# Calibration Weighting for Nonresponse with Proxy Frame Variables (So that Unit Nonresponse Can Be Missing Not at Random)

Phillip S. Kott
RTI International, 6110 Executive Blvd., Rockville, MD 20852

**Abstract**
When adjusting for unit nonresponse in a survey, it is common to assume that the response/nonresponse mechanism is a function of variables known either for the entire sample before unit response or at the aggregate level for the frame or population. Often, however, some of the variables governing the response/nonresponse mechanism can only be proxied by variables on the frame while they are measured (more) accurately on the survey itself. For example, an address-based sampling frame may contain area-level estimates for the median annual income and the fraction home ownership in a Census block group, while a household's income category and ownership status are reported on the survey itself for the housing units responding to the survey. A relatively new calibration-weighting technique employed by the WTADJX procedure in SUDAAN® 11 allows a statistician to calibrate the sample using proxy variables while assuming the response/ nonresponse mechanism is a function of the analogous survey variables. We will demonstrate how this works with data from the Residential Energy Consumption Survey National Pilot, a nationally representative web-and-mail survey of U.S. households sponsored by the U.S. Energy Information Administration.

**Key Words**: Model variable, Weight-adjustment function, Selection bias.

## 1. Introduction

Calibration weighting is a useful tool for treating unit nonresponse in a survey. It can implicitly estimate the probability of response given a known form the response model. Moreover, the resulting weights tend to more efficient than the weights produced using maximum-likelihood methods to estimate the response model (Kim and Riddles 2012).

Deville (2000) has shown how calibration weighting can be used to treat unit (element-level) nonresponse that can be either missing at random (MAR) or missing not at random (MNAR). The former means that nonresponse is a function entirely of variables with either known population totals of known values for the entire sample, while the latter allows nonresponse to be at least partially a function of variables known only for responding sampled elements. Unfortunately, there is no statistical way to determine whether or not nonrespondents are missing at random (Molenberghs et al. 2008). As a result many have argued that techniques like Deville's are best suited for sensitivity analyses. See, for example, National Research Council (2010).

There are some situations, however, where unit nonresponse can logically be inferred to be not missing at random. In a survey of housing units (HUs), for example, unit nonresponse may be a function of whether or not the HU is owned by the household residing in it and by the income of that household. This information can be collected on the survey itself (assuming no item nonresponse), but can only be proxied for the sample as a whole. Such proxies are useful because Deville's method requires that there be variables on which to calibrate the respondent sample so that the weighted sum of those variables among respondents equal a known population total or a weighted total computed from the full sample (including nonrespondents). A potential source for proxy variables in the US is the American Community Survey, which makes available estimates at the Census-block-group level of the average median annual income and the fraction of owned HUs.

This paper looks at data from the 2015 national pilot of the Residential Energy Consumption Survey (RECS) which was conducted my mail and web. It compares results of calibration weighting assuming nonresponse is missing at random using some proxy variables available on the frame among its arguments with results of calibration weighting where survey variables replace the proxy variables.

Section 2 will review the underlying theory of calibration weighting assuming (for simplicity) a logistic response function. Section 3 will describe the RECS National Pilot and how it is being weighted for nonresponse assuming that unit respondents are missing at random. Section 4 compare some estimates and their estimated standard errors using the National-Pilot method and then alternative methods that assume nonresponse is not missing at random. Section 5 offers some concluding remarks.

## 2. An Overview of Calibration Weighting Assuming a Logistic Response Function

To simplify matters, let us assume that there is only one type of unit nonresponse, and it takes place at the element level, denoted by the subscript $k$. Moreover, there is no coverage problems with the sampling frame nor is their any item nonresponse among element respondents.

Suppose the unit (element) response mechanism can be represented by an independent logistic function that depends on a vector of values for each element. Letting $\rho_k$ be the probability that element $k$ responds, and $\mathbf{x}_k$ the vector of (response) *model* variables governing that probability, which includes unity or the equivalent (i.e., a linear combination of the components of $\mathbf{x}_k$ is 1), we have

$$\rho_k = \rho(\mathbf{x}_k^T\boldsymbol{\gamma}) = 1/[1 + \exp(\mathbf{x}_k^T\boldsymbol{\gamma})], \qquad (1)$$

for some unknown vector $\boldsymbol{\gamma}$.

Calibration weighting begins with the calibration equation:

$$\sum_R d_k\,[1 + \exp(\mathrm{x}_k^T\mathrm{g})]\,\mathrm{z}_k = \mathrm{T}_\mathrm{z} \qquad (2)$$

where $R$ denotes the respondent sample, $d_k$ the design (initial sampling) weight of element $k$, $\mathbf{z}_k$ a vector of *calibration* variables, each having either a known population total or a total can be estimated from the full sample (including the unit nonrespondents), $\mathbf{T}_\mathbf{z}$ the vector of (estimated) totals for the components of $\mathbf{z}_k$. Finally, $\mathbf{g}$ is a consistent estimator for $\boldsymbol{\gamma}$ under mild conditions, determined by solving for it in calibration equation (2) using Newton's method (repeated linearization).

The calibration weight for element $k$ resulting from the solution of equation (2) is

$$w_k = d_k\alpha(\mathbf{x}_k{}^T\mathbf{g}) = d_k[1 + \exp(\mathbf{x}_k{}^T\mathbf{g})].$$

The expression $\alpha(\mathbf{x}_k{}^T\mathbf{g})$ is called the *weight-adjustment function* because it converts the design weight $d_k$ into the nonresponse-adjusted or calibration weight $w_k$. The estimated total of a survey variable $y$ using calibration weights is $t_y = \sum_R w_k y_k$.

In most applications, the components of calibration vector $\mathbf{z}_k$ are assumed to coincide with the components of the model vector $\mathbf{x}_k$. This means unit nonrespondents are assumed to be missing at random. When that is the case, the calibration equation in (2) will almost always have a solution so long as unit nonresponse is truly a logistic function of the components of $\mathbf{x}_k$. When the components of $\mathbf{z}_k$ and $\mathbf{x}_k$ do not coincide, the calibration equation may not have a solution, especially if a component of $\mathbf{x}_k$ is linearly independent of all the components of $\mathbf{z}_k$.

Chang and Kott (2008) generalized the notion of calibration weighting to allow more calibration variables than model variables, but Kott and Liao (2016) maintained that a prudent approach would be to include in $\mathbf{z}_k$ all the components of $\mathbf{x}_k$ for which population or full-sample estimates are known. The rest they called *shadow variables*, which they suggested should be proxies for the *model-only variables* in $\mathbf{x}_k$ that could not themselves be calibration variables in $\mathbf{z}_k$.

In our RECS National Pilot example, some variables, such as whether or not an HU $k$ is in an urban area, can be in both the model vector and the calibration vector, while other variables, such home ownership (yes or no), are model-only variables in $\mathbf{x}_k$. At the same time, a reasonable proxies for each model-only variable, like the fraction of homes owned in its Census block group, can be a shadow variables in $\mathbf{z}_k$.

When the calibration equation has a solution, it is not hard to show that an asymptotically unbiased estimator for the variance of $\mathbf{g}$ under mild conditions is

$$\mathbf{V_g} = \mathbf{F}\,\mathbf{var}\{\textstyle\sum_R d_k\,[1 + \exp(\mathbf{x}_k{}^T\boldsymbol{\gamma})]\,\mathbf{z}_k \,|\, \mathbf{T_z}\}\ \mathbf{F}^T, \tag{3}$$

where $\mathbf{F} = [\sum_R d_k \exp(\mathbf{x}_k{}^T\mathbf{g})]\ \mathbf{z}_k\mathbf{x}_k{}^T]^{-1}$ , and $\mathbf{var}\{\mathbf{q}\,|\,\mathbf{T_z}\}$ is an estimator for variance-covariance matrix for $\mathbf{q}$ as an estimator for $\mathbf{T_z}$ . To compute it, one treats $p_k = 1/[1 + \exp(\mathbf{x}_k{}^T\mathbf{g})]$ as if it equaled $\rho_k$ in equation (1).

An asymptotically unbiased estimator for the variance in $t_y = \sum_R w_k y_k$ is

$$v_y = v\{\textstyle\sum_S d_k[\mathbf{z}_k{}^T\mathbf{b} + \alpha(\mathbf{x}_k{}^T\mathbf{g})e_k]\} \tag{4}$$

where $e_k = y_k - \mathbf{z}_k{}^T\mathbf{b}$, $\mathbf{b} = [\sum_R d_j\alpha'(\mathbf{x}_j{}^T\mathbf{g})\,\mathbf{x}_j\mathbf{z}_j{}^T]^{-1}\sum_R d_j\alpha'(\mathbf{x}_j{}^T\mathbf{g})\,\mathbf{x}_j y_j$, and $\alpha(\mathbf{x}_k{}^T\mathbf{g}) = [1 + \exp(\mathbf{x}_k{}^T\mathbf{g})]$ when $k \in R$ and 0 otherwise is treated as a constant within the probability-sampling variance estimator $v\{.\}$. For the variance of $m_y = \sum_R w_k y_k / \sum_R w_k$ , replace $y_k$ by $(y_k - m_y)/ \sum_R w_j$. The reader is directed to Kott (2014) for the proofs and further details.

It is easy to see that due to calibration $\sum_R w_k y_k - \sum_S d_k y_k = \sum_R w_k e_k$ . We thus have this estimate for the increase in variance due to nonresponse:

$$\mathrm{var}\{\textstyle\sum_R d_k\,[1+\exp(\mathbf{x}_k^T\boldsymbol{\gamma})]\,y_k \,|\, \sum_S d_k y_k\,\} = \sum_R d_k^2(1/p_k^2)(1-p_k)e_k^2 \qquad (5)$$
$$= \textstyle\sum_R d_k^2\,[1+\exp(\mathbf{x}_k^T\mathbf{g})]\,\exp(\mathbf{x}_k^T\mathbf{g})e_k^2,$$

The estimate assumes the probabilities of element response are independent of each other. Again, the reader can consult Kott (2014) for proofs and details.

## 3. The RECS National Pilot

The RECS National Pilot was an attempt to convert what historically has been an in-person interview survey into one conducted by web and mail. More information on it can be found elsewhere (Berry and O'Brien 2016). For our purposes, the RECS National Pilot (hereafter the "National Pilot") used four randomly-assigned protocols and two randomly-assigned incentive levels in data collection from a stratified, two-stage sample drawn using an address-based sampling frame with mail invitation and up to six mailings.

The protocols were, 1, web only, 2, choice of web or mail, 3, choice of web or mail but with an added \$10 incentive to respond via web, and, 4, web in the first mailing followed by a choice in subsequent mailings. The two incentive levels both provided the sampled HU \$5 initially. One provided an extra \$10 upon completion while the other provided an extra \$20. There was a shortened mail follow-up survey (NRFU) for nonrespondents, but that does not concern us here – except in a design-weight adjustment to be described shortly – nor does the poststratification designed to capture HUs not on the address-based sampling frame.

Two issues with the enumerations of the National Pilot do have an impact on our analysis. Not all HUs in the sampling frame were occupied, and some were occupied but not primary residents. Only data from primary residents were to be used in making National-Pilot estimates.

A latent-variable model (Biemer et al. 2016) has been used to estimate the probability that a sampled HU was occupied based on frame characteristics, the disposition of the first three mailings, and whether they responded to the survey. Those estimates have been incorporated into the design weights (the $d_k$ in equation (2)). Also incorporated into the design weights are the inverse of an estimated probability of a non-vacant HU being a primary residence. All responding primary residences had an estimated probability of 1, and all HU determined not to be primary residences a probability of 0. The rest have been assigned a probability of being a primary residence based on a logistic regression conducted among partially or fully responding HUs to either the National Pilot or its NRFU survey for which primary residence status could be determined.

Turning to nonresponse adjustment, after investigating a longer list of candidate variables, the logistic model used to fit a response model in the National Pilot contains indicators for 17 geographic area (groups of states), indicators for the 4 protocols, indicators for the 2 incentive levels, an urbanicity indicator, an indicator of whether the HU is a single-family dwelling units from the frame, the fraction of HUs owned in the Census block group (CBG) containing the HU, and the fraction of HUs in its CBG with annual incomes less than \$60,000. The latter two are estimated from the 2010 American Community Survey.

The WTADJUST procedure in SUDAAN® (Research Triangle Institute 2012) has been used to compute the calibration weights for the National Pilot. The procedure removes the extraneous calibration variables that would cause a singularity in matrix inversion (e.g., because the four protocol levels and two incentive levels cannot all define non-singular calibration variables).

WTADJUST has also been used to choose the variables for the National Pilot's missing-at-random logistic response model, which assumed the components of $\mathbf{x}_k$ in equations (1) and (2) were the same as those in $\mathbf{z}_k$. WTADJUST fits a logistic model very much like SUDAAN's pseudo-maximum-likelihood LOGISTIC (RLOGIST) procedure but with a different estimating equation (WTADJUST uses equation (2) to solve for $\mathbf{g}$ rather than $\sum_R d_k\mathbf{z}_k = \sum_S \{d_k/[1 + \exp(-\mathbf{z}_k^T\mathbf{g})]\}\mathbf{z}_k$). The output for the estimates of the components of $\mathbf{g}$ looks the same except that the estimates themselves differ. In fact, their signs will usually be reversed (because $p_k = 1/[1 + \exp(\mathbf{z}_k^T\mathbf{g})]$ using WTADJUST and $1/[1 + \exp(-\mathbf{z}_k^T\mathbf{g})]$ using LOGISTIC). The logistic functional form is, in fact, only a special case of the weight-adjustment functions fit by WTADJUST, but we restrict our attention to that form here until the concluding section.

## 4. Converting Proxy Variables into Model-Only Variables

The response model fit for the National Pilot contains three model variables that logic suggests would be better replaced by survey variables: the frame indicator for a single-family dwelling unit, the CBG fraction of owned HUs, and the CBG fraction of HUs with annual income less than $60,000.

Using the model variables described in the previous section as the calibration variables in fitting a missing-at-random (MAR) logistic response model, Table 1 shows the adjusted *F* values and their associated *p*-values produced by the WTADJUST (which uses equation (3) to estimate variances by setting DESIGN = WR ADJUST = NONRESPONSE and NEST _ONE_). All the model variables are significant at the .15 level and have an *F* value greater than 2.5.

**Table 1.** MAR: Model Variable and Calibration Variables are the Same

| Variable | Adjusted Wald F | *p*-value |
|---|---|---|
| GEOGRAPHICAL AREA | 4.63 | 0.0000 |
| INCENTIVE | 17.63 | 0.0000 |
| PROTOCOL | 8.76 | 0.0000 |
| URBANICITY INDICATOR | 3.19 | 0.0741 |
| CENSUS BLOCK GROUP INCOME ≤ 60K? | 8.44 | 0.0037 |
| FRACTION OWNED IN CBG | 2.52 | 0.1128 |
| SINGLE-FAMILY UNIT–FRAME | 6.95 | 0.0000 |

Table 2 show what happens when the three survey variables discussed above replace their proxy frame values as model variables but not calibration variables. This is NMAR1 fit using WTADJX. Only annual income less that $60,000 remains significant at the .15 level, while the *F* values of the other two fall below 1. This partly due to their collinearity. Table 3, NMAR2, removes whether the HU is a single-family dwelling unit from the model. All the remaining variable are significant at the .1 level. It should be noted that estimation treats mobile homes and attached single-family units as single-family dwelling units. Removing one of both does not meaningfully change the results however.

**Table 2.** MNAR1: Model Variables Analogous to MAR Model Variables

| Variable | Adjusted Wald F | *p*-value |
|---|---|---|
| GEOGRAPHICAL AREA | 4.51 | 0.0000 |
| INCENTIVE | 14.43 | 0.0001 |
| PROTOCOL | 7.37 | 0.0001 |
| URBANICITY INDICATOR | 2.71 | 0.0996 |
| INCOME ≤ 60K? | 3.30 | 0.0695 |
| HOUSING UNIT OWNED | 0.28 | 0.5938 |
| SINGLE-FAMILY UNIT−SURVEY | 0.00 | 0.9548 |

**Table 3.** MNAR2: MNAR1 with One Insignificant Model Variable Removed

| Variable | Adjusted Wald F | *p*-value |
|---|---|---|
| GEOGRAPHICAL AREA | 4.53 | 0.0000 |
| INCENTIVE | 14.89 | 0.0001 |
| PROTOCOL | 7.98 | 0.0000 |
| URBANICITY INDICATOR | 2.89 | 0.0894 |
| INCOME ≤ 60K? | 5.60 | 0.0179 |
| HOUSING UNIT OWNED | 4.73 | 0.0297 |

A fourth fit, NMAR3, is very similar to the previous one and not shown. It replaces the two shadow calibration variables in NMAR2, the CBG fraction of owned HUs and the CBG fraction of HUs with incomes less than $60,000, with ordinary-least-squares (OLS) predictions of the probability of HU ownership and the probability of having an income less than $60,000, as suggested in Kott and Liao (2016). The regressors in those OLS predictions are the two CBG fractions and the frame indicator of the HU being a single-family dwelling unit.

Tables 4 and 5 displays some estimated means and standard errors computed (using SUDAAN with NEST _ONE replaced by NEST STRATUM PSU) first assuming missingness is completely at random (MCAR; i.e., unit response does not depend on any frame or survey variables), then missing at random as in Table 1, and after that missing not at random under the MNAR assumption and using the MNAR methods described above. All five methods treat the original sample as a stratified two-stage sample, with the original design's 19 strata collapsed into 17 variance strata to avoid variance strata containing only a single primary sampling unit (Census PUMAs). The adjustments for the vacancies and non-primary residences are treated as part of the design weights. Although this is a simplification, it is the same simplification for all nonresponse-adjustment methods.

The results suggest that in all cases, using the MAR method of nonresponse adjustment often corrects for part, but not all, of the bias that failing to make any adjustment due to the characteristics of the nonrespondents, who tend to have less income and are less likely to own their own homes. The bias is worse for the two model variables: fraction of owned housing units and the fraction with annual incomes less than $60,000.

**Table 4.** Estimated Means for Alternative Nonresponse Adjustment Methods

| Variable | MCAR Model | MAR Model | MNAR Models: Full | MNAR Models: Dropped SFDU | MNAR Models: OLS Version |
|---|---|---|---|---|---|
| Bedrooms | 2.9096 | 2.8412 | 2.8078 | 2.8084 | 2.8079 |
| Fraction of HUs with a Dryer | 0.8569 | 0.8344 | 0.8262 | 0.8263 | 0.8261 |
| Fraction of HUs with Central Air | 0.6871 | 0.6750 | 0.6654 | 0.6652 | 0.6651 |
| Fraction of SFDUs | 0.8169 | 0.7830 | 0.7778 | 0.7785 | 0.7782 |
| Fraction of HUs Owned | 0.7134 | 0.6803 | 0.6468 | 0.6451 | 0.6445 |
| Fraction of HUs with Income < 60K | 0.5356 | 0.5516 | 0.6181 | 0.6176 | 0.6173 |

HU – Housing Unit; SFDU – Single-Family Dwelling Unit

**Table 5.** Estimated Standard Errors for Alternative Nonresponse Adjustment Methods

| Variable | MCAR Model | MAR Model | MNAR Models: Full | MNAR Models: Dropped SFDU | MNAR Models: OLS Version |
|---|---|---|---|---|---|
| Bedrooms | 0.0331 | 0.0346 | 0.0344 | 0.0358 | 0.0345 |
| Fraction of HUs with a Dryer | 0.0094 | 0.0107 | 0.0113 | 0.0117 | 0.0113 |
| Fraction of HUs with a Central Air | 0.0136 | 0.0131 | 0.0137 | 0.0134 | 0.0135 |
| Fraction of HUs with a SFDU | 0.0113 | 0.0130 | 0.0148 | 0.0147 | 0.0134 |
| Fraction of HUs Owned | 0.0107 | 0.0114 | 0.0419 | 0.0190 | 0.0162 |
| Fraction of HUs with Income < 60K | 0.0148 | 0.0144 | 0.0302 | 0.0274 | 0.0270 |

HU – Housing Unit; SFDU – Single-Family Dwelling Unit

The mean estimates from using the three MNAR methods are very similar. The standard errors tend to decrease as one goes from using all three model variables through dropping single-family dwelling units from the model and the calibration variables and finally to using the frame values for single-family dwelling units to create proxy calibration variables for the two remaining model-only variables. In most cases, however, even the final NMAR method produced larger estimated standard errors than the MAR and no-model methods. Those standard errors, despite evidence and logic to the contrary, assume that the MAR model and MCAR models, respectively are correct.

## 5. Concluding Remarks

In the RECS example, element response was at first modeled as a function of variables with known values for the entire sample, where some were of those obvious proxies for variables with known values only for respondents. When those proxies were replaced by their model-only analogues in a calibration-weighting equation, one was found no longer to be a contributor of response. Still, following Kott and Liao (2016), this variables was shown to have value in creating shadow variables for model-only values using OLS. As Kott and Liao demonstrated, the resulting calibration-weighted estimator retains its near quasi-probability-sampling unbiasedness despite the somewhat ad-hoc use of OLS.

It is a simple matter to extend the methodology used in the text to other element response functions. In SUDAAN, the weight adjustment function in equation (2) can be replaced by:

$$\alpha(\mathbf{x}_k^T\mathbf{g}) = [L + \exp(\mathbf{x}_k^T\mathbf{g})]/[1 + U^{-1}\exp(\mathbf{x}_k^T\mathbf{g})],$$

the inverse of which is a truncated logistic response model where the probabilities of element response are bound between $1/U \geq 0$ and $1/L \leq 1$. Other smooth monotonic functions can also be $\alpha(.)$, but the user may have to do his/her programming for that.

## Acknowledgements

Much of this work was supported by a grant from the National Science Foundation, award number SES-1424492. The author sincerely thank Dan Liao for her help in preparing this document and in what we hope will be a version published in the refereed literature.

## References

Berry, C. and O'Brien E. (2016). Managing the Fast-Track Transformation of a 35-Year Old Federal Survey, Presented at the 2016 FedCASIC Workshop, Washington DC., https://www.census.gov/fedcasic/fc2016/ppt/2_2_Speed.pdf

Biemer, P., Kott, P., and Murphy, J. (2016). Estimating Mail or Web Survey Eligibility for Undeliverable Addresses: A Latent Class Analysis Approach, Proceedings of the ASA Survey Research Methods Section, to appear.

Chang, T. and Kott, P.S. (2008). Using Calibration Weighting to Adjust for Nonresponse Under a Plausible Model, *Biometrika*, 95, 557-571.

Deville, J. C. (2000), Generalized Calibration and Application to Weighting for Non-response, *COMPSTAT: Proceedings in Computational Statistics, 14th Symposium, Utrecht, The Netherlands,* J.G. Bethlehem and P.G.M. van der Heijden, (Eds.), New York: Springer-Verlag.

Kim, J.K. and Riddles, M. (2012). Some Theory for Propensity Scoring Adjustment Estimator, *Survey Methodology*, 38, 157-165.

Kott, P.S. (2014). Calibration Weighting When Model and Calibration Variables Can Differ. In *Contributions to Survey Statistics* - ITACOSM 2013 *Selected Papers* (pp. 1-18). Cham: Springer, Contributions to Statistics.

Kott, P.S. and Liao, D. (2016). Calibration Weighting for Nonresponse That is Not Missing at Random: Allowing for More Calibration than Response-model Variables, forthcoming, *Journal of Survey Statistics and Methodology.*

Molenberghs, G., Beunckens, C., and Sotto, C. (2008), "Every Missingness Not at Random Model Has a Missingness at Random Counterpart with Equal Fit," *Journal of Royal Statistical Society B*, 70 Part 2, 371-388.

National Research Council. (2010), *The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials*. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Research Triangle Institute (2012), *SUDAAN Language Manual*, Volumes 1 and 2, Release 11. Research Triangle Park, NC: Research Triangle Institute.