

4.1 Response Rates

The Census Bureau uses standard formulas for calculating unit and item response rates for economic survey programs. The unit response rate (URR) is the unweighted proportion of valid responses received from eligible units. The item response rate, or total quantity response rate (TQRR) as it is referred to by economic programs, is the weighted proportion of an estimate received from reported and "equivalent quality" sources (U.S. Census Bureau 2012). The December 2008 MWTS response rates are given in Table 2.

Table 2. December 2008 MWTS Response Rates, Overall and by Certainty Status

	Program	Certainty	Non-certainty
URR	75.5	81.7	71.0
Sales TQRR	71.8	81.3	54.1
Inventories TQRR	74.3	88.5	61.5

We see that the response rates among certainty units are higher than non-certainty units. This is a typical result for a business survey, because follow-up effort focuses largely on certainty units which tend to contribute more to estimates of totals.

4.2 Proxy Models

Using December 2008 MWTS respondent-only data, we fit models for sales, relative change in sales, inventories, and relative change in inventories by regressing on a set of covariates from the 2007 Economic Census (business-level employment, annual payroll, sales, and operating expenditures). We included main effects and all two-way interaction terms and accounted for the complex sample design.

At the sector level (Wholesale NAICS 42), the correlation between sales and its proxy was 0.72. The difference between the full sample and respondent-only proxy means was -$487,695, with a corresponding 95% confidence interval (CI)—assuming the differences are Normally distributed—of (-$723,309, -$252,081). The difference in proxy mean sales is substantial given a relative absolute difference of 43% and that zero is not covered in the CI. Note that although we did not account for covariance, the covariance would have been positive resulting in narrower CIs.

We saw similar results by subsector. In subsector 423 (Durable Goods), the correlation between sales and its proxy was 0.76. The difference between the full sample and respondent-only proxy means was -$374,950 with a CI of (-$515,266, -$234,634) and a relative absolute difference of 34%. In subsector 424 (Non-durable Goods), the correlation between sales and its proxy was 0.72. The difference between the full sample and respondent-only proxy only means was -$697,381 with a CI of (-$1,251,696, -$143,066) and a relative absolute difference of 59%.

At the sector level, the correlation between inventories and its proxy was 0.74 with a difference in full sample and respondent-only mean inventories of -$835,365 (a relative absolute difference of 52%) and a corresponding CI of (-$1,136,996, -$533,734). Using the same evaluation criteria for inventories, the difference in the full sample and respondent-only proxy means is also substantial.

In subsector 423, the correlation between sales and its proxy was 0.77. The difference between the full sample and respondent-only proxy means was -$683,128 with a CI of (-$860,267, -$505,989) and a relative absolute difference of 47%. In subsector 424, the correlation between sales and its proxy was 0.74. The difference between the full sample and respondent-only proxy only means was -$1,142,679 with a CI of (-$1,750,317, -$535,041) and a relative absolute difference of 62%.

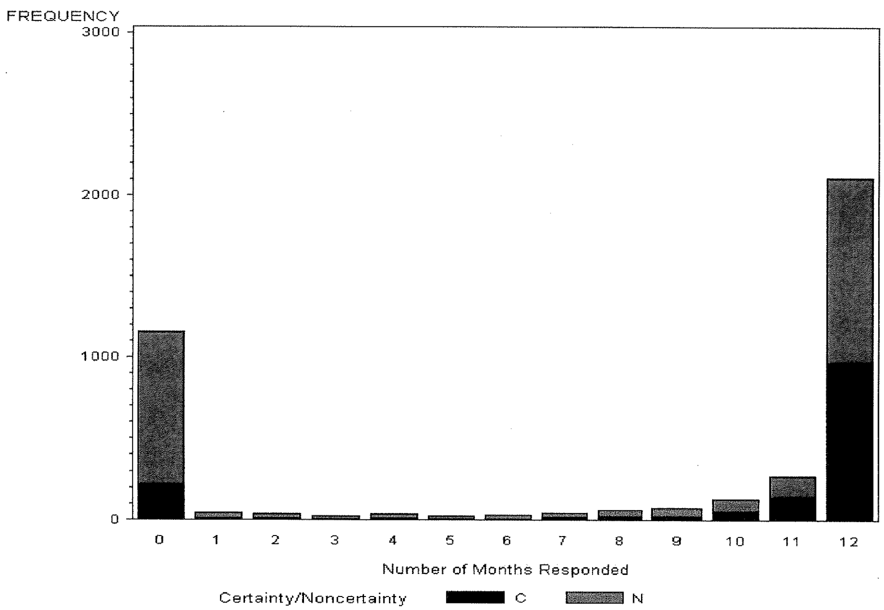
The correlation between relative change in sales and its proxy and relative change in inventories and its proxy was zero. We tried to improve the fit of all our models by including American Community Survey data, but our model fits stayed relatively unchanged. Because we could not directly evaluate the potential for nonresponse bias in relative change estimates, we turned to the theory of bias in relative change estimates. (See Appendix A for the derivation.) It tells us that in order for the relative change estimator to be approximately unbiased, the population level and the bias on the level must change at the same rate from month-to-month.

Although the results are not shown here, we also fit models for sales, inventories, relative change in sales and relative change in inventories by 4-digit NAICS (i.e., industry group), and we saw similar results. In all, there is evidence of nonresponse bias for both sales and inventories; however, there are strong proxies to correct for the bias. Based on our current results, we cannot say anything definitive about the potential for nonresponse bias in relative change estimates.

4.3 Nonresponse Mitigation

Recall that industry and MOS (reported value from the same month in a prior year) are used to form adjustment cells. For both sales and inventories, we took into account the complex sample design and regressed the current month value on the prior month value crossed with adjustment cell. Similarly, we regressed the propensity to report on adjustment cell. All four models had an R^2 or pseudo- R^2 greater than 0.98 which would seem ideal. Note, however, that no nonrespondent information is used in the imputation model (neither in the formation of adjustment cells, nor in the imputation ratio). Additionally, over the 12-month period January 2009 to December 2009 most businesses fit into one of two categories: a) responded in all 12 months, or b) did not respond at all during the 12 months (see Figure 1). Therefore, even though the adjustment cell variables are highly associated with nonresponse and outcome, we cannot conclude that they are minimizing bias.

Figure 1. Count of 2009 MWTS Respondents by Number of Months by Certainty Status



5. Discussion

The key statistic for the MWTS is month-to-month relative change in inventories. Also important are estimates of monthly total sales, month-to-month relative change in sales, and monthly total inventories. In this paper, we examined the potential for nonresponse bias in all four estimates.

Given a set of covariates from the Economic Census and the American Community Survey, we used the technique proposed by Andridge and Little (2011) to investigate the potential for nonresponse bias if we assume no nonresponse adjustment. There is evidence of potential for nonresponse bias in sales and inventories unadjusted estimates. We used a technique proposed by Little and Vartivarian (2005) to evaluate the potential for current nonresponse adjustments to minimize possible nonresponse bias. This technique assumes respondent *and* nonrespondent information is used in the formation of adjustments cells which is not the case for the MWTS. As a result, the MWTS' nonresponse adjustments are not necessarily minimizing bias, and there is evidence of potential nonresponse bias in adjusted estimates of sales and inventories.[2]

There is not enough information to draw conclusions about the potential for nonresponse bias in estimates of change in sales and change in inventories. What we do know is that the relative change estimator is approximately unbiased if the population level and the bias on the level change at the same rate from month-to-month. Additionally, if sales and inventories estimates are unbiased, the change estimates are also unbiased. We are investigating ways to evaluate the relative change estimates directly.

Future research for the MWTS will involve investigation of adjustment techniques that we can use the rich set of available covariates to adjust for nonresponse. Multiple imputation using chained equations is one approach we will investigate.

References

Andridge, R.R. and Little, R.J.A. (2011), Proxy Pattern-Mixture Analysis for Survey Nonresponse, *Journal of Official Statistics*, 27, 153-180.

Groves, R., & Brick, J. (2005), Practical Tools for Nonresponse Bias Studies, Course notes.

Federal Committee on Statistical Methodology, How to Do Nonresponse Bias Analyses in Household and Establishment Surveys, Federal Committee on Statistical Methodology Workshop held June 10, 2009, Washington, D.C.

Lineback, J.F. and Thompson, K.J. (2010), Conducting Nonresponse Bias Analysis for Business Surveys, *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

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

Sarndal, C-E, Swensson, B., Wretman, J. (2003), *Model Assisted Survey Sampling*, Springer, New York.

U.S. Census Bureau (2012), Statistical Quality Standards, www.census.gov/quality/standards/index.html.

[2] In this paper, we do not address the potential effects of benchmarking and seasonal adjustments.

Attachment A

Bias in Change and Relative Change Estimates of Levels

Let
Y_t = The population total at time *t*,
$\hat{Y}_t$ = An estimate of Y_t at time *t*, and
B_t = The bias of $\hat{Y}_t$.

Now if $\hat{Y}_t$ is a biased estimator of Y_t, then for some $B_t \neq 0$

$$E[\hat{Y}_t] = Y_t + B_t$$

<u>Bias in the Estimate of Change</u>
Let
$\Delta Y_t = Y_t - Y_{t-1}$ and
$\widehat{\Delta Y_t} = \hat{Y}_t - \hat{Y}_{t-1}$.

Then if $\hat{Y}_t$ is a biased estimator of Y_t in both time *t* and *t-1*

$$\begin{aligned} E[\widehat{\Delta Y_t}] &= E[\hat{Y}_t - \hat{Y}_{t-1}] \\ &= E[\hat{Y}_t] - E[\hat{Y}_{t-1}] \\ &= Y_t + B_t - Y_{t-1} - B_{t-1} \\ &= (Y_t - Y_{t-1}) + (B_t - B_{t-1}) \\ &= \Delta Y_t + (B_t - B_{t-1}) \end{aligned}$$

For $\widehat{\Delta Y_t}$ to be an unbiased estimator of ΔY_t the bias on the level at time *t* and the bias on the level at time *t-1* have to be equal (i.e., $B_t = B_{t-1}$). That is, the bias is independent of the magnitude of *Y*.

<u>Bias in the Estimate of Relative Change</u>
Assuming $Y_{t-1} \neq 0$,

let
$R_t = \frac{Y_t}{Y_{t-1}}$,
$\hat{R}_t = \frac{\hat{Y}_t}{\hat{Y}_{t-1}}$,
$Rel(\Delta Y_t) = \frac{Y_t - Y_{t-1}}{Y_{t-1}} = \frac{Y_t}{Y_{t-1}} - 1 = R_t - 1$, and
$Rel(\widehat{\Delta Y_t}) = \frac{\hat{Y}_t - \hat{Y}_{t-1}}{\hat{Y}_{t-1}} = \frac{\hat{Y}_t}{\hat{Y}_{t-1}} - 1 = \hat{R}_t - 1$.

For $Rel(\widehat{\Delta Y_t})$ to be an unbiased estimator of $Rel(\Delta Y_t)$, $\hat{R}_t$ has to be an unbiased estimator of R_t.

Now note that the first order Taylor Series approximation of $\hat{R}_t$ expanded about R_t is

$$\hat{R}_t \cong R_t - \left(\frac{Y_t}{Y_{t-1}^2}\right)(\hat{Y}_{t-1} - Y_{t-1}) + \frac{1}{Y_{t-1}}(\hat{Y}_t - Y_t)$$

so

$$\begin{aligned} E[\hat{R}_t] &\cong E\left[R_t - \left(\frac{Y_t}{Y_{t-1}^2}\right)(\hat{Y}_{t-1} - Y_{t-1}) + \frac{1}{Y_t - 1}(\hat{Y}_t - Y_t)\right] \\ &\cong R_t - \left(\frac{Y_t}{Y_{t-1}^2}\right)E[(\hat{Y}_{t-1} - Y_{t-1})] + \frac{1}{Y_{t-1}}E[(\hat{Y}_t - Y_t)] \\ &\cong R_t - \left(\frac{Y_t}{Y_{t-1}^2}\right)B_{t-1} + \frac{1}{Y_{t-1}}B_t \end{aligned}$$

So for $\hat{R}_t$ to be an approximately unbiased estimator of R_t,

$$\frac{1}{Y_{t-1}}B_t - \left(\frac{Y_t}{Y_{t-1}^2}\right)B_{t-1} = 0.$$

This implies that

$$R_t = \frac{B_t}{B_{t-1}}.$$

In other words, the bias and population levels must change at the same rate from time *t-1* to *t* for the $Rel(\widehat{\Delta Y_t})$ to be an approximately unbiased estimator of $Rel(\Delta Y_t)$. For both $\widehat{\Delta Y_t}$ and $Rel(\widehat{\Delta Y_t})$ to be unbiased, $B_t = B_{t-1} = 0$ which implies that $\hat{Y}_t$ must be an unbiased estimate of the population total.