Calibration for Nonresponse Treatment Using Auxiliary Information at Different Levels

Per Gösta Andersson, Carl-Erik Särndal[1]

Abstract

This paper explores the different ways in which auxiliary information can be put to use in calibrated weighting adjustment under survey nonresponse. Information is often present at two levels, the population level and the sample level. The many options available in executing the calibration derive from several factors: One is the order in which the two sources of information enters into calibration, a choice of a bottom-up as opposed to a top-down approach. Another is whether the calibration should be carried out sequentially in two steps, or in one single step with the combined information. A third question is whether one can simplify the procedure, at no major loss of accuracy, by transcribing individual population auxiliary data from the register to the sample units only. We make a systematic list of the possibilities arising for calibration adjustment in this setting. An empirical study concludes the paper.

Key Words: Bottom-up, moon vector, star vector, top-down

1. Introduction

The nonresponse affecting most sample surveys today continues to pose methodological challenges because of the bias caused in estimates of population parameters. Nonresponse rates are high and continue to rise, exacerbating the problem. Nonresponse adjustment weighting is commonly used in the estimation. The broad possibilities that this technique offers, especially for calibration weighting, are explored in this article.

We need to distinguish different levels of availability of variable values: The population level, the sample level, the response level. The sample is drawn by probability sampling from the population. The response is the subset of the sample for which the study variables values (the y-variable values) are individually observed.

Auxiliary variables are essential. To qualify as auxiliary, a variable must contain information at a higher level than the response, and its value must be known individually for all units in the response.

The use of auxiliary variables contributes to two important objectives in estimation: A reduction of variance and a reduction of nonresponse bias. A considerable literature exists. A recent review paper [1] brings up many of the issues in unit nonresponse weighting adjustments and is a useful starting point for reading.

We agree with the assessment of Brick (2013), p. 330: "... survey estimates may be biased even after the adjustments. Nonresponse also causes a loss in the precision of survey estimates, primarily due to reduced sample size and secondarily as the result of increased variation of the survey weights. However, bias is the dominant component of the nonresponse-related error in the estimates, and nonresponse bias generally does not decrease as the sample size increases. Thus bias is often the largest component of mean square error of the estimates even for subdomains when the sample size is large". The same author, p. 334, notes: "The auxiliary variables are very valuable for adjusting the design weights to account for nonresponse."

For this article, two features of an auxiliary variable must be distinguished:

[1]Per Gösta Andersson, Department of Statistics, Stockholm University, email: per.gosta.andersson@stat.su.se. Carl-Erik Särndal, email: car.sarndal@telia.com.

- The *information* about the variable. It specifies the detailed knowledge about the variable available at the outset; it indicates the potential use in estimation.

- The *role* of the variable in the calibration estimation technique.

We consider two possible *roles* in estimation of the auxiliary variable:

- As a known population total.

- As variable values known for all units in the sample.

This distinction has been described as population-based as opposed to sample-based, as Kalton and Kasprzyk (1986); or Info-U as opposed to Info-S, as in Lundström and Särndal (1999) and Särndal and Lundström (2005).

The *information* can be of several types, depending on the survey environment. Different countries, and their national statistical agencies, have different auxiliary resources for their production of national statistics. Examples of types of information are the following:

- The auxiliary variable value is known individually for all units in the population. In the Scandinavian countries, many auxiliary potential auxiliary variables fit this description, emanating from exhaustive registers of the whole population and transcribable to the survey data file using the unique personal identifier as a key. This generates two roles of the information in the calibration: As a known population total, obtainable by adding up the individual values; As values used for the sample units only, although individually known for all population units. This deliberate decision to forego part of the information is justified when squared bias rather than variance is the main contributor to mean squared error and/or important savings in time and other production factors are realized by transcribing the auxiliary values from the registers (where they reside for the whole population) to the much smaller sample file only.

- Population totals are taken from reliable sources outside the survey itself; totals are known (imported), but individual values are unknown at the population level. A typical example is the long time use in the Canadian Labour Force Survey of up-to-date results from demographic modeling, hence not exact, but deemed sufficiently close, about the population count in categories by sex, age group and region. These counts are treated as known population counts in the calibration. Brick (2013), p. 334 emphasizes: ”The population-based adjustment is especially useful when characteristics for the entire sample are not available but the population totals are known, because these adjustments only require capturing the data from the respondents.”

- The auxiliary variable value is known (observed) for the sample units only, and thereby known individually for the respondents also. Data of this kind are known as paradata, consisting e.g. of information from the data collection process. Examples include data for a sample unit such as the number of (telephone) contact attempts attempted before the unit responds or is declared nonrespondent, the identity of the interviewer handling the unit, and others. Brick (2013) p. 334 also notes: ”Sample-based adjustments need data for the full sample but do not require knowing control totals for the entire population. Sample-based and population-based adjustments are equally effective for dealing with nonsense bias. (Särndal and Lundström (2005) and Brick and Jones (2008).”

A much practiced simple use of auxiliary information is the weighting class technique. The inverse of the response proportion in sample subgroups is the basis for this weighting. Kalton (1983), p. 63 states: "Among the potential variables for use in forming weighting classes, the ones that are most effective in reducing nonresponse bias are those that are highly correlated both with the survey variables and the (0,1) response variable." However, more generally, the adjustments need not be based on a grouping of units.

This article focuses on calibration estimation, a general technique for the use of auxiliary information. Basic calibration theory is found in Kalton (1983); its uses in connection with nonresponse has been explored by a number of authors including Deville (1995), Kott (2006), Kott and Chang (2010), Kott and Liao (2012), Lundström and Särndal (1999) and Särndal and Lundström (2005).

A wide spectrum of possibilities are included in the class of calibration estimators. As Brick (2013) notes: "A wide variety of nonresponse adjustment estimators are in this class, including poststratification, raking, and generalized regression estimators". In the context of nonresponse, a variety of uses of calibration arises because information may be available at the population level and at the sample level.

Should the information be combined and be put to use in a single step of calibration? Or should the two sources of information enter into calibration one at a time, and then in which order? In the latter case, bottom-up calibration is one possibility: The first calibration is then carried out from the response level to the sample level, using only the sample information. The second calibration step uses the population information. In top-down calibration, the first calibration step links the population to the sample using the population level information only. The second step then usues the sample information. In either case, both the sample information and, if available, the population information should in the end have come to use.

The article is arranged as follows: Section 2 introduces notation and theoretical arguments for the different types of calibration estimators that we will consider. Based on these proposed estimators, Section 3 then discusses some of the scenarios outlined in the Introduction. In Section 4 properties of the estimators are studied through a simulation study.

2. Notation and outline of estimators

Some notation: The population is $U = \{1, \dots, k, \dots N\}$; s denotes a probability sample from U; the inclusion probability for unit k is π_k, and its sampling weight is $d_k = 1/\pi_k$; r is the response set obtained from s; $r \subset s \subset U$. The value y_k is observed for $k \in r$ only.

We study and compare a number of different alternatives for obtaining the calibrated weights w_k in estimators of the general form $\hat{Y} = \sum_r w_k y_k$ of the population total $Y = \sum_U y_k$. We expect to find differences between those alternatives, both with regard to bias and to variance.

The auxiliary variables are of two types, depending on the information available; it can be at the population level or at the sample level. Variables of the first type make up a "star vector" $\mathbf{x}_k^*$, those of the second type make up a "moon vector" $\mathbf{x}_k^o$. The vector values $\mathbf{x}_k^*$ and $\mathbf{x}_k^o$ are known for $k \in s$, that is, for respondents as well as for nonrespondents. Further we assume about $\mathbf{x}_k^*$ that it is known for all $k \in U$ (as when taken from a complete population register) or, in some situations, that at least the population total $\sum_U \mathbf{x}_k^*$ is known from a reliable source. On the other hand, $\sum_U \mathbf{x}_k^o$ is unknown, but the important fact that it can be estimated without bias by $\sum_s d_k \mathbf{x}_k^o$ contributes to a reduction of the nonresponse bias of $\hat{Y}$.

The final weights w_k are obtained by calibrating on a vector denoted $\mathbf{x}_k$. It can have different forms. Three components need to be specified for the computation of the final weights w_k:

- The specification of the auxiliary vector $\mathbf{x}_k$;
- The calibration constraint denoted $\mathbf{X}$ for that given $\mathbf{x}$-vector;
- The starting weights for the calibration that gives the final weights w_k.

The calibration constraint states the auxiliary information that is available, or that we decide to utilize, in computing the final weights w_k. The starting weights for the final calibration can be of two kinds:

- Direct, meaning that the starting weights are the already known sampling weights d_k;
- Intermediary, meaning that a preliminary calibration step is carried out, resulting in preliminary weights to be used in the calibration that leads to the final weights w_k.

We consider three different specifications of the x-vector:

$$A: \quad \mathbf{x}_k = \begin{pmatrix} \mathbf{x}_k^* \\ \mathbf{x}_k^o \end{pmatrix}$$

$$B: \quad \mathbf{x}_k = \mathbf{x}_k^*$$

$$C: \quad \mathbf{x}_k = \mathbf{x}_k^o$$

All vectors $\mathbf{x}_k$ used here are of the form $\boldsymbol{\mu}'\mathbf{x}_k = 1$, for all k, for some constant vector $\boldsymbol{\mu}$. For the vector specification A we consider three forms for the calibration constraint $\mathbf{X}$, giving rise to the cases $A1$ to $A3$:

$$A1: \quad \boldsymbol{X} = \begin{pmatrix} \sum_U \mathbf{x}_k^* \\ \sum_s d_k \mathbf{x}_k^o \end{pmatrix}$$

$$A2: \quad \boldsymbol{X} = \begin{pmatrix} \sum_s d_k \mathbf{x}_k^* \\ \sum_s d_k \mathbf{x}_k^o \end{pmatrix}$$

$$A3: \quad \boldsymbol{X} = \begin{pmatrix} \sum_U \mathbf{x}_k^* \\ \sum_s w_{sk} \mathbf{x}_k^o \end{pmatrix}$$

where

$$w_{sk} = d_k \times (\sum_U \boldsymbol{x}_k^*)'(\sum_s d_k \boldsymbol{x}_k^* \boldsymbol{x}_k^{*'})^{-1} \boldsymbol{x}_k^*, \tag{1}$$

which is achieved through a calibration from the sample s to the population U.

For the vector specification B we consider also two variations for the calibration constraint $\mathbf{X}$, giving rise to the cases $B1$ and $B2$:

$$B1: \quad \boldsymbol{X} = \sum_U \mathbf{x}_k^*$$

$$B2: \quad \boldsymbol{X} = \sum_s d_k \mathbf{x}_k^*$$

Finally, for the vector specification C, we consider

$$C1: \quad \boldsymbol{X} = \sum_s d_k \mathbf{x}_k^o$$

$$C2: \quad \boldsymbol{X} = \sum_s w_{sk} \mathbf{x}_k^o$$

2.1 The Direct approach

The weights w_k in $\hat{Y} = \sum_r w_k y_k$ are computed directly in a single-step calibration, using the sampling weights d_k as starting weights. Weights calibrated to the specified information $\mathbf{X}$ are given by

$$w_k = d_k \times \boldsymbol{X}'(\sum_r d_k \boldsymbol{x}_k \boldsymbol{x}_k')^{-1} \boldsymbol{x}_k$$

We get five cases, $A1dir$, $A2dir$, $B1dir$, $B2dir$ and $C1dir$.

2.2 Two-step approaches

As an alternative to the direct approach we have two approaches in two steps, called bottom-up and top-down.

2.2.1 The Bottom-up approach

The name indicates that intermediate weights are obtained by calibrating first from the response set r to the sample s, using the auxiliary vector $\boldsymbol{x}_k^o$ and the constraint $\sum_s d_k \boldsymbol{x}_k^o$. The population information plays no role in that step. Then in the final (second) step, the $\boldsymbol{x}_k^*$ vector is also brought into the picture. The intermediate weights are

$$w_k^o = d_k \times (\sum_s d_k \boldsymbol{x}_k^o)'(\sum_r d_k \boldsymbol{x}_k^o \boldsymbol{x}_k^{o'})^{-1} \boldsymbol{x}_k^o$$

These are used as starting weights in computing the final weights

$$w_k = w_k^o \times \boldsymbol{X}'(\sum_r w_k^o \boldsymbol{x}_k \boldsymbol{x}_k')^{-1} \boldsymbol{x}_k$$

needed for the estimator $\hat{Y} = \sum_r w_k y_k$. This gives another four cases, $A1BU$, $A2BU$, $B1BU$ and $B2BU$.

2.2.2 The Top-down approach

The name indicates that intermediate weights are obtained by calibrating first from the sample s to the population U, using the population information $\sum_U \boldsymbol{x}_k^*$. The response set r and the vector $\boldsymbol{x}_k^o$ play no role in that step, which gives the intermediary weights

$$w_{sk} = d_k \times (\sum_U \boldsymbol{x}_k^*)'(\sum_s d_k \boldsymbol{x}_k^* \boldsymbol{x}_k^{*'})^{-1} \boldsymbol{x}_k^*$$

as in (1). In the final (second) step, the response set comes into play, by a calibration from r to s. The constraints to be used are A1, A2, A3, C1 and C2.

In the second step, we can, for the constraints given by A3 and C2, choose between using the sampling weights d_k and using the intermediary weights w_{sk}. The first case gives

$$w_k = d_k \times \boldsymbol{X}'(\sum_r d_k \boldsymbol{x}_k \boldsymbol{x}_k')^{-1} \boldsymbol{x}_k,$$

leading to A3TD1 and C2TD1.

For the constraints A1, A2, A3, C1 and C2 we can also use the intermediary w_{sk} and get

$$w_k = w_{sk} \times \boldsymbol{X}'(\sum_r w_{sk} \boldsymbol{x}_k \boldsymbol{x}_k')^{-1} \boldsymbol{x}_k,$$

leading to A1TD2, A2TD2, A3TD2, C1TD2 and C2TD2.

For both cases, these w_k are the final weights used in $\hat{Y} = \sum_r w_k y_k$.

(As we have assumed that $\boldsymbol{\mu}'\mathbf{x}_k = 1$, for all k, we do not need to consider the case where $\sum_s d_k \boldsymbol{x}_k^*$ is used in place of $\sum_U \boldsymbol{x}_k^*$ in (1), since this leads to $w_{sk} = d_k$.)

In total we thus have 16 cases, arranged in tabular form as follows:

Table 1: Organization of the calibration approaches

Auxiliary vector	Vector with constraint	Estimators			
		Direct	Two-step		
			Bottom-up	Top-down1	Top-down2
A	A1	A1dir	A1BU	-	A1TD2
A	A2	A2dir	A2BU	-	A2TD2
A	A3	-	-	A3TD1	A3TD2
B	B1	B1dir	B1BU	-	-
B	B2	B2dir	B2BU	-	-
C	C1	C1dir	-	-	C1TD2
C	C2	-	-	C2TD1	C2TD2

3. Preliminary discussion

Only the A1 cases use the full available information $\boldsymbol{X} = \begin{pmatrix} \sum_U \mathbf{x}_k^* \\ \sum_s d_k \mathbf{x}_k^o \end{pmatrix}$; the other cases forego parts of it. This has consequences for bias and variance that we want to explore. The variance of $\hat{Y}$ can be expected to be greater in A2, B1 and B2, compared with A1, which benefits from the known population total $\sum_U \boldsymbol{x}_k^*$. But the consequences for the bias are difficult to foresee without a more detailed investigation. A justification for A2 is a simplified estimation task: Although the values $\boldsymbol{x}_k^*$ may be available in the administrative register for the whole population, A2 requires them to be transcribed only to the sample data file, which saves time and effort; we need to see if the result is an appreciable loss of accuracy in the estimates compared with A1. Cases A2, B1 and B2 imply seemingly severe reductions of the full information content, especially A2 and B2. Nevertheless, the implications with regard to bias and variance are not a priori obvious and need to be assessed.

The B1BU approach is commonly used in practice. Some statisticians justify it by saying that nonresponse adjustment should be carried out first, resulting in nonresponse adjusted weights, w_k^o, to be used in a calibration on the population information, and on only that information. Hence some are prepared to argue that B1BU is better than A1BU, and this despite the fact that A1BU uses more information for the final calibration. Our prediction before doing any empirical testing is that the bias for A1BU is likely to be smaller, but the variance perhaps greater, than for B1BU, so that a trade-off situation, lower bias as against higher variance, may occur. Another question is whether the choice of starting weights is important, that is, whether deriving intermediary weights really serves a useful purpose; could one not calibrate directly on the stated information in one single step? This calls for pair-wise comparisons: Is A1BU/TD better than A1dir, is A2BU/TD better than A2dir, and so on.

4. A simulation study

4.1 Notation and outline

We studied the behaviour of the estimators in Table 1 by a small Monte Carlo simulation. The population used is a dataset called KYBOK, consisting of $N = 832$ clerical municipalities in Sweden in 1992. (We have closely followed the format of the simulation studies reported in Särndal and Lundström (2005).) The municipalities are divided into four groups according to size, from smallest to largest, with group sizes $N_1 = 218$, $N_2 = 272$, $N_3 = 290$ and $N_4 = 52$. The study variable y_k is ”Expenditure on administration and maintenance” (with total $Y = 1\ 025\ 983$), the moon vector is $\mathbf{x}_k^o = (x_{1k}^o, \ldots, x_{4k}^o)'$, where $x_{ik}^o = 1$, if the element k belongs to population group i, and otherwise 0. The star vector is $\mathbf{x}_k^* = (1, x_k)'$, where x_k is the square root of ”Revenue advances”. We use response probabilities determined by $\theta_k = 1 - \exp(-0.0318 x_k)$, $k \in U$, which (for this population) leads to an average non-response probability of 0.14.

10 000 simple random samples, each of size $n = 300$, were independently generated. For each such sample s, a response set r was created by performing 300 independent Bernoulli trials, one for each unit, with probability θ_k of success (response), for $k \in s$.

We assess the properties of the estimators, in terms of their simulation bias, variance and mean squared error. For an arbitrary population total estimator $\hat{Y}$, these were computed, with $K = 10\ 000$, as:

$$\text{Bias}_{\text{SIM}}(\hat{Y}) = \text{E}_{\text{SIM}}(\hat{Y}) - Y = \frac{1}{K}\sum_{j=1}^{K} \hat{Y}_j - Y$$

$$\text{Var}_{\text{SIM}}(\hat{Y}) = \frac{1}{K}\sum_{j=1}^{K} (\hat{Y}_j - \text{E}_{\text{SIM}}(\hat{Y}))^2$$

$$\text{MSE}_{\text{SIM}}(\hat{Y}) = \frac{1}{K}\sum_{j=1}^{K} (\hat{Y}_j - Y)^2 = \text{Var}_{\text{SIM}}(\hat{Y}) + (\text{Bias}_{\text{SIM}}(\hat{Y}))^2$$

In addition to the estimators mentionend earlier, we added, as a benchmark, the expansion (exp) estimator $\hat{Y} = N\bar{y}_r$, which would be used in a complete lack of auxiliary information. Formally, it results from a calibration with $\mathbf{x}_k = \mathbf{x}_k^* = 1, k \in U$ and $\mathbf{X} = N$. (As expected, the simulation shows this estimator to be inferior to the other choices.)

Please note that because bias can have either sign; the term ”smaller bias” (”greater bias”) should be interpreted in the following sections as ”smaller (greater) in absolute value”.

4.2 Results and discussion

The simulation confirms some of our conjectures in Section 3. A key factor for interpreting the results is that the quantitative star variable x_k is highly explanatory for y, whereas the moon vector indicating the four groups is not. The variance is the dominating component of the MSE, for essentially all of the estimators.

The decisive factor for the variance and the MSE is whether calibration on x_k takes place at the population level, as in A1dir, B1dir, A1BU and B1Bu, or at the sample level, as in A2dir, B2dir, A3BU and B2BU.

Table 3 (for variance) and Table 4 (for MSE) show much lower numbers for the former group of four estimators compared with the latter group of four. Much accuracy is lost here by not taking advantage of the available population total of x_k. The practice of transcribing the register information to the sample file only would not be recommendable here.

If we look within the high accuracy group A1dir, B1dir, A1BU and B1BU, we see little difference in variance but interesting differences in bias. The much practiced B1BU to adjust first for bias, then to calibrate only on the the population total has the highest bias and the highest MSE in this group of four. B1BU has marginally lower variance but not enough to offset its greater bias. Both of the direct appoaches, A1dir and B1dir, have lower MSE. The lowest MSE in this group of four is achieved by A1BU, showing that it is important to let the second calibration step incorporate the full auxiliary information, although part of it - the group vector part - was already used in the first step. It is recommended, as in A1BU, to make the repeated use of that information.

Among the top-down alternatives, only A1TD2 shows a MSE comparable with the best choices. As expected, the alternatives C1 and C2 do not perform well since they calibrated on weak information.

The empirical study confirms at least some of our conjectures in the preceding discussion. The variance is indeed much greater for the cases of $A2$ and $B2$ compared with $A1$, whereas there is a negligible difference between $A1$ and $B1$. This is not surprising since the moon indicator vector carries less information about the study variable than does the quantitative star variable. Furthermore, for $A1$, $A2$, $B1$, and $B2$ there is almost no difference with regard to variance between the direct and two-step procedures. Concerning bias, we get slightly higher bias for $B1$ and $B2$ compared with $A1$ and $A2$. The two-step approach produce estimators with slightly less bias for $A1$ and $A2$ compared with the direct approach and this tendency is reversed for $B1$ and $B2$. We note in particular that the two-step case B1BU of "adjusting first for the nonresponse, then calibrating to the population total" leads, for these data, to a higher bias (-3.03) than the direct calibration (-2.76). Although not large, the difference may suggest that the two-step approach should be used with some caution. Comparing the two-step approaches, there are small differences, with respect to both bias and variance, between A1BU and A1TD2 and between A2BU and A2TD2. Within the top-down approach, we may however observe that A3TD1 have much larger variance than A3TD2 and C2TD1 have much larger variance than C2TD2. The observations for variances are carried on to the MSE for the estimators, due to the fact that the variance dominates the bias. A trade-off of lower bias against higher variance for $B1$ compared with $A2$ can thus not be observed for this data set.

Table 2: BIAS: $\text{Bias}_{\text{SIM}}(\hat{Y})\text{x}10^{-4}$

		Estimators			
Auxiliary vector	Vector with constraint	Direct	Two-step		
			Bottom-up	Top-down1	Top-down2
A	A1	-2.44	-2.31	-	-2.41
A	A2	-2.26	-2.13	-	-2.27
A	A3	-	-	-2.25	-2.46
B	B1	-2.76	-3.03	-	-
B	B2	-2.59	-2.86	-	-
C	C1	5.96	-	-	5.84
C	C2	-	-	5.94	5.80
Exp		9.80			

Table 3: VARIANCE: $\text{Var}_{\text{SIM}}(\hat{Y})\text{x}10^{-9}$

Auxiliary vector	Vector with constraint	Estimators: Direct	Two-step: Bottom-up	Two-step: Top-down1	Two-step: Top-down2
A	A1	0.84	0.85	-	0.84
A	A2	5.14	5.15	-	5.22
A	A3	-	-	5.02	1.53
B	B1	0.83	0.82	-	-
B	B2	5.17	5.14	-	-
C	C1	5.92	-	-	2.14
C	C2	-	-	6.32	1.23
Exp		7.23			

Table 4: MSE: $\text{MSE}_{\text{SIM}}(\hat{Y})\text{x}10^{-9}$:

Auxiliary vector	Vector with constraint	Estimators: Direct	Two-step: Bottom-up	Two-step: Top-down1	Two-step: Top-down2
A	A1	1.44	1.38	-	1.43
A	A2	5.65	5.60	-	5.74
A	A3	-	-	5.53	2.14
B	B1	1.59	1.74	-	-
B	B2	5.85	5.95	-	-
C	C1	9.47	-	-	5.55
C	C2	-	-	9.85	4.59
Exp		16.84			

5. Closing remarks

Our motivation for this article was to show that calibration on two sources of auxiliary information, one at the population level, one at the sample level, gives rise to a number of possibilities to compute the calibrated weights. We have attempted to account systematically for all the possible cases (Table 1). We have shown, in an empirical illustration (Tables 2, 3 and 4) that the numerical differencies (in terms of bias, variance and MSE) between the cases are sometimes considerable. Depending on the practical situation, a user of calibration methodology will find some guidance in our results.

REFERENCES

Brick JM. Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics*. 2013; 29:329-353.

Brick JM, Jones ME. Propensity to respond and nonresponse bias. Metron–International *Jorunal of Statistics*. 2008; 66: 51-73.

Deville JC. Generalized calibration and application to weighting for nonresponse. In: Bethlehem JG, van der Heijden PGM, editors. Compstat, *Proceedings in Computational Statistics, 14th symposium, Utrecht, Heidelberg*; Physica Verlag; 1995. p. 65-76.

Deville JC, Särndal CE. Calibration estimators in survey sampling, *Journal of the American Statistical Association*. 1992; 87:376-382.

Kalton G. *Compensating for Missing Survey Data*. Ann Arbor: University of Michigan Press; 1983.

Kalton G, and Kasprzyk D. The treatment of missing data. *Survey Methodology*. 1986; 12:1-16.

Kott PS. Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*. 2006; 32:133-142.

Kott PS, Chang T. Calibration weighting to adjust for nonignorable unit nonresponse, *Journal of the American Statistical Association*. 2010; 105:1265-1275.

Kott PS, Liao D. Providing double protection for unit nonresponse with a nonlinear calibration-weighting routine. *Survey Research Methods*. 2012; 6:105-111.

Lundström S, Särndal CE. Calibration as standard method for treatment of nonresponse. *Journal of Official Statistics*. 1999; 15:305-327.

Särndal CE, Lundström S. *Estimation in Surveys with Nonresponse*. Chichester, UK: Wiley; 2005.