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

Phillip S. Kott[1] and Dan Liao[1]

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 annual 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 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 can be done with data from the Residential Energy Consumption Survey National Pilot, a nationally representative web-and-mail survey of American households sponsored by the U.S. Energy Information Administration.

Key words: Model variable; calibration 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 of 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 not missing at random (NMAR). The former means that nonresponse is a function entirely of variables with either known population totals or 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. The calibration-weighting framework in Särndal and Lundström (2005) also allows nonresponse to be not missing at random.

Unfortunately, there is no statistical way to determine whether or not nonrespondents are missing at random. Molenberghs et al. (2008) show that any data set fit by a model

assuming nonrespondents are not missing at random could also be fit by a model assuming nonrespondents are missing at random. As a result, many have argued that techniques like Deville's are best suited for sensitivity analyses. National Research Council (2010; 48, 59) discusses the limitations of what it calls the "inverse probability weighting" method for handling not-at-random missingness.

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 annual 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 United States 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.

Using data from the 2015 national pilot of the (United States) Residential Energy Consumption Survey (RECS) which was conducted by mail and web, we demonstrate how one can compare results of calibration weighting assuming nonresponse is missing at random using proxy variables available on the frame as response model variables with results of calibration weighting where survey variables, more logically related to response than their proxies, replace the proxy variables in the response model, showing in the process how to choose which survey variables to include in the response model.

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 to compensate 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 their alternatives 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 there any item nonresponse among element respondents.

In this article, we follow the quasi-randomization approach in Chang and Kott (2008) and treat unit response as an additional phase of probability sampling, where the response probabilities need to be estimated from the data. Although Kott and Chang (2010) showed that the methods they had proposed have good prediction-model properties, we will not discuss those here.

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 ρ_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

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

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

Calibration weighting begins with the calibration equation:

$$\sum_R d_k \left[1 + \exp\left(\mathbf{x}_k^T \mathbf{g}\right)\right] \mathbf{z}_k = \mathbf{T_z} \tag{2}$$

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 that can be estimated in the full sample (including the unit nonrespondents), $\mathbf{T_z}$ 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 linearizations).

In practice, a $\mathbf{g}$ exists when one can be found satisfying (2). Moreover, that $\mathbf{g}$ is consistent for survey designs under which the expansion estimator for $\mathbf{T_z}$ in the absense of nonresponse is consistent. The appendices in Chang and Kott (2008) lay out the theoretical conditions for $\mathbf{g}$ to exist and be consistent. A more lucid account of the theory underpinning this section can be found in Kott (2014).

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

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

The expression $\alpha\left(\mathbf{x}_k^T \mathbf{g}\right)$ 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 (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 (2017) maintained that a prudent approach would be to include in $\mathbf{z}_k$ all the components of $\mathbf{x}_k$ for which population totals 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 $\mathbf{z}_k$.

Some variables in the RECS National Pilot sample, such as an indicator of 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 as home ownership (yes or no), are model-only variables in $\mathbf{x}_k$. At the same time, a reasonable proxy for each model-only variable, like the fraction of homes owned in its Census block group, can be a shadow variable 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

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

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

An asymptotically unbiased estimator for the quasi-probability variance of $t_y = \sum_R w_k y_k$ (again under mild conditions) is

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

where $e_k = y_k - \mathbf{z}_k^T\mathbf{b}$, $\mathbf{b} = \left[\sum_R d_j\alpha'\left(\mathbf{x}_j^T\mathbf{g}\right)\mathbf{x}_j\mathbf{z}_j^T\right]^{-1}\sum_R d_j\alpha'\left(\mathbf{x}_j^T\mathbf{g}\right)\mathbf{x}_j y_j$, and $\alpha\left(\mathbf{x}_k^T\mathbf{g}\right) = \left[1 + \exp\left(\mathbf{x}_k^T\mathbf{g}\right)\right]$ 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 $(y_k - m_y)/\sum_R w_j$.

It is easy to see that due to calibration $\sum_R w_k y_k - \sum_S d_k y_k = \sum_R w_k e_k$ (which also provides a heuristic justification for Equation (4)). We thus have the following estimate for the increase in quasi-probability variance due to nonresponse and nonresponse adjustment:

$$\mathrm{var}\left\{\sum_R d_k\left[1 + \exp\left(\mathbf{x}_k^T\boldsymbol{\gamma}\right)\right] y_k \Big| \sum_S d_k y_k\right\} = \sum_R d_k^2\left(1/p_k^2\right)(1 - p_k)e_k^2$$

$$= \sum_R d_k^2\left[1 + \exp\left(\mathbf{x}_k^T\mathbf{g}\right)\right]\exp\left(\mathbf{x}_k^T\mathbf{g}\right)e_k^2, \tag{5}$$

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 of 9,650 dwelling units 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 USD 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 USD 5 initially. One provided an extra USD 10 upon completion while the other provided an extra USD 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.

Roughly 40% of eligible HUs responded to the RECS National Pilot. 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 four protocols, indicators for the two 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 USD 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 regression procedure (RLOGIST) but with a different estimating equation (WTADJUST solves for $\mathbf{g}$ in Equation (2) rather than in $\sum_R d_k \mathbf{z}_k = \sum_S \left\{ d_k / \left[1 + \exp\left(-\mathbf{z}_k^T \mathbf{g}\right) \right] \right\} \mathbf{z}_k$). 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 more reasonably 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 USD 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 2 show what happens when the three survey variables discussed above replace their proxy frame values in the model vector but not in the calibration vector. This is denoted as NMAR1 and fitted using WTADJX. Only annual income less than USD 60,000 remains significant at the .15 level, while the *F* values of the other two fall below 1. This is partly due to collinearity among them. In Table 3, NMAR2 removes whether the HU is a single-family dwelling unit from the model vector. All the remaining variables 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.

A fourth fit, NMAR3, containing the same model variables as NMAR2 with similar results is not shown. It replaces the two shadow calibration variables in NMAR2, the CBG fraction of owned HUs and the CBG fraction of HUs with annual incomes less than USD 60,000, with ordinary-least-squares (OLS) predictions of the probability of HU ownership and the probability of having an annual income less than USD 60,000, as suggested in Kott and Liao (2017). The regressors in those OLS predictions are the two CBG fractions and the frame indicator of the HU being a single-family dwelling unit.

Table 4 displays a number of estimated means and (quasi-probability) standard errors computed (with SUDAAN and NEST _ONE_ replaced by NEST STRATUM PSU to capture stratification and clustering effects on the estimated means) first assuming missingness is completely at random (MCAR; i.e., unit response does not depend on any frame or survey variables and both model and calibration vectors only have an intercept), then missing at random as in Table 1, and after that missing not at random under the NMAR assumption and using the three NMAR 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 (PSU). The PSUs in the RECS National Pilot design are 2010 US Census Public Use Microdata Areas (PUMAs, http://www.census.gov/geo/reference/puma.html).

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
CBG ANNUAL INCOME ≤ $60K?	8.44	0.0037
FRACTION OWNED IN CBG	2.52	0.1128
SINGLE-FAMILY UNIT(FRAME	6.95	0.0000

CBG – Census Block Group.

Table 2.   NMAR1: three model-only variables and three shadow proxies.

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
ANNUAL INCOME ≤ $60K?	3.30	0.0695
HU OWNED	0.28	0.5938
SINGLE-FAMILY UNIT(SURVEY	0.00	0.9548

HU – Housing Unit.

The adjustments for the vacancies and non-primary residences are treated in variance estimation here as part of the design weights. Although this is a simplification, it is the same simplification for all five nonresponse-adjustment methods.

The results in Table 4 are summarized in Tables 5 and 6 and extended to three domain estimates: one for owned HUs (a model-only variable in the NMAR models), one for detached standing HUs, and one for HUs built before 1970. The measure $\log(X) - \log(Y) = \mathrm{Log}(X/Y)$ used in those tables is close to the percent difference between X and Y when that difference is less than 40% ($\mathrm{Log}(X/Y) \approx (X - Y)/Y$). Unlike percent differences, however, it is a symmetric measure (i.e., $\mathrm{Log}(X/Y) = -\mathrm{Log}(Y/X)$).

In Table 5, we see that the estimates from using the three NMAR methods always fall within 0.5% of each other. Assuming that these models more reasonably reflect reality than the MAR model, which in turn is more reasonable than the MCAR model, it appears that adjusting for nonresponse using an MAR model removes more than half of the bias relative to not adjusting at all (i.e., assuming unit nonresponse is completely at random). The sizes of the relative biases vary, with those associated with the two model-only variables (the fractions of HU owned and with annual income less than USD 60K) being the largest. Observe that the relative biases tend to be smaller for a domain related, or correlated to, the model-only variables (e.g., having a detached HU is correlated with both ownership and HU annual income).

In Table 6, we see that the estimated standard errors are, on average, lowest when the MAR is used, except for the domain of owned HUs. Using NMAR1 has, on average, the highest estimated standard errors while using NMAR3 has, on average, the lowest among the three NMAR methods but still higher estimated standard errors than when the MAR is used. The results appear to vary by variable, however.

Table 3.   NMAR2: NMAR1 with an insignificant model-only 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
ANNUAL INCOME ≤ 60K?	5.60	0.0179
HU OWNED	4.73	0.0297

HU – Housing Unit.

Table 4.   Comparing alternative models of nonresponse adjustment.

Housing unit variable	Models				
	MCAR	MAR	NMAR1	NMAR2	NMAR3
Estimated means					
FRACTION BUILT BEFORE 1970	0.3802	0.3864	0.3946	0.3947	0.3947
FRACTION BUILT AFTER 1999	0.2227	0.2194	0.2124	0.2124	0.2124
DAYS SOMEONE IS AT HOME	3.5096	3.4940	3.5209	3.5203	3.5198
FRACTION WITH AN ATTIC	0.4953	0.4726	0.4607	0.4609	0.4607
NUMBER OF BEDROOMS	2.9096	2.8412	2.8078	2.8084	2.8079
FRACTION WITH A CELLAR	0.3504	0.3187	0.3153	0.3155	0.3154
FRACTION WITH CENTRAL AIR	0.6871	0.6750	0.6654	0.6652	0.6651
FRACTION WITH CLOTHES WASHER	0.8706	0.8499	0.8429	0.8431	0.8429
NUMBER OF DESKTOPS	0.5518	0.5404	0.5293	0.5292	0.5291
FRACTION WITH DISHWASHER	0.7417	0.7305	0.7119	0.7117	0.7116
FRACTION WITH DRYER	0.8569	0.8344	0.8262	0.8263	0.8261
FRACTION WITH HOME HEATING	0.9585	0.9538	0.9530	0.9530	0.9530
FRACTION WITH INTERNET	0.8745	0.8725	0.8616	0.8617	0.8617
FRACTION OWNED	0.7134	0.6803	0.6468	0.6451	0.6445
FRACTION WITH ANNUAL INCOME < $60K	0.5356	0.5516	0.6181	0.6176	0.6173
HUMBER OF HOUSEHOLD MEMBERS	2.5583	2.5372	2.5236	2.5255	2.5257
NUMBER OF REFRIGERATORS	1.4204	1.3941	1.3774	1.3774	1.3773
NUMBER OF LAPTOPS	1.0701	1.0676	1.0327	1.0333	1.0335
NUMBER OF TABLETS	0.9964	0.9783	0.9418	0.9425	0.9426
FRACTION WITH A GARAGE	0.4653	0.4391	0.4270	0.4272	0.4270
NUMBER OF ROOMS	6.5633	6.4211	6.3422	6.3429	6.3420
NUMBER OF COLORS TVS	2.3651	2.3290	2.3009	2.3015	2.3013
FRACTION OF DETACHED STANDING UNITS	0.6677	0.6330	0.6236	0.6239	0.6236
FRACTION WITH A CENTRAL FURNACE	0.6227	0.6081	0.6066	0.6067	0.6067
FRACTION WITH NATURAL GAS HEATING	0.4884	0.4764	0.4721	0.4721	0.4720
FRACTION WITH ELECTRIC HEATING	0.3541	0.3655	0.3691	0.3692	0.3693

Table 4. *Continued.*

			Models		
Housing unit variable	MCAR	MAR	NMAR1	NMAR2	NMAR3
Estimated standard errors					
FRACTION BUILT BEFORE 1970	0.0148	0.0142	0.0135	0.0142	0.0140
FRACTION BUILT AFTER 1999	0.0113	0.0113	0.0115	0.0114	0.0112
DAYS SOMEONE IS AT HOME	0.0384	0.0371	0.0418	0.0387	0.0388
FRACTION WITH AN ATTIC	0.0119	0.0093	0.0086	0.0093	0.0086
NUMBER OF BEDROOMS	0.0331	0.0263	0.0257	0.0259	0.0251
FRACTION WITH A CELLAR	0.0147	0.0130	0.0129	0.0136	0.0129
FRACTION WITH CENTRAL AIR	0.0136	0.0122	0.0122	0.0119	0.0122
FRACTION WITH CLOTHES WASHER	0.0088	0.0071	0.0068	0.0072	0.0069
NUMBER OF DESKTOPS	0.0133	0.0124	0.0125	0.0118	0.0116
FRACTION WITH DISHWASHER	0.0129	0.0112	0.0114	0.0103	0.0103
FRACTION WITH DRYER	0.0094	0.0075	0.0071	0.0074	0.0073
FRACTION WITH HOME HEATING	0.0061	0.0065	0.0067	0.0068	0.0067
FRACTION WITH INTERNET	0.0070	0.0065	0.0071	0.0068	0.0067
FRACTION OWNED	0.0107	0.0073	0.0535	0.0201	0.0127
FRACTION WITH ANNUAL INCOME < $60K	0.0148	0.0120	0.0351	0.0322	0.0321
HUMBER OF HOUSEHOLD MEMBERS	0.0283	0.0269	0.0588	0.0349	0.0342
NUMBER OF REFRIGERATORS	0.0152	0.0125	0.0114	0.0114	0.0115
NUMBER OF LAPTOPS	0.0260	0.0246	0.0288	0.0249	0.0249
NUMBER OF TABLETS	0.0237	0.0210	0.0290	0.0239	0.0238
FRACTION WITH A GARAGE	0.0164	0.0125	0.0111	0.0108	0.0113
NUMBER OF ROOMS	0.0683	0.0537	0.0493	0.0511	0.0495
NUMBER OF COLORS TVS	0.0315	0.0283	0.0326	0.0294	0.0297
FRACTION OF DETACHED STANDING UNITS	0.0132	0.0092	0.0091	0.0110	0.0090
FRACTION WITH A CENTRAL FURNACE	0.0111	0.0102	0.0104	0.0108	0.0105
FRACTION WITH NATURAL GAS HEATING	0.0151	0.0135	0.0136	0.0137	0.0134
FRACTION WITH ELECTRIC HEATING	0.0135	0.0129	0.0134	0.0131	0.0128

NMAR3 – ols fits of the two model-only variables using the three shadow variables.

Table 5.   Summarizing the relative percent differences of the estimated means using alternative models of nonresponse adjustment across 26 variables.

	Mean	Median	3rd Q	Max
All				
$\|\log(\text{MAR}) - \log(\text{MCAR})\| \times 100$	2.56	2.15	2.96	9.48
$\|\log(\text{NMAR1}) - \log(\text{MAR})\| \times 100$	2.09	1.24	2.58	11.38
$\|\log(\text{NMAR2}) - \log(\text{NMAR1})\| \times 100$	0.04	0.02	0.04	0.26
$\|\log(\text{NMAR3}) - \log(\text{NMAR2})\| \times 100$	0.02	0.02	0.03	0.10
Owned housing unit				
$\|\log(\text{MAR}) - \log(\text{MCAR})\| \times 100$	0.86	0.66	0.95	5.58
$\|\log(\text{NMAR1}) - \log(\text{MAR})\| \times 100$	1.42	0.73	1.17	12.80
$\|\log(\text{NMAR2}) - \log(\text{NMAR1})\| \times 100$	0.05	0.03	0.06	0.22
$\|\log(\text{NMAR3}) - \log(\text{NMAR2})\| \times 100$	0.02	0.01	0.02	0.11
Detached housing unit (excludes mobile homes)				
$\|\log(\text{MAR}) - \log(\text{MCAR})\| \times 100$	0.72	0.47	0.80	5.13
$\|\log(\text{NMAR1}) - \log(\text{MAR})\| \times 100$	1.76	1.09	2.11	14.09
$\|\log(\text{NMAR2}) - \log(\text{NMAR1})\| \times 100$	0.04	0.02	0.04	0.24
$\|\log(\text{NMAR3}) - \log(\text{NMAR2})\| \times 100$	0.01	0.00	0.01	0.08
Built before 1970				
$\|\log(\text{MAR}) - \log(\text{MCAR})\| \times 100$	2.61	2.25	3.27	8.21
$\|\log(\text{NMAR1}) - \log(\text{MAR})\| \times 100$	1.77	0.97	1.50	9.68
$\|\log(\text{NMAR2}) - \log(\text{NMAR1})\| \times 100$	0.05	0.03	0.06	0.32
$\|\log(\text{NMAR3}) - \log(\text{NMAR2})\| \times 100$	0.03	0.02	0.04	0.11

Table 7 tries to get a cleaner picture of the impact of unit nonresponse and the alternative methods of adjusting for it when estimating means for all occupied residences. It computes the square root of the estimated added variance due to nonresponse adjustment computed using Equation (5) (which could not be done in SUDAAN). This measure ignores the impact of any correlation, whether real or random, between the sampling and nonresponse errors. Similarly, it ignores the impact of any within PSU correlations across HUs. The conclusions from Table 6 are amplified. The additional estimated variance from using MAR is always less than that from using any NMAR method. The additional estimated variance from using MAR is also less than that from doing nothing (i.e., MCAR) at least 75% of the time (since the third quartile is negative), contrary to popular belief, a possibility pointed out by Little and Vartivarian (2005). Similarly, the added estimated variances drop in over 75% of the cases (in fact, all but one case) when NMAR2 replaces NMAR1 and NMAR3 replaces NMAR2.

5.   Concluding Remarks

The primary purpose of this article was to show how the theoretical and simulation results from Kott and Liao (2017) could be applied to a real survey suffering from a relatively large fraction of unit nonresponse (roughly 60%). In creating calibration weights to compensate for units nonresponse to RECS National Pilot survey, element response was at first modeled as a function of variables with known values for the entire sample, where

Table 6. Summarizing the relative differences of the estimated standard errors using alternative models of nonresponse adjustment across 26 variables.

	Mean	Min	1st Q	Median	3rd Q	Max
All						
(log(MAR) − log(MCAR)) × 100	−14.19	−38.80	−22.27	−11.78	−5.69	6.67
(log(NMAR1) − log(MAR)) × 100	16.40	−11.16	−3.36	1.17	11.87	199.48
(log(NMAR2) − log(NMAR1)) × 100	−6.96	−97.94	−8.69	−1.26	3.61	18.75
(log(NMAR3) − log(NMAR2)) × 100	−3.74	−45.83	−2.99	−1.85	0.11	4.51
Owned housing unit						
(log(MAR) − log(MCAR)) × 100	1.94	−5.51	−0.40	0.58	3.36	13.36
(log(NMAR1) − log(MAR)) × 100	17.51	−0.44	4.99	17.58	21.83	99.71
(log(NMAR2) − log(NMAR1))) × 100	−10.48	−25.02	−16.48	−9.74	−1.99	−0.33
(log(NMAR3) − log(NMAR2)) × 100	−0.56	−3.62	−1.10	−0.47	0.01	1.98
Detached housing unit (excludes mobile homes)						
(log(MAR) − log(MCAR)) × 100	−3.45	−13.84	−7.26	−3.16	−0.09	6.57
(log(NMAR1) − log(MAR)) × 100	21.32	−7.56	−0.13	3.60	22.63	194.02
(log(NMAR2) − log(NMAR1)) × 100	−10.32	−117.46	−12.08	−2.92	0.39	6.40
(log(NMAR3) − log(NMAR2)) × 100	−1.81	−17.75	−2.40	−0.64	−0.06	1.40
Built before 1970						
(log(MAR) − log(MCAR)) × 100	−5.08	−15.61	−8.86	−4.81	−1.73	10.44
(log(NMAR1) − log(MAR)) × 100	12.72	−5.11	−2.83	2.27	8.68	154.28
(log(NMAR2) − log(NMAR1)) × 100)	−6.93	−92.88	−7.75	−1.32	2.23	6.93
(log(NMAR3) − log(NMAR2)) × 100	−2.69	−34.99	−3.17	−0.37	0.17	1.75

Table 7. Summarizing the relative differences of the increases of estimated standard errors using alternative models of nonresponse adjustment across 26 variables*.

	Mean	Min	1st Q	Median	3rd Q	Max
(log(MAR) − log(MCAR**)) × 100	−19.02	−78.53	−36.08	−12.52	−4.19	6.50
(log(NMAR1) − log(MAR))) × 100	66.57	8.71	19.03	30.45	60.70	432.61
(log(NMAR2) − log(NMAR1)) × 100	−10.96	−205.68	−28.86	−1.97	23.38	76.69
(log(NMAR3) − log(NMAR2)) × 100	−16.95	−88.08	−27.98	5.23	−2.27	0.04
(log(NMAR3) − log(NMAR1)) × 100	−27.90	−285.54	−34.27	−6.42	−2.59	0.85
(log(NMAR3) − log(MAR)) × 100	38.67	2.58	10.03	19.72	45.93	261.25

*The increases are computed as the square root of the right-hand side of Equation (5).
**MCAR treats the intercept as the lone model and calibration variable.

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 (2017), this type of 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.

With this data, there appeared to be gains in bias reduction from assuming reasonably that nonresponse was a logistic function of survey variables rather than their frame proxies (which were several years old when based on the ACS). With the largest bias reductions in those survey variables added to the response model. There was, however, a marked tendency for the standard errors to increase when NMAR modeling replaced MAR modeling. The reader should keep in mind that the results from the RECS National Pilot may not generalize to other surveys.

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

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

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 used $\alpha(.)$, but the user may have to do his/her own programming for that. Choosing an appropriate form for the response function and the penalty for failing to do so is an area for future research.

Finally, the reader should be aware that there are packages in R that can implement calibration weighting similar to the routine in SUDAAN. One such is 'Sampling' (Tille and Matei 2013).

6. References

Berry, C. and E. O'Brien. 2016. "Managing the Fast-Track Transformation of a 35-Year Old Federal Survey." Presented at the 2016 FedCASIC Workshop, Washington DC, May 4, 2016. Available at: https://www.census.gov/fedcasic/fc2016/ppt/2_2_Speed.pdf (accessed September 2016).

Biemer, P., P. Kott, and J. Murphy. 2016. "Estimating Mail or Web Survey Eligibility for Undeliverable Addresses: A Latent Class Analysis Approach." Proceedings of the Survey Research Methods Section: American Statistical Association, Chicago, IL, August 2016, 1166–1172. Available at: https://ww2.amstat.org/sections/srms/Procee-dings/y2016/files/389587.pdf.

Chang, T. and P.S. Kott. 2008. "Using Calibration Weighting to Adjust for Nonresponse under a Plausible Model." *Biometrika* 95: 557–571. Doi: http://dx.doi.org/10.1093/biomet/asn022 (A version for the full appendices can be found here: http://ageconsearch.umn.edu/handle/234362).

Deville, J.C. 2000. "Generalized Calibration and Application to Weighting for Non-response." COMPSTAT: Proceedings in Computational Statistics, 14th Symposium,

Utrecht, The Netherlands, edited by J.G. Bethlehem and P.G.M. van der Heijden. New York: Springer-Verlag. Doi: http://dx.doi.org/10.1007/978-3-642-57678-2.

Kim, J.K. and M. Riddles. 2012. "Some Theory for Propensity Scoring Adjustment Estimator." *Survey Methodology* 38: 157–165.

Kott, P. 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. Doi: https://doi.org/10.1007/978-3-319-05320-2_1.

Kott, P. and T. Chang. 2010. "Using Calibration Weighting to Adjust for Nonignorable Unit Nonresponse." *Journal of the American Statistical Association* 105: 1265–1275. Doi: http://dx.doi.org/10.1198/jasa.2010.tm09016.

Kott, P. and D. Liao. 2017. "Calibration Weighting for Nonresponse that is Not Missing at Random: Allowing for More Calibration than Response-model Variables." *Journal of Survey Statistics and Methodology* 5(2): 159–174. Doi: https://doi.org/10.1093/jssam/smx003.

Little, R. and S. Vartivarian. 2005. "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology* 31: 161–168.

Molenberghs, G., C. Beunckens, and C. Sotto. 2008. "Every Missingness Not at Random Model has a Missingness at Random Counterpart with Equal Fit." *Journal of Royal Statistical Society B* 70: 371–388. Doi: http://dx.doi.org/10.1111/j.1467-9868.2007.00640.x.

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. Doi: http://dx.doi.org/10.17226/12955.

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

Särndal, C-E. and S. Lundström. 2005. *Estimation in Surveys with Nonresponse*. John Wiley & Sons: Chichester. Doi: http://dx.doi.org/10.1002/0470011351.

Tille, Y. and A. Matei. 2013. Package 'Sampling.' A software routine available at" http://cran.r-project.org/web/packages/sampling/sampling.pdf (accessed October 2017).