A More Unified Statistical Approach to Nonresponse and Weighting Adjustments

Ismael Flores Cervantes [1]
[1]Westat, 1600 Research Blvd, Rockville, MD 20850

Abstract
As nonresponse continues to increase, weighting adjustments increasingly rely on implicit (or explicit) mathematical models for explaining nonresponse. Although it is crucial to understand this relationship and its impact on survey estimates, the literature that describes the different weighting adjustments is dispersed and sometimes contradictory. In this research, we explore a more unified approach for understanding these relationships. We begin by proposing an expression for nonresponse bias for estimates computed using weights that incorporate nonresponse adjustments. We explain why this revised expression may be preferred for examining adjusted survey statistics under the Total Survey Error conceptual framework. Starting with this expression, we determine the common characteristics and relationships among the different types of weighting adjustments. We argue that weighting for nonresponse should be seen as an estimation task, and once the statistical models have been identified, classical statistical tools such as goodness of fit and model diagnostics can be used to evaluate the quality of nonresponse adjusted weights. This approach may enable us to evaluate the effect of model misspecification from incorrect functional forms or omitted variables in the models on survey estimates.

Key Words: Nonresponse, nonresponse weighting adjustment, Total Survey Error, response statistical model, model misspecification, weighting

1. Bias in Unadjusted Estimates of Ratio Means

We begin with the expression for the nonresponse bias in estimates of unadjusted ratio means in the presence of nonresponse under the stochastic model described by Oh & Scheuren (1983) and Lessler & Kalsbeek (1992). The expression of the relative bias of an unadjusted ratio mean is

$$\mathrm{RB}\left(\hat{\bar{y}}_{un}\right)=\frac{\mathrm{B}\left(\hat{\bar{y}}_{un}\right)}{\bar{Y}}\approx\frac{\rho_{\phi y}\sigma_{\phi}\sigma_{y}}{\bar{\phi}\bar{Y}}=\rho_{\phi y}\,\mathrm{CV}(\phi)\mathrm{CV}(Y), \tag{1}$$

where $\bar{Y}$ is the population mean of the variable of interest (i.e., $\{y_1,..,y_N\}$ such as $\bar{Y}=\sum_{i\in U} y_i/N$), σ_y is the population standard deviation of Y, ϕ_i is the propensity to respond for the *i*-th member in the frame (i.e., $\{\phi_1,..,\phi_N\}$), $\bar{\phi}=\sum_{i\in N}\phi_i/N$ is the population mean of response propensities, σ_ϕ is the population standard deviation of ϕ; and $\rho_{\phi y}$ is the correlation between ϕ and Y.

Equation (1) is based on a stochastic response model that assumes each element in the frame has a positive propensity to respond, i.e., $\phi_i > 0$ (Bethlehem, 1988). In other words, each person is potentially a respondent or nonrespondent and the decision to participate is the realization of a stochastic process. Under this view, response is a random variable with an assumed probability mass function. Although the specific outcome cannot be predicted with certainty, its behavior can be observed through repeated sampling. As any statistical model, the values of the parameters (in this case, the response propensities ϕ_is) are not known but can be estimated for the observed data. Bethlehem imposes the conditions in the model that $\phi_i > 0$ for all members in the frame; however, in practice, there will be a fraction of hard core nonrespondents who never answer the survey (Kott, 1994).

The expression (1) has been cited in numerous articles when describing the impact of nonresponse on the quality of survey estimates (see Brick, 2013; Groves, 2006; Brick & Montaquila, 2009; and Kreuter & Olson, 2011). Groves (2006) presents a simplified version of this formula in describing nonresponse bias in household surveys. This expression is also presented when reporting response rates in surveys (California Health Interview Survey, 2011). This expression is used for explaining the bias of estimates due to nonresponse, one of non-observational errors within the survey cycle within the Total Survey Error (TSE) framework (Groves, et al., 2009).

The stochastic expression (1) is useful for understanding when nonresponse becomes an important source of bias in estimates of unadjusted means. Most of the literature focuses on the relationship between the magnitude of the bias and the correlation between the response propensity and the dependent variable. This relationship or association is represented by the correlation $\rho_{\phi y}$ between y, the variable of interest in the survey, and ϕ, the propensity to respond. The correlation $\rho_{\phi y}$ measures the extent to which the two variables fluctuate together. Larger absolute values of $\rho_{\phi y}$ result in larger bias in the unadjusted estimates of the mean of Y. In the extreme case, when the correlation of the response propensity and the variable of interest is zero ($\rho_{\phi y} = 0$), the unadjusted mean is unbiased. Although unlikely, it is possible to compute unbiased estimates of means without any nonresponse adjustment in this situation.

There are other factors that affect the size of the bias, which are not generally cited in the literature but can be inferred from expression (1). For example, for a fixed correlation, the bias also depends on the coefficient of variation of the response propensities, or $\mathrm{CV}(\phi) = \sigma_\phi / \bar{\phi}$. The bias can be large if the response propensities are highly variable even for small values of $\rho_{\phi y}$. This condition may occur in some adaptive designs where variability of the response propensities is sometimes induced on purpose by selecting specific subsets of nonrespondents to follow.

One drawback of equation (1) is that it only describes the bias for unadjusted means and is not useful to examine the bias for other estimators such as totals. Totals and domain totals are also important statistics as many surveys are conducted to determine the sizes of special subpopulations or domains. The relative bias for unadjusted estimates totals is

$$\mathrm{RB}\left(\hat{y}_{un}\right)=\frac{\mathrm{B}\left(\hat{y}_{un}\right)}{Y}=\bar{\phi}\,\rho_{\phi y}\mathrm{CV}\left(\phi\right)\mathrm{CV}\left(Y\right)+\bar{\phi}-1.$$

Unlike equation (1), the expression of the relative bias for totals is exact. Unadjusted estimates of totals are always biased including the case when the unadjusted mean is unbiased.

The simplicity of the stochastic model in equations (1) and (2) is a contributing factor to its extensive use for modeling response. Estimating the response propensities based on this model has motivated the development of numerous methods for adjusting sampling weights for nonresponse (Brick & Kalton, 1996; Da Silva & Opsomer, 2004; Little, 1986; Schouten & de Nooij, 2005; and Buskirk & Kolenikov, 2015). However, some basic assumptions behind the model may no longer hold in practice as new data collection modes and innovative sample designs are implemented. We describe the impact of some of these developments on the assumptions of the model and bias of unadjusted estimates.

- The value of the response propensities depends on the data collection mode. For example, the propensity of a unit *i* to respond on a telephone survey is ϕ_k' while for a mail survey it is ϕ_k'', and $\phi_k' \neq \phi_k''$. In some multimode surveys, where different modes are used in different parts of the sample, respondents may have different propensities increasing the variability of ϕ or $\mathrm{CV}\left(\phi\right)$. This variability also increases when the same potential respondents are presented different modes in sequential data collection designs. On the other hand, the variability in the overall propensities can be reduced in some adaptive designs if high propensity modes are used in low propensity cases and low propensity modes with high propensity cases to reduce variability in the overall propensities (Schouten, Calinescu, & Luiten, 2013).

- The value of the response propensities may change within the same mode during data collection. Survey practitioners have some influence on the response propensity values, by varying the contact procedures, offering incentives, or shortening the questionnaire. Although the influence on nonresponse in general is limited (for example, procedures such as incentives increase the likelihood to respond by some percentages points), these procedures may increase the variability to response propensities if they are only applied to parts of the sample. On the other hand, in some adaptive designs, incentives are only offered to low propensity cases in an effort to equalize the propensities reducing the variability.

- The value of response propensities may change during data collection. In the stochastic model, response is the result of a trial or stochastic realization based on a model with a parameter, or propensity, ϕ_i. However, in many surveys, a second contact attempts is made to elicit response (i.e., a second callback, subsequent mail-out, etc.). Based on the model, the observed respondent is the result of a second stochastic trial. In this case, the overall response probability to respond in two attempts when the sampled case refuses at the first attempt is $\phi_i^* = \phi_i + (1-\phi_i)\,\phi_i$ assuming that the response propensities remain unchanged after the first contact for all units in the population. In this case, the overall response rate is 1.0 if an infinite number of attempts are made. This is proven by noticing that the limit of the sum of the sequence of the number of completes as the number of contact attempts increases is

$$Completed\ sample = n\phi\ + n(1-\phi)\phi\ + n(1-\phi)^2\phi + ... = n\phi\sum_{k=1}^{\infty}(1-\phi)^{k-1} = n\phi\frac{1}{\phi} = n\ .(2)$$

This result provides a justification for making repeated attempts to increase both the response rate and the number of completed interviews. However, only a fixed number of contact attempts are made in practice. The main reason is that the response propensities are likely to change or decay after each attempt, including taking values of 0 after a few attempts (i.e., some will never respond to the survey). In other words, the conditional response propensities are not constant across multiple contact attempts, a fact which is not reflected in the stochastic model. As an example of the former situation in an adaptive design where a subsample of nonrespondents is contacted for a second attempt, the propensity to respond is $\phi_i^* = \phi_i + (1-\phi_i)\phi_i = 2\phi_i(1-\phi_i)$ while the response propensity remains ϕ_i for the remainder of the sample. On the other hand, if the response propensity changes after the first attempt, then the response propensity after the second attempt is $\phi_i^* = \phi_{1i} + (1-\phi_{1i})\phi_{2i}$ where ϕ_{1i} and ϕ_{2i} are the response propensities in the first and second attempts respectively. In both situations, the data collection procedures and the sample design increase the variability of ϕ and $\mathrm{CV}(\phi)$. If the variable of interest is correlated to those cases that received additional contact attempts, then the estimates may be biased if these cases are not weighted properly.

Similar issues to those discussed above are raised by Brick (2013), who advocates the inclusion of the relevant data collection activities in the definition of propensities when the data are weighted. Unbiased estimates of the response propensities can be used to adjust the sample for nonresponse with this information. Despite these observations, the stochastic model is still valid but could be adapted to reflect the impact on the response propensities from the new contact protocols and innovative survey designs.

2. Total Survey Error and Nonresponse Adjusted Estimators

Within the Total Survey Error (TSE) context, where all sources of errors in the survey cycle and the effect of their accumulation on estimates are studied, a more suitable mathematical expression for the bias in estimates should include the effect of nonresponse adjustments made to the weights after the data have been collected. After all, most analyses are conducted using nonresponse-adjusted weights. This adjustment is made during weighting and is the last opportunity that survey statisticians have to implement procedures that mitigate the effect of both nonresponse error and other errors accumulated during the life cycle of the survey. Examining response rates or studying expressions of bias of unadjusted estimates such as those presented in the previous section only inform about the quality of the data originally collected. However, these expressions do not provide any insights about the quality of the adjusted estimates, adjustment factors, or what elements have a large influence on the bias and errors of adjusted estimates.

For example, it is possible for a survey to achieve a high overall response rate with a positive correlation between the response propensity and the variable of interest, but with very small achieved sample sizes in some domains due to low propensities in those domains. In this example, equations (1) and (2) do not reflect the very large adjustments needed to reduce the bias.

Groves, et al. (2009) identifies the *adjustment error* or the error associated with the adjusted mean within the TSE framework. The adjustment error is the bias of the adjustment mean computed as

$$B\left(\bar{y}_{rw}\right) = \bar{y}_{rw} - \bar{Y}\,, \tag{3}$$

where $\bar{y}_{rw}$ is the adjusted sample mean computed as $\bar{y}_{rw} = \sum_{i \in r} w_i^* y_i \Big/ \sum_{i \in r} w_i^*$ and w_i^* is the nonresponse-adjusted weight. Although TSE provides a name and an expression for the error for adjusted means, equation (4) is not informative because it does not give us any insights about the bias in adjusted estimates in the same way as equations (1) and (2) do for unadjusted estimates.

The need for a more informative expression for studying the adjustment error is not new. Brick & Jones (2008) have proposed expressions of the bias for several estimators adjusted for nonresponse using poststratification. However, in this article we take a more general approach and examine the adjustment error for any type of adjustments.

3. Need for a More Unified Approach to Nonresponse Adjustments

One goal of this article is to explore a more unified framework for the study of nonresponse. Ideally, this framework could be used to explain the conditions where the different methods used for adjusting for nonresponse are successful at reducing bias. As mentioned before, many methods for adjusting for nonresponse have been proposed in the literature, and although these methods share the same objective of removing the bias in the estimates, there are important differences in the assumptions behind them. Furthermore, there is no clear consensus in the literature on how to choose the best method for a specific survey.

Many researchers have empirically studied different nonresponse adjustment methods through simulation studies (Iachan, Lee, & Peters, 2014; Kreuter & Olson, 2011; Little & Vartivarian, 2003; Sukasih, Jang, Vartivarian, Cohen, & Zhang, 2009; and West, 2009). However, simulation studies of artificial populations are limited because their results depend on the simulated data, making it difficult to generalize their findings. Another type of empirical study used to evaluate nonresponse involves comparisons of estimates from the same survey produced by different nonresponse adjustment methods (Rizzo, Kalton, & Brick, 1996; and Yang & Wang, 2008). One disadvantage of this approach is that the theoretical value or “gold standard” is not known. As a result, these studies are limited to reporting statistics related to the differences of estimates produced by the different methods, which cannot be generalized to other surveys.

Since most of the research on nonresponse adjustments has been focused on simulation studies or empirical comparisons, there are still several open questions and issues that could be better addressed with more appropriate statistical tools. We illustrate these issues with two examples. First, some researchers advocate the use of unweighted response rates for adjusting for nonresponse using weighting classes (Little & Vartivarian, 2003). With the appropriate tools, we could determine the conditions when this approach is best. This justification would be based on examining the characteristics of an appropriate response model instead of generalizing results from limited simulations based on specific conditions. Furthermore, several articles compare weighting classes to methods that compute response propensities, either with or without sampling weights, without describing the rationale and assumed response model that justifies either type of analysis (Grau, Potter, Williams, & Diaz-Tena, 2003; Ekholm & Laaksonen, 1991; and Lohr, Hsu, & Montaquila, 2015). Either weighted or unweighted adjustments need to be justified by an appropriate theoretical framework.

As a second example, some of the methods used for nonresponse adjustments such as weighting class nonresponse weighting seem to be ad hoc procedures. In most cases, the classes are formed by simply tabulating response rates for the available auxiliary variables (Valliant, Dever, & Kreuter, 2013). In this situation, there is no guarantee that the response propensities will be homogeneous within each weighting class as postulated by the response model unless the goodness of fit of the model is examined. In other words, there is need for diagnostic tools to determine whether the model holds. As shown by Flores Cervantes & Brick (2016), when the response model is misspecified, the weighting class adjustments produce biased estimates. Checking and evaluating models should be part of any statistical analysis, including those analyses based on survey data.

There are other more sophisticated methods for adjusting for nonresponse currently being used in practice (e.g., stratification of propensity scores and classification algorithms); however, there are some basic issues, such as whether the assumptions behind these methods are appropriate that need to be justified within the survey sampling context (Brick & Montaquila, 2009).

In summary, there is no methodology in the literature that justifies some of the current approaches or provides guidance on how they can be evaluated beyond empirical comparisons or simulation studies. However, the existence of numerous articles on evaluating and comparing nonresponse adjustments indicates that there is a strong need for a more formalized approach within the survey sampling context.

The development of such framework is described in the following sections. Nevertheless, most of the discussion is heuristic and based on a close examination of a proposed expression for the bias in nonresponse adjusted estimates. These observations will be verified and formalized in future articles, including the practical applications of some of the presented ideas.

4. Exploring a Parametric Approach to Nonresponse Adjustment

We begin our discussion by differentiating two different types of random processes. The first type is the result of unit randomization according to a sample design. The random outcomes are the results of how a sample is selected under such design. This “physical” randomization or randomization based on permutations is the cornerstone of sampling theory (Cochran, 1977). The second type of random variable is the variation result of stochastic or probabilistic process. These random outcomes are assumed to have a known probability distribution or pattern that can be analyzed statistically. This is the foundation of classical statistics. In the development of parametric approach to nonresponse, we explicitly incorporate these two processes in both derivation and evaluation of the nonresponse-adjusted estimates.

4.1 Data Generation Under Nonresponse

We begin by describing the data generation process accounting for nonresponse. Let $\mathbf{U}$ be a finite population or finite collection of N elements with labels $\{u_1,...,u_k,...,u_N\}$ where $N \in \mathbb{N}^+$ is finite and known. For simplicity, each element in U is represented by its label $U = \{1,...,k,...,N\}$. Let $\mathbf{y} \in \mathbb{R}^N$ the vector of the unknown variable y defined as $\mathbf{y} = (y_1,..., y_N)^t$, such that y_i is the value y of *i*-th element of population U. In the same way, define $\mathbf{X} \in \mathbb{R}^{N\times p}$ as a matrix of auxiliary variables defined for each element in the frame.

Consider the set of events for the outcome when an element $k \in U$ will respond or not, then R_k is a discrete random variable result of a realization of a stochastic process with an underlying model ζ that maps these events as

$$R_k = \begin{cases} 1 & \text{if unit } k \text{ responds} \\ 0 & \text{if unit } k \text{ does not respond} \end{cases}.$$

The probability mass function of the discrete random variable R_k defined by model ζ is

$$p_{R_k}(r_k) = p(R_k = r_k) = \begin{cases} \phi_k & \text{if } r_k = 1 \\ 1-\phi_k & \text{if } r_k = 0 \end{cases},$$

where $\phi_k \in \mathbb{R}$ is the probability (or propensity) to respond such as $0 < \phi_k \leq 1 \ \forall k \in U$. Let $\boldsymbol{\phi} \in \mathbb{R}^N$ be the vector with the response propensities for each element in the frame. Since we are focusing on a parametric approach, ϕ_k is the parameter of an assumed (or known) probability mass function of R_k. As in any other parametric statistical analysis, the parameter ϕ_k is unknown but can be estimated using the observed data by $\hat{\phi}_k$. It is desirable to compute the estimate $\hat{\phi}_k$ such as $E_\xi(\hat{\phi}_k) = \phi_k$ where the operator E_ξ is the expected value under the model ξ.

Let $p(s)$ be a sample design where the probability that element k is selected in the sample s for such design is defined by π_i and $1-\pi_i$ otherwise such as $0 < \pi_k \leq 1$ $\forall k \in U$. We draw a sample or subset $s \subset U$ according to the sample design $p(s)$. The random variable I_k is the indicator that maps the event for whether the unit *k* being in the sample is in the sample or not, as

$$I_k = \begin{cases} 1 & \text{if } k \in S = s \\ 0 & \text{if } k \notin S = s \end{cases}. \tag{4}$$

The probability mass function I_k when the case where the element k is in the selected sample for all possible samples is $p(I_k = 1) = \Pr(k \in S = s) = \sum_{s \in S_k} p_S(s)$ and $p(I_k = 0) = 1 - p(I_k = 0)$ otherwise or

$$p_{I_k}(i_k) = p(I_k = i_k) = \begin{cases} \pi_k & \text{if } i_k = 1 \\ 1 - \pi_k & \text{if } i_k = 0 \end{cases} \tag{5}$$

Once the sample is drawn, we only are able to observe data $\mathbf{y}_r$ on the subset $r \subset s$ (i.e., respondents).

4.2. Estimation in the Presence of Nonresponse

Define Y as the total of the population U as the $Y = \sum_U y_i$. In the absence of nonresponse, an estimate of the total based on the sample s drawn under the design $p(s)$ defined above is

$$\hat{Y}_\pi = \sum_s \frac{y_i}{\pi_i} = \sum_U \frac{y_i I_i}{\pi_i} = \sum_U d_i y_i I_i \ , \tag{6}$$

where $d_i = \pi_i^{-1}$ is the sample design weight. This estimator is known as the π -weighted total of y. In this case, the estimation and inference goes from the sample to the population.

In the presence of nonresponse, the unadjusted estimate of the total can be written as

$$\hat{Y}_u = \sum_{k \in r} d_k y_k = \sum_U d_k y_k R_k I_k \ . \tag{7}$$

This estimate is biased under repeated sampling because $\mathrm{E}(\hat{Y}_u) = \sum_U y_k \phi_k < Y$. The expected value can be seen as a weighted total of y where ϕ_k is the weight (i.e., ϕ_k -weighted total). A natural way to remove the bias in the estimate of the total is weighting by the inverse of the propensities ϕ_k as

$$\hat{Y} = \sum_{k \in r} \frac{d_k}{\phi_k} y_k \ . \tag{8}$$

However, as mentioned before, the propensities $\phi_k s$ are unknown and need to be estimated from the observed distribution of the observed $R_k s$.

4.2. Estimation of the Response Propensities

Suppose that there is no sample selection and the complete population is observed. In this situation, the only source of random variation is from the response values $\mathbf{R}_U$, which are stochastic realizations under the model ξ . Any appropriate classical statistical method can be used to not only estimate ϕ_k by $\hat{\phi}_k$ but also to make statistical inferences and evaluate the goodness of fit of the model of ϕ_k.

The estimation of the response propensity is more complex when a sample is drawn under sample design $p(s)$. We not only observe a smaller subset of responses R_k, $\mathbf{R}_s$, but the collected data are correlated. This correlation is the result from the way the sample is drawn under the sample design. In this case, the estimation process needs to reflect both random processes (i.e., randomization by design and the stochastic process of responding). To facilitate the description of the estimation steps, we use a two-step process. We first condition on a single realization of stochastic process described by model ξ for ϕ_k for all elements in the frame. In other words, we observe realized values $R_k = r_k, \forall k \in U$ (i.e., fixed values when conditioned on ξ). Any appropriate classical statistical method can be used to compute $\hat{\phi}_k, \forall k \in U$. In this case, $\hat{\phi}_k$ is fixed and is not different from any other variable y_k in the population. If we draw repeated samples under the same design $p(s)$, then each sample can be used to estimate these fixed model-based propensities as with any other variable of interest under repeated sampling. In this case, the conditional expectation of the estimated propensity $\hat{\phi}_k$ is $\mathrm{E}_p\left(\hat{\phi}_k\right) = \hat{\phi}_k$ under the sample design $p(s)$. In the second step, we remove the conditionality of the response propensities on ξ. Any stochastic realization can be used to estimate ϕ_k. In other words, the unconditional expectation is $\mathrm{E}_\xi \, \mathrm{E}_p\left(\hat{\phi}_k\right) = \phi_k$ [1]. Fortunately, there are several statistical methods available for computing estimates of parameters of distributions such as $\hat{\phi}_k$ under both a complex sample design and a stochastic model. Two of these approaches are described in Binder (1983) and Chambers, Steel, Wang, & Welsh (2012).

4.3. Estimating Totals

Once the estimates of the response propensities $\hat{\phi}_k$ are computed, we can modify the expression (9) to estimate the total in the population. The nonresponse adjusted estimator of the total based on estimated response propensities is

$$\hat{Y}_a = \sum_{k \in r} \frac{1}{\hat{\phi}_k} d_k y_k = \sum_U \frac{1}{\hat{\phi}_k} y_k d_k R_k I_k \,. \tag{9}$$

Under repeated sampling, the expected value of the estimate of the total is

$$\mathrm{E}_p\left(\hat{Y}_a\right) = \sum_U \frac{R_k}{\hat{\phi}_k} y_k = \sum_U \hat{\tau}_k y_k \tag{10}$$

[1] One consequence of this approach is that the variance of the estimates includes two terms to account for the two sources of variation as

$$\mathrm{V}\left(\hat{\phi}_k\right) = \mathrm{V}_\xi\left[\mathrm{E}_p\left(\hat{\phi}_k\right)\right] + \mathrm{E}_\xi\left[V_p\left(\hat{\phi}_k\right)\right].$$

where $\hat{\tau}_k = \frac{R_k}{\hat{\phi}_k}$ is the ratio of two random variables, R_k, the result of the stochastic process described by model ξ with a distribution with parameter ϕ_k as described above, and $\hat{\phi}_k$, the result of estimating the parameter of the distribution of R_k. Equation (11) is the design based expected value of the nonresponse adjusted total or the $\hat{\tau}_k$-weighted total of y.

4.4. Bias of Nonresponse Adjusted Estimators

We begin by deriving the expression similar to equation (2) but for the nonresponse adjusted estimator of the total. The relative bias of the nonresponse adjusted total under repeated sampling is defined as

$$\mathrm{RB}_p\left(\hat{Y}_a\right) = \frac{\mathrm{E}_p\left(\hat{Y}_a\right) - \hat{Y}}{Y}. \tag{11}$$

Substituting equation (11) and after some algebraic manipulation, it is easy to show that the relative bias of the nonresponse-adjusted total under repeated sampling is

$$\mathrm{RB}_p\left(\hat{Y}_{\hat{\phi}}\right) = \bar{\hat{\tau}}\, \rho_{\hat{\tau}Y} \mathrm{CV}\left(\hat{\tau}\right) \mathrm{CV}\left(Y\right) + \bar{\hat{\tau}} - 1 \tag{12}$$

In a similar way, the relative bias of the nonresponse-adjusted mean of y is approximately

$$\mathrm{RB}_p\left(\hat{\bar{Y}}_{\hat{\phi}}\right) \approx \rho_{\hat{\tau}Y} \mathrm{CV}\left(\hat{\tau}\right) \mathrm{CV}\left(Y\right), \tag{13}$$

where the ratio $\hat{\tau}_k$ is defined as $\hat{\tau}_k = R_k / \hat{\phi}_k$. Both expressions are useful for studying the effect of the nonresponse adjustment on the relative bias of adjusted estimates of the total and means.

When we compare equations (2) and (13) we notice that equation (13) is similar to the expression of the relative bias of unadjusted estimate of total of a population except that the "response propensity" is defined as $\hat{\tau}_k$.[2] However, unlike the response propensities ϕ_k in (2), which depends on the population/survey, the value of the ratio $\hat{\tau}_k$ depends on the method that the statistician uses to estimate ϕ_k. Furthermore, unlike the size of the bias described in equation (1), which is very difficult to influence because it requires changes in the data collection protocol to affect the values of the propensities ϕ_k, it is very easy to change the size of the bias in (13) and (14) by substituting any value of $\hat{\tau}_k$ in

---

[2] As a side note, this analogy of regarding the $\hat{\tau}_k$ ratios as response propensities is not entirely correct because the $\hat{\tau}_k$ ratios can be greater than 0. However, this analogy helps the reader when examining when the conditions when the adjusted estimators are biased. Values greater than 1 correspond to adjustments for undercoverage.

the equation. Not all $\hat{\tau}_k$ values will completely remove the bias of the total; however, some will have better properties than others.

There are different methods to compute $\hat{\phi}_k$ and these estimates will have a different impact on the $\hat{\tau}_k$ ratios and ultimately on the relative bias of the estimates of the total of y. Those methods have different assumptions for stochastic response models and the parameters of the distribution of the model. In the following paragraph, we list some observations about expression (13) without focusing on any particular method. Instead, we describe what we expect the method to accomplish to reduce the bias. This discussion is similar to identifying the conditions when the nonresponse unadjusted estimates are unbiased, as presented in Section 1. Although this discussion may only be for purely academic interest because we are commenting on unobservable parameters, it provides different insights about the conditions when the bias is reduced.

- The first condition when the estimates is unbiased is when $\hat{\phi}_k = \phi_k$ so that $E\left(\hat{\tau}_k\right) = 1$. In other words, the propensities are estimated without error. This condition highlights the difficulty of adjusting for nonresponse. Unlike other types of estimation where only one parameter is estimated, we need to estimate correctly ϕ_k for <u>all elements in the sample</u>. In practice, it is more likely that propensities ϕ_k are estimated correctly for some domains and not for others. As a result, the estimates will be biased for the domains where $E\left(\hat{\phi}_k\right) \neq \phi_k$.

When the ϕ_k are estimated without error, the response mechanism can be modeled as Poisson sampling (PO) with known probabilities of selection $\pi_k = \phi_k$ (Särndal, Swensson, & Wretman, 1992). Under this design, the estimate of the total is $t_{PO} = \sum_n y_k / \pi_k$ with a variance given by

$$\mathrm{V}_{PO}\left(t\right) = \sum_U \left(\frac{1}{\pi_k} - 1\right) y_k^2 \ . \tag{14}$$

Särndal, Swensson, & Wretman (1992) warns that the variance of this estimator may be unduly large because of variability in the sample size. However, the variance also increases if selection probabilities(π_k s) vary and if they are close to 0. The variability of the estimate is not the result of the nonresponse adjustment, but the result of the variability of the propensities. Even if these were known, the nonresponse adjusted estimate may have a large variance. In other words, the variability of the estimates does not arise from computing weights and weighting the data.

The variance of the estimate under PO sampling in (15) is minimized if $\pi_k \propto y_k$. Translating these observations to the response propensities, we expect a reduction in bias and variance when the estimated propensities are highly correlated with y_k as reported in the literature (see Little & Vartivarian, 2005).

- In the second situation we consider, $\hat{\phi}_k$ is an estimate such that $\mathrm{E}\left(\hat{\phi}_k\right)=\phi_k$ with an associated error, or $\mathrm{V}\left(\hat{\phi}_k\right)=\sigma_k^2$. When this is the case, the expected value of $\hat{\tau}_k$ is biased since

$$\mathrm{E}\left(\hat{\tau}_k\right)=\mathrm{E}_\xi\left(\frac{R_k}{\hat{\phi}_k}\right)\neq\frac{\mathrm{E}\left(R_k\right)}{\mathrm{E}_\xi\left(\hat{\phi}_k\right)}=1\,. \tag{15}$$

Using the Taylor expansion approximation, we can approximate the expected value as

$$\mathrm{E}\left(\hat{\tau}_k\right)=\mathrm{E}_\xi\left(\frac{R_k}{\hat{\phi}_k}\right)\approx\frac{\mathrm{E}_\xi\left(R_k\right)}{\mathrm{E}_\xi\left(\hat{\phi}_k\right)}-\frac{\mathrm{COV}_\xi\left(R_k.\hat{\phi}_k\right)}{\left[\mathrm{E}_\xi\left(\hat{\phi}_k\right)\right]^2}+\frac{\mathrm{E}_\xi\left(R_k\right)\mathrm{V}_\xi\left(\hat{\phi}_k\right)}{\left[\mathrm{E}_\xi\left(\hat{\phi}_k\right)\right]^3}=1-\frac{\mathrm{COV}_\xi\left(R_k.\hat{\phi}_k\right)}{\phi_k^2}+\frac{\mathrm{V}_\xi\left(\hat{\phi}_k\right)}{\phi_k^2}\,. \tag{16}$$

We expect that the last two terms in (17) will become smaller as the sample size increases when computing $\hat{\phi}_k$. As a result, some of the methods compute adjustments only with some minimum number of respondents. A more detailed analysis is beyond the observations made in this paper; however, equation (17) is a starting point for studying the effect of small sample sizes on the bias of the adjusted estimate.

- As in any modeling exercise, the response model can be misspecified, that is $\mathrm{E}\left(\hat{\phi}_k\right)=\phi_k+b_k$ where $b_k\neq 0$ or where the different models have different fits when $\hat{\phi}_{1k}$ and $\hat{\phi}_{2k}$ are estimated using two different models but where $\mathrm{V}\left(\hat{\phi}_{1k}\right)>\mathrm{V}\left(\hat{\phi}_{2k}\right)$. In these cases, $\mathrm{E}\left(\hat{\tau}_k\right)\neq 1$ and the bias becomes smaller as the sample size increases. Flores Cervantes & Brick (2016) illustrated the bias in stratified designs with misspecified models.

- Equation (13) highlights the “Achilles’ heel” of this approach for adjusting for nonresponse. Since in the ratio $\hat{\tau}_k$ where the estimates $\hat{\phi}_k$ can take values close to zero including zero, the $\hat{\tau}_k$ ratio becomes very unstable. This and the variability of $\hat{\phi}_k$ have a large impact on $\mathrm{CV}\left(\hat{\tau}\right)$. A possible approach is to restrict the value of the ratios for the domains where the ratios are unstable while maintaining the remaining adjustments as estimated in domains where the model is good. This approach will not produce biased estimates for all those domains where the ratio is stable. Some of the methods used to adjust for nonresponse do not use this approach.

The discussion so far has been on the conditions where the adjusted estimates are unbiased. We expect that a similar expression to (13) can be derived for the variance of adjusted estimates. Such expression may be useful to study the “estimated weights” paradox; that is, using estimated weights rather than known weights reduces variance (Kim & Kim, 2007 and Lumley, Shaw, & Dai, 2011).

4.5. Auxiliary Variables

In the previous paragraphs we examine the bias in the estimates in general. In practice, we rely on auxiliary variables for computing $\hat{\phi}_k s$. In other words, we build a model for ϕ_k based on auxiliary variables, $\mathbf{x}_k$ which should be available for respondents and nonrespondents. Before exploring the role of the auxiliary in the estimation of ϕ_k, we examine another approximation for the bias of the estimates. Assuming that $\mathrm{E}\left(\bar{\tau}\right) \approx 1$ and $\mathrm{CV}\left(\hat{\tau}\right) \approx 1$, and if we focus on the numerator term of the correlation between $\hat{\tau}_k$ and y_k, an approximation of the bias of the estimates is

$$\mathrm{RB}_p\left(\hat{\bar{Y}}_{\hat{\phi}}\right) \propto \sum_U \left(\hat{\tau}_k - 1\right)\left(y_k - \bar{y}\right). \tag{17}$$

Expression (18) is similar to the expression of the near bias in Särndal & Lundström (2005); although the expression of near bias is a function of response propensities and not the $\hat{\tau}_k$ ratios.[3] Similar conclusions for unbiased estimates presented in the previous section can be drawn examining equation (18).

For the analysis of auxiliary variables and as a way to simplify the analysis, we assume that the auxiliary variables $\mathbf{x}_k$ are uncorrelated[4], that is $\mathrm{Cor}\left(x_{ki}, x_{kj}\right) = 0 \quad \forall i \neq j$ and $i, j \in \{1,.., p\}$. Suppose that there is a model that generates the population response propensities such as $\phi_k = \mathrm{f}\left(\boldsymbol{\lambda}\mathbf{x}_k^t\right)$ where $0 < \mathrm{f}\left(\boldsymbol{\lambda}\mathbf{x}_k^t\right) \leq 1$. Based the model presented before, we focus only on two forms of model misspecification: incorrect functional form and omitted variable[5] (Asteriou & Hall, 2011).

In the first form of misspecification, the functional form of the model generating ϕ_k is not identified correctly. That is when $\phi_k = \mathrm{f}\left(\boldsymbol{\lambda}\mathbf{x}_k^t\right)$ is not approximated well by $\hat{\phi}_k = \mathrm{g}\left(\hat{\boldsymbol{\lambda}}\mathbf{x}_k^t\right)$, that is $\hat{\tau}_k = \dfrac{\mathrm{f}\left(\boldsymbol{\lambda}\mathbf{x}_k^t\right)}{\mathrm{g}\left(\hat{\boldsymbol{\lambda}}\mathbf{x}_k^t\right)} \neq 1$. For example when $\mathrm{f}\left(\boldsymbol{\lambda}\mathbf{x}_k^t\right) = \boldsymbol{\lambda}\mathbf{x}_k^t$ and $\hat{\phi}_k = \mathrm{expit}\left(\hat{\boldsymbol{\lambda}}\mathbf{x}_k^t\right)$. That is, the propensities are linear $\mathbf{x}_k^t$ while the fitted function is a sigmoid function on $\mathbf{x}_k^t$. The bias in the estimates in this case depends on how dissimilar are the functions f and g. There are several statistical methods and techniques to determine whether the adopted functional form is a good fit for the observed data.

[3] Särndal & Lundström, (2005) mention that the near bias expression is not based on a model, unlike what is described in equation (18).

[4] In practice, this can be done by replacing the original vector of auxiliaries with a principal components or principal factors.

[5] Other forms of misspecification are overfit or the inclusion of irrelevant variables, simultaneity bias where the dependent variable is part of a system of simultaneous equations, and measurement errors in the independent variables. These misspecification forms can also be studied in these response models.

In the second type of misspecification, important auxiliary variables are omitted or excluded in the model. For example, suppose that estimated response propensities are estimates using the correct functional form $\hat{\phi}_k = \mathrm{f}\left(\hat{\boldsymbol{\lambda}}\mathbf{x}_k^{*t}\right)$ for $\phi_k = \mathrm{f}\left(\boldsymbol{\lambda}\mathbf{x}_k^t\right)$ but where an important variable $x_{ok} \notin \mathbf{x}_k^*$ is omitted or excluded in the model for $\hat{\phi}_k$. The estimates $\hat{\phi}_k$ will be biased if the omitted variable x_{ok} is a determinant of the response propensity ϕ_k (i.e., $\left|\lambda_{ok}\right| \square\ \lambda_{jk}$) and correlated with the response propensity (i.e., $\mathrm{COV}\left(x_{ok}, \phi_k\right) \neq 0$). Omitted variable bias is one of the most common problems in linear regression and current methods for dealing with it may not be readily applicable to binary regression used for response modeling. Additional theoretical development may be needed to handle this situation. In the worst case scenario, there may be some variables that should be in the model but are not found in the set of auxiliary variables. In this situation, the estimates are expected to be biased. Unfortunately, this may be the usual situation.

All these previous observations provide a good starting point for describing the relationships between response model, auxiliary variables, and dependent variables, and the conditions when the adjusted estimate is unbiased. In addition to the pedagogical value, it helps us make predictions when different nonresponse adjustments reduce the bias in adjusted estimates. This is important in simulation studies because we can predict whether estimates will be biased based on how the response models are created in the simulations. Unexpected results of simulations in published articles are incorrectly attributed to extraneous reasons, when in reality these anomalies are the result of unintended interactions among the response model, auxiliary variables, and dependent variables in the simulation.

5. Final Comments

There is a need to evaluate nonresponse adjustment methods to use beyond empirical approaches such as simulations and comparison of estimates produced using different weighting methods. We develop an expression for the relative bias for nonresponse adjusted totals and means. The expression can be used as a pedagogical tool since it provides insights on how the nonresponse adjustments work. We discuss the conditions when the bias is minimized based on relationships between the model, response, and auxiliary variables. Based on these observations, we are able to determine the desired properties of the methods for reducing bias in adjusted estimates. Starting from this analysis, we begin investigating the use of a parametric approach for nonresponse adjustments that reflect both randomization from the sample design and the stochastic process and random draws from probability distributions. This approach let us take advantage of classical statistical tools such as statistical tests for parameters and goodness of fit of the response model for determining the best way to adjust for nonresponse. The main goal is to develop a framework where nonresponse can be seen as an estimation exercise based on statistical methods such as exploratory analyses, variable selection, model evaluation, and checking assumptions.

The proposed parametric approach has been used in nonresponse adjustments based on weighting classes and the results look promising. We will continue developing this approach for more nonresponse adjustment methods. However, one of the main obstacles is the development of specialized software. Current results and further developments in addition to the formalized observations described in this article will be presented in future articles.

Acknowledgments

Special thanks to Dr. Roger Tourangeau for his comments on this work and Dr. J. Michael Brick for his support.

References

Asteriou, D., & Hall, S. G. (2011). *Applied econometrics.* New York, NY: Palgrave Macmillan.

Bethlehem, J. G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics, 4*, pp. 251-260.

Binder, D. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review, 51*(3), 279-292.

Brick, J. M. (2013). Unit nonresponse and weighting adjustments: a critical review. *Journal of Official Statistics, 29*, pp. 329-353.

Brick, J. M., & Jones, M. E. (2008). Propensity to respond and nonresponse bias. *Metron-International Journal of Statistics, LXVI*, 51-73.

Brick, J., & Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research, 5*, pp. 215-238.

Brick, J., & Montaquila, J. (2009). Nonresponse and weighting. In D. Pfeffermann, & C. Rao, *Handbook of statistics_29A: Sample surveys: Design, methods and applications* (p. 722). North Holland.

Buskirk, T. D., & Kolenikov, S. (2015). Finding respondents in the forest: a comparison of logistic regression and random forest models for response propensity weighting and stratification. *Insights: Methods from the field, weighting: Practical Issues and "How to" approach*. doi:10.13094/SMIF-2015-00003.

California Health Interview Survey. (2011). CHIS 2009: Methodology Series: Report 4 - Response Rates. Los Angeles, CA. Retrieved September 18, 2013, from http://healthpolicy. ucla.edu/chis/design/Pages/methodology.aspx.

Chambers, R. L., Steel, D. G., Wang, S., & Welsh, A. H. (2012). *Maximum likelihood estimation for sample surveys.* Boca Raton, Florida: CRC Press.

Cochran, W. G. (1977). *Sampling techniques.* (3rd ed.). New York: John Wiley & Sons.

Da Silva, D., & Opsomer, J. (2004). Properties of the weighting cell estimator under a nonparametric response mechanism. *Survey Methodology, 30*(1), 45-55.

Ekholm, A., & Laaksonen, S. (1991). Weighting via response modeling in the Finnish Household Budget Survey. *Journal of Official Statistics, 7*(3), 325-337.

Flores Cervantes, I., & Brick, J. M. (2016). Nonresponse adjustments with misspecified models in stratified designs. *Survey methodology*.

Grau, E., Potter, F., Williams, S., & Diaz-Tena, N. (2003). Nonresponse adjustment using logistic regression: to weight or not to weight? *Proceedings of the Annual Meeting of the American Statistical Association, Section on Survey Research Methods.* American Statistical Association.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 646-675.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). New York: Wiley.

Iachan, R., Lee, H. R., & Peters, K. (2014). A Comparison of weighting adjustment methods for nonresponse. *Proceedings of the Annual Meeting of the American Statistical Association, Section on Survey Research Methods.* American Statistical Association.

Kim, J. K., & Kim, J. J. (2007). Nonresponse weighting adjustment using estimated response probability. *The Canadian Journal of Statistics, 35*(1), 501-514.

Kott, P. (1994). A note on handling nonresponse in sample surveys. *Journal of the American Statistical Association, 89*(426), 693-699.

Kreuter, F., & Olson, K. (2011). Multiple auxiliary variables in nonresponse adjustment. *Sociological Methods & Research, 40*, 311-332.

Lessler, J., & Kalsbeek, W. (1992). *Nonsampling errors in surveys.* New York: John Wiley and Sons.

Little, R. J. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review, 54*(2), 139-157.

Little, R. J., & Vartivarian, S. (2003). On weighting the rates in nonresponse weights. *Statistics in Medicine, 22*, 1589-1599.

Little, R. J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology, 31*(2), 161-168.

Lohr, S., Hsu, V., & Montaquila, J. (2015). Using classification and regression trees to model survey nonresponse. *Joint Statistical Meetings, Proceedings of the Survey Research Methods.* American Statistical Association.

Lumley, T., Shaw, P. A., & Dai, J. Y. (2011). Connections between Survey Calibration Estimators and Semiparametric Models for Incomplete Data. *International Statistical Review, 79*(2), 200-220. doi:doi:10.1111/j.1751-5823.2011.00138.x.

Oh, H., & Scheuren, F. (1983). Weighting adjustments for unit nonresponse. In W. Madow, I. Olkin, & D. Rubin, *Incomplete data in sample surveys* (Vol. 2, pp. 143-184). Academic Press.

Rizzo, L., Kalton, G., & Brick, J. (1996). A comparison of some weighting adjustments for panel nonresponse. *Survey Methodology, 22*, 43-53.

Särndal, C.-E., & Lundström, S. (2005). *Estimation in surveys with nonresponse.* New York: Wiley.

Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling.* New York: Springer.

Schouten, B., & de Nooij, G. (2005). *Nonresponse adjustment using classification trees.* Statistics Netherlands, Voorburg/Heerlen.

Schouten, B., Calinescu, M., & Luiten, A. (2013). Optimizing quality of response through adaptive survey. *Survey Methodology, 39*(1), 29-58.

Sukasih, A., Jang, D., Vartivarian, S., Cohen, S., & Zhang, F. (2009). A simulation study to compare weighting methods for nonresponses in the National Survey of Recent College Graduates. *Proceedings of the Survey Research Methods Section, American Statistical Association.* Retrieved October 21, 2013, from www.amstat.org/sections/srms/proceedings/y2009/ Files/304345.pdf.

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples.* New York: Springer.

West, B. T. (2009). A simulation study of alternative weighting class adjustments for nonresponse when estimating a population mean from complex sample survey data. *Proceedings of the Survey Research Methods Section.* American Statistical Association. Retrieved 10 21, 2013, from www.amstat.org/sections/srms/proceedings/y2009/Files/305394.pdf

Yang, Y. M., & Wang, Y. (2008). Weighting class versus propensity model approaches to nonresponse adjustment: the SDR experience. *Joint Statistical Meetings, Proceedings of the Section on Survey Research Methods.* American Statistical Association.