Generalised calibration for nonresponse and non-probability samples

Alina Matei
University of Neuchâtel, Switzerland

2 December 2025, Rome,
Fourth Workshop on Methodologies for Official Statistics
Organized by ISTAT

Contents

- Nonresponse framework.
- Generalised calibration for handling nonignorable nonresponse.
- Latent variables in connection with generalised calibration.
- Extension to non-probability samples.
- Conclusions.

1) Missing data and nonresponse bias

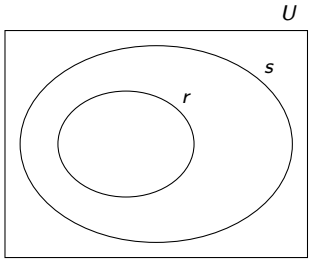
- Missing data arise in surveys under two forms:
  1. Unit nonresponse: a sampled unit does not respond to the entire survey;
  2. Item nonresponse: a sampled unit does not respond to a particular item (question).
- Nonresponse can create bias in estimates.
- In general, nonresponse bias is never completely eliminated, but one tries to reduce it as far as possible.
- Here we focus on unit nonresponse and bias reduction.
- Unit nonresponse also decreases the sample size, so the variance of the estimates can increase.

Framework

- Finite population with labels $U = \{1, 2, \ldots, k, \ldots, N\}$.
- Sample s (of size n, without replacement) selected from U using a sampling design $p(s)$
  - $\pi_k = P(k \in s) = \sum_{s; s \ni k} p(s)$: first-order inclusion probability of unit k in the sample;
  - $\pi_k > 0$ for all $k \in U$,
- y - variable of interest (y_k - value of y for the survey unit k)
- $\mathbf{x}_k = (x_{k1}, x_{k2}, \ldots, x_{kq})'$ - vector of auxiliary variables is assumed to be available for each survey unit k in the population or at least in the sample.

Framework

- Unit nonresponse:
  - $r \subset s$
    r: set of survey *respondents*
  - $\overline{r} = s \setminus r$
    $\overline{r}$: set of survey *nonrespondents*



- Two-phase approach: the set r is seen as a random sample selected from s using a sampling design $q(r|s)$.
- Suppose that y_i observed for $i \in r$ and it is missing for $i \in \overline{r}$.

Framework

- $R_i = 1$ if $i \in r$ and 0 otherwise (responding/not responding)
- R_i are independent Bernoulli variables (the respondents answer the survey independently one of others) with probabilities
- $p_i = P\left(R_i = 1 | i \in s\right)$
- $p_i > 0$ for all $i \in s$.

Nonignorable nonresponse:

Response probabilities

$$p_i = P(R_i = 1 | y_i, \mathbf{x}_i) \text{ or } p_i = P(R_i = 1 | y_i). \quad (1)$$

Examples: typical of, but not limited to, surveys with sensitive topics (income, drug abuse, politics, sexual attitudes, etc).

Estimation

- Goal: estimate $Y = \sum_{k \in U} y_k$ or $\overline{Y} = \sum_{k \in U} y_k / N$
- Without unit nonresponse: Y is estimated by the Horvitz-Thompson estimator: $\widehat{Y}_{HT} = \left(\sum_{k \in s} y_k / \pi_k\right)$. while $\overline{Y}$ can estimated by
  - the Horvitz-Thompson estimator: $\widehat{\overline{Y}}_{HT} = \left(\sum_{k \in s} y_k / \pi_k\right) / N$.
  - the Hájek estimator: $\widehat{\overline{Y}}_{Hajek} = \left(\sum_{k \in s} y_k / \pi_k\right) / \left(\sum_{k \in s} 1 / \pi_k\right)$.
- If $d_i = \frac{1}{\pi_i}$, unit nonresponse may be treated by reweighting (adjustment), using the following adjusted weights

$$\frac{d_i}{p_i}, \text{ with } p_i > 0, i \in r.$$

Estimators adjusted for nonresponse

- for the total: the Horvitz-Thompson type estimator adjusted for nonresponse:

$$\widehat{Y}_{HT,p} = \sum_{i \in r} \frac{y_i}{\pi_i p_i}.$$

- for the mean: the Hájek type estimator adjusted for nonresponse:

$$\widehat{\overline{Y}}_p = \left(\sum_{i \in r} \frac{y_i}{\pi_i p_i} \right) / \sum_{i \in r} \left(\frac{1}{\pi_i p_i} \right)$$

- In practice, the response probabilities p_i are never known; they are estimated.
- Central problem: estimate p_i by $\widehat{p}_i$ to compute

$$\sum_{i \in r} \frac{y_i}{\pi_i \widehat{p}_i} \text{ or } \sum_{i \in r} \left(\frac{y_i}{\pi_i \widehat{p}_i} \right) / \sum_{i \in r} \left(\frac{1}{\pi_i \widehat{p}_i} \right).$$

Calibration - without nonresponse

(Deville and Särndal, 1992)

- A set of weights w_k is constructed by modifying the initial weights $d_k = 1/\pi_k, k \in s$, such that a distance between d_k and the cal. weights is minimized, while satisfying the calibration equations

$$\sum_{k \in s} w_k \mathbf{x}_k = \sum_{k \in U} \mathbf{x}_k = \text{known}. \tag{2}$$

- One uses a function $F(.)$ with suitable properties, like $F(\boldsymbol{\lambda}^\top \mathbf{x}_k) = 1 + \boldsymbol{\lambda}^\top \mathbf{x}_k$ or $F(\boldsymbol{\lambda}^\top \mathbf{x}_k) = exp(\boldsymbol{\lambda}^\top \mathbf{x}_k)$, with $\boldsymbol{\lambda}$ a vector to determine by solving the calibration equations (2).
- The final weights are

$$w_k = d_k F(\widehat{\boldsymbol{\lambda}}^\top \mathbf{x}_k) = F(\widehat{\boldsymbol{\lambda}}^\top \mathbf{x}_k)/\pi_k.$$

- The calibration estimator of $\overline{Y}$ is

$$\widehat{\overline{Y}}_{\text{cal}} = \left(\sum_{k \in s} w_k y_k \right) / \left(\sum_{k \in s} w_k \right).$$

Calibration – MAR nonresponse

MAR: $p_i = P(R_i = 1|\mathbf{x}_i)$.
There are two different approaches (Särndal and Lundström, 2005):
1) two-steps approach:

1. estimate the response probabilities (e.g. logistic regression) using the model

$$p_i = \mathrm{E}(R_i|\mathbf{x}_i) = \frac{1}{1 + \exp\left[-\left(\alpha_0 + \mathbf{x}_i^\top \boldsymbol{\alpha}\right)\right]},$$

   and form the intermediate adjusted weights $d_i/\widehat{p}_i$;

2. obtain the final weights by calibration

$$w_i = d_i F(\widehat{\boldsymbol{\lambda}}^\top \mathbf{x}_i)/\widehat{p}_i$$

   such that the following **calibration equations** are satisfied:

$$\sum_{i \in r} w_i \mathbf{x}_i = \sum_{k \in U} \mathbf{x}_k \text{ or } \sum_{i \in r} w_i \mathbf{x}_i = \sum_{k \in s} d_k \mathbf{x}_k, \tag{3}$$

   assuming that $\sum_{k \in U} \mathbf{x}_k$ or $\sum_{k \in s} d_k \mathbf{x}_k$ are known.

2) one-step approach:

- obtain the final weights

$$w_i = d_i F(\widehat{\boldsymbol{\lambda}}^\top \mathbf{x}_i), i \in r$$

  such that the calibration equations (3) are satisfied.
- $F(\widehat{\boldsymbol{\lambda}}^\top \mathbf{x}_i)$ is an adjustment factor that it is implicitly an estimation of p_i^{-1}.
- The same variables $\mathbf{x}_i$ are used in the adjustment factor $F(.)$ and in the calibration equations.

In both approaches, the calibration estimator of $\overline{Y}$ is

$$\widehat{\overline{Y}}_{cal} = \left(\sum_{i \in r} w_i y_i\right) / \left(\sum_{i \in r} w_i\right).$$

Differences between the two approaches: Andersson and Särndal (2016).

Generalised calibration

Calibration (one-step approach)

$$w_i = d_i F(\widehat{\boldsymbol{\lambda}}^\top \mathbf{x}_i) \quad \rightarrow \quad \widehat{p}_i^{-1} = F(\widehat{\boldsymbol{\lambda}}^\top \mathbf{x}_i)$$

Generalised calibration

$$w_i = d_i F(\widehat{\boldsymbol{\lambda}}_g^\top \mathbf{z}_i) \quad \rightarrow \quad \widehat{p}_i^{-1} = F(\widehat{\boldsymbol{\lambda}}_g^\top \mathbf{z}_i)$$

- $\mathbf{x}$: q calibration variables
- $\mathbf{z}$: q 'model' variables

- calibration equations:

$$\sum_{i \in r} d_i F(\boldsymbol{\lambda}_g^\top \mathbf{z}_i)\mathbf{x}_i = \sum_{k \in U} \mathbf{x}_k$$

or

$$\sum_{i \in r} d_i F(\boldsymbol{\lambda}_g^\top \mathbf{z}_i)\mathbf{x}_i = \sum_{k \in s} d_k \mathbf{x}_k$$

- The generalised calibration estimator of $\overline{Y}$

$$\widehat{\overline{Y}}_{gcal} = \left(\sum_{i \in r} d_i F(\widehat{\boldsymbol{\lambda}}_g^\top \mathbf{z}_i) y_i\right) / \left(\sum_{i \in r} d_i F(\widehat{\boldsymbol{\lambda}}_g^\top \mathbf{z}_i)\right).$$

- *Remark* : There are two different points of view:
  1) $\mathbf{z} = q$ 'model' variables (Deville, 2000; Kott, 2006) - $F(\cdot)$ is a functional form, no model
  2) $\mathbf{z} = q$ model variables (Haziza and Beaumont, 2017) - a calibration equation is an estimating equation.

y as 'model' variable

- The values of $\mathbf{z}_i$ should be known only for respondents. Thus, it can simply be y_i. (Deville, 2000; Kott, 2006).
- $\mathbf{z}_i = (1, y_i)^T$ or $z_i = y_i, i \in r$. So, y is the variable of interest, but also a 'model variable' in the 'response model' (Kott, 2006).
- Thus, we can assume for instance $p_i = P(R_i = 1|y_i)$.

Features - 1

1. The choice of F in the one-step approach calibration and in generalised calibration is relevant in the presence of nonresponse because it influences significantly the properties of the final estimator (see also Haziza and Lesage, 2016).
2. Different forms of $F(\cdot)$ (e.g. $exp(\cdot), 1 + exp(\cdot)$) provide different estimators that are ***not asymptotically equivalent*** to the one based on linear form.
3. No explicit model is specified for the treatment of the nonresponse; it is given by the procedure: p_i is implicitly assumed to be $1/F(\boldsymbol{\lambda}^\top \mathbf{z}_i)$.
4. In practice, the functional form of $F(\cdot)$ is a question of 'guessing' since there is no way to determine its real form such that $F(\boldsymbol{\lambda}^\top \mathbf{z}_i) = p_i^{-1}$.

Features - 2

1. Haziza and Lesage (2016) showed by Monte Carlo simulation that an incorrect functional form of $F(\cdot)$ in one-step calibration may result in significant biases.
2. For generalised calibration, the problem of the misspecification of $F(\cdot)$ was not studied in detail.
3. The number of calibration variables should be equal to the number of 'model' variables. Chang and Kott (2008) and Kott and Liao (2017) provide solutions for cases where these numbers differ (see also Kott, 2014, for a review).

Variance estimation for a generalised calibration estimator

- reverse approach (Fay, 1991; Shao and Steel, 1999; Lesage et al., 2019).
- Kott (2006) also suggested using the jackknife method and a two-phase approach to estimate the variance.
- See also Kott and Chang (2010) and Ranalli et al. (2023).

Consistency of the generalised estimator $\widehat{Y}_{gcal}$ of the total

If F is correctly specified:

- Studied by Kott (2006) who assumes the following model

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, i \in U, \tag{4}$$

  with $E(\varepsilon_i|(\mathbf{x}_j, \mathbf{z}_j, R_j), j \in U) = 0$. If $\mathbf{z}_i$ contains y_i, $E(\varepsilon_i|(\mathbf{x}_j, \mathbf{z}_j, R_j), j \in U) \neq 0$.
- See also Kott and Chang (2010).
- Lesage et al. (2019) - asymptotic framework approach with random vectors $(R_i, \mathbf{X}_i^T, Y_i, \mathbf{Z}_i^T), i \in U$, with $E(R_i|\mathbf{Z}_i) = 1/F(\boldsymbol{\lambda}^T \mathbf{Z}_i)$. The proof is based on several assumptions, where the main one is:

$$Cov(\mathbf{X}_i, R_i|\mathbf{Z}_i) = 0 \text{ and } Cov(Y_i, R_i|\mathbf{Z}_i) = 0.$$

- Lesage et al. (2019) discuss cases in which the first condition is not met.
- If $\mathbf{Z}_i$ contains Y_i, the second condition is always fulfilled.

Remarks

If $z = y$, two important ingredients are required for the efficiency of the generalised calibration estimator:

1. a large $cor(x, y)$ for variance reduction,
2. independence between x and R conditional on y for bias reduction.

Examples - Monte Carlo simulation

- We estimate $\overline{Y}$.
- A census is used ($d_k = 1$, for all $k \in U$; 10,000 sets r are directly drawn from U using Poisson sampling with probabilities p_i), since we focus here on nonresponse.
- $N = 1000$ and 4 settings, with different correlation degrees between x and y :
  1. Setting 1: $x_i \sim N(2, 4^2), y_i = 0.3 + 0.3x_i + \varepsilon_i, \varepsilon_i \sim N(0, 1)$, iid,
  2. Setting 2: $x_i \sim N(2, 4^2), y_i = 0.3 + 0.1x_i + \varepsilon_i, \varepsilon_i \sim N(0, 1)$, iid,
  3. Setting 3: $y_i \sim N(2, 4^2), u \sim N(2, 4^2)$, independent
     $x_i = 0.3 + 0.1y_i + 0.3u_i + \varepsilon_i, \varepsilon_i \sim N(0, 1)$, iid,
  4. Setting 4: $x \sim N(2, 4^2), y_i \sim N(0, 1)$, independent,

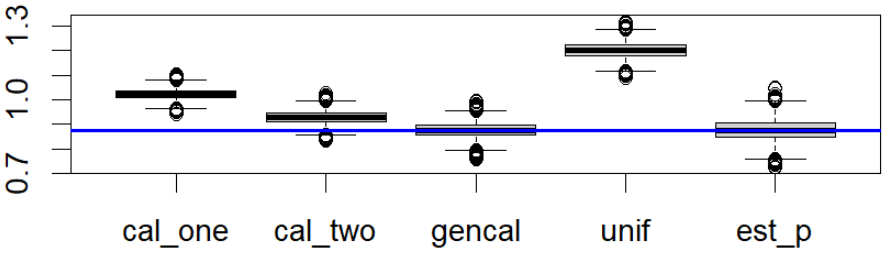
  with $i = 1, \ldots N$.

- In Settings 1, 2: $p_i = exp(0.5 + 0.5y_i)/(1 + exp(0.5 + 0.5y_i))$, $i = 1, \ldots N$.
- In Setting 3: $p_i = exp(0.5 + 0.5y_i + 0.1u_i)/(1 + exp(0.5 + 0.5y_i + 0.1u_i))$, $i = 1, \ldots N$.
- In Setting 4: $p_i \sim Unif(0, 1), i = 1, 2, \ldots, N$.
- We set $\mathbf{x}_i = (1, x_i)^T$ and $\mathbf{z}_i = (1, y_i)^T, i = 1, \ldots, N$.
- In each run, the estimated response probabilities have the form $exp(\widehat{\boldsymbol{\lambda}}^T \mathbf{z}_i)/(1 + exp(\widehat{\boldsymbol{\lambda}}^T \mathbf{z}_i))$ - only in Settings 1, 2, p_i are respected.

Setting 1 - a good case for gen. calib. estimator

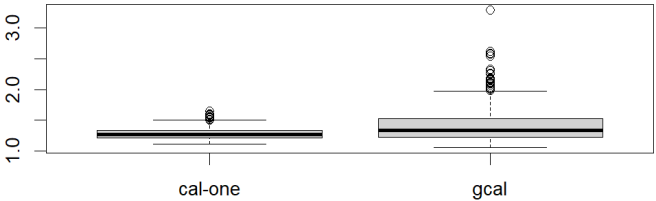
$cor(x, y) = 0.75$. **Conditions 1 and 2 fulfilled**.

5 Hajèk type estimators of $\overline{Y}$ are used: one-step cal., two-steps cal., gen. cal., uniform and the estimator using the true p. The blue line shows the value of $\overline{Y}$.



Comparison of weights in Setting 1

- Setting 1, cor(x,y)=0.75, $F(\cdot) = 1 + exp(\cdot)$, a single run

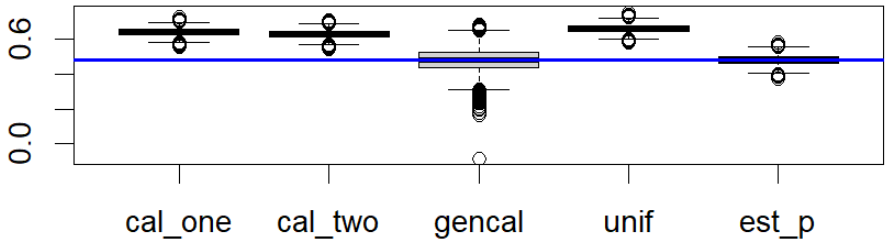


Larger variance of the gen. calib. estimator than that of the one-step calib., but smaller MSE.

Setting 2 - increase in variance

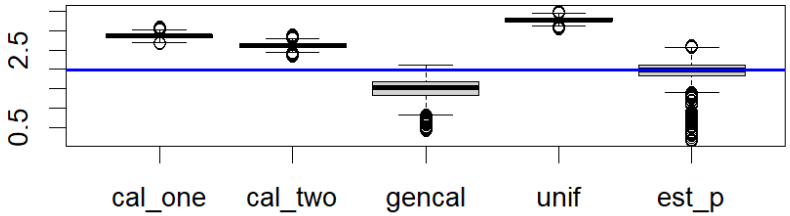
$cor(x, y) = 0.35$. **Condition 2 fulfilled**.

5 Hajèk type estimators of $\overline{Y}$ are used: one-step cal., two-steps cal., gen. cal., uniform and the estimator using the true p. The blue line shows the value of $\overline{Y}$.



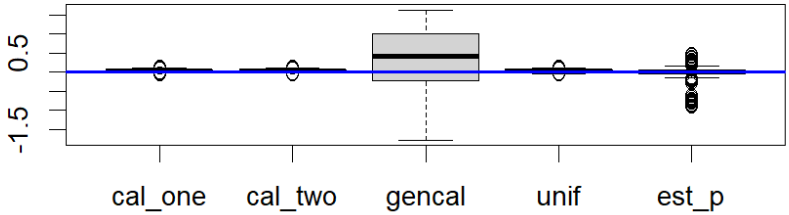
Setting 3 - bias amplification

This setting corresponds to the case where x and R are not independent given y because of u.
$cor(x, y) = 0.58, cor(x, u) = 0.62, cor(p, y) = 0.90, cor(p, u) = 0.19$.
The blue line shows the value of $\overline{Y}$.
5 Hajèk type estimators of $\overline{Y}$ are used: one-step cal., two-steps cal., gen. cal., uniform and the estimator using the true p. The blue line shows the value of $\overline{Y}$.



Setting 4 - worst case

$cor(x, y) \sim 0, cor(p, y) \sim 0, \overline{p} = 0.50$
The blue line shows the value of $\overline{Y}$.
5 Hajèk type estimators of $\overline{Y}$ are used: one-step cal., two-steps cal., gen. cal., uniform and the estimator using the true p. The blue line shows the value of $\overline{Y}$.



R package 'sampling'

Tillé and Matei (2025), version 2.11

- generalised calibration is implemented in the function gencalib()
- 4 methods:
  - linear: $F(\cdot) = 1 + (\cdot)$
  - raking: $F(\cdot) = exp(\cdot)$
    (for $F(\cdot) = 1 + exp(\cdot), p = exp(\cdot)/(1 + exp(\cdot))$ is obtained by calibrating on $\sum_{k \in U} \mathbf{x}_k - \sum_{i \in r} d_i \mathbf{x}_i$ instead of $\sum_{k \in U} \mathbf{x}_k$)
  - truncated: $F(\cdot) = 1 + (\cdot) \in (L, U)$
  - logit.

Generalized exponential adjustment

- method in gencalib(): logit
(Kott, 2006; Folsom and Singh, 2000)

$$p_i^{-1} = F(\boldsymbol{\lambda}^T \mathbf{z}_i) = \frac{L(U-C) + U(C-L)\exp(A\boldsymbol{\lambda}^T \mathbf{z}_i)}{(U-C) + (C-L)exp(A\boldsymbol{\lambda}^T \mathbf{z}_i)} \in (L, U),$$

where

- $A = (U-L)/((C-L)(U-C))$ and $1 \leq L < C < U \leq \infty$,
- if $C = 1$ we have $L < 1 < U$, like in Deville and Särndal (1992),
- for $L = 1, C = 2$ and $U = \infty$, $F(\boldsymbol{\lambda}^T \mathbf{z}_k)$ is close to $1 + \exp(\boldsymbol{\lambda}^T \mathbf{z}_i)$,
- for $C = 1, L = 0, U = \infty$, $F(\boldsymbol{\lambda}^T \mathbf{z}_i)$ is close to $\exp(\boldsymbol{\lambda}^T \mathbf{z}_i)$.
- The adjustment factor F is centered at the value C (not necessary 1), and the estimated response probabilities $1/F(.)$ are between $(1/U, 1/L)$.

2) Low response rate in prob. samples

- Based on the international experience (Lundström and Särndal, 2009), the nonrespondents are in general: metropolitan residents, single people, members of childless households, young people, divorced / widowed people, people with lower educational level, self-employed people, persons of foreign origin.
- Characteristics of respondents
  - General: educational level, age, gender, ethnicity or race, civil status, etc.;
  - Specific: personality traits, motivation, relevance of survey topic, attitudes, etc.

Does a propensity for response in surveys really exist?

"This is interesting, 70% of the respondents to our survey said they don't respond to surveys."

- Perhaps, but it is not observable: it is a latent variable. It is called *tendency to respond* to the survey (De Menezes and Bartholomew, 1996) or the *will to respond to the survey* (Matei and Ranalli, 2015). Moustaki and Knott (2000) note:

  "Weighting through latent variable modeling is expected to perform well under non-ignorable nonresponse where conditioning on observed covariates only is not enough."

Generalised calibration with latent variables

Ranalli, Matei and Neri (2023)

- The idea is to use latent variables as model variables $\mathbf{z}_i$ in generalised calibration.
- One can uses the 'will to respond to the survey' latent variable as $\mathbf{z}_i$ (this can be estimated by using an item response model, for all $i \in r$).
- Nonresponse may be related to the survey topic itself (see Groves et al., 2006) (e.g. surveys on attitudes towards politics). A latent variable called *attitude towards politics* can better explain the response probability of a unit to answer the survey questionnaire than any observed variable.

Italian Survey on Household Income and Wealth (SHIW)

Ranalli, Matei, Neri, 2023

- It is conducted by the Italian Central Bank every two years and collects detailed information on household income, wealth and consumption.
- The survey collects information on two topics that are generally considered highly sensitive: income and household wealth.
- y is the household wealth.
- 12 manifest variables with values known only for $i \in r$: the wealth class (ordinal variable with five levels), the number of total call attempts needed to make the interview (ranging between 1 and 4) and six dummy indicators for the ownership of a secondary dwelling, of bonds, of agricultural and of non-agricultural land, of other non-residential buildings and for the household living in a deluxe dwelling.
- 5 latent classes of response were created starting from the 12 manifest variables (that yield a **probabilistic clustering approach**) and employed as homogeneous response groups.
- The predicted latent class memberships (binary variables) were used as model variables $\mathbf{z}_i$ in generalised calibration.

Remark

- Usual assumption : $|\mathbf{x}| = |\mathbf{z}|$
- Chang and Kott (2008) - use a symmetric, positive definite matrix $W_{p\times p}$ and compute the weights $w_i = d_i F(\boldsymbol{\lambda}^T z_i)$ that minimize

$$\left(\sum_{k\in s} d_k \mathbf{x}_k - \sum_{i\in r} d_i F(\boldsymbol{\lambda}^T \mathbf{z}_i)\mathbf{x}_i\right)^T W \left(\sum_{k\in s} d_k \mathbf{x}_k - \sum_{k\in r} d_i F(\boldsymbol{\lambda}^T \mathbf{z}_i)\mathbf{x}_i\right),$$

  where $(\mathbf{x}_i)_{p\times 1}, (\mathbf{z}_i)_{q\times 1}$, and $p > q$.
- Different choices for W are discussed by Chang and Kott (2008).
- The method allows to use $|\mathbf{x}| > |\mathbf{z}|$ but the calibration equations are not necessary satisfied.

Double calibration

- Calibration variables: age (5 classes), gender, education (3 levels), nationality (italian/foreigner), job status (employed/unemployed/inactive), geographical area (north/centre/south), etc.
- Less number of model variables.
- We applied a double calibration:
  1. Use the method of Chang and Kott (2008) and provide intermediate weights $d_i F(\widehat{\boldsymbol{\lambda}}_1^T \mathbf{z}_i), i \in r$
  2. Use calibration with all $\mathbf{x}$ variables (to fulfill the calibration equations) and obtain the final weights as:

$$w_i = d_i F(\widehat{\boldsymbol{\lambda}}_1^T \mathbf{z}_i) F(\widehat{\boldsymbol{\lambda}}_2^T \mathbf{x}_i), i \in r,$$

Estimated mean of the households' wealth using different estimators, ratio to the macro estimate of the mean (368,840 euros), estimated jackknife standard error, coefficient of variation (in %), standard deviation of the set of calibrated weights

Method	Estimated Mean	Ratio to macro est.	Jackknife St. Err.	%CV	Weights St. Dev.
(1) Hajek estimator	252,407.35	0.684	7,780.42	3.08	2,465.59
(2) Two–phase estimator	296,012.21	0.803	8,862.29	2.99	3,349.27
(3) Classical calibration	282,378.61	0.766	6,732.86	2.38	2,550.33
Two–step generalised calibration:					
(4) Model variables – latent classes	307,762.18	0.834	7,843.80	2.55	2,697.16
(5) Model variables – manifest variables	329,457.08	0.893	26,458.18	8.03	8,306.05
(6) Model variables – classes of y	319,804.76	0.867	8,393.59	2.62	3,075.03

(1) Hajek estimator (without nonresponse adjustment); (2) two-phase estimator - response probabilities estimated via a logistic model that uses covariates known also for nonrespondents.

3) Non-probability samples

- non-probability sample $\subseteq U$ as follows: the selection of the sample is not random, but the decision of unit i to answer the survey questionnaire is the outcome of a random experiment.
- That, there exists a Bernoulli variable I_i that takes value 1 if unit i answers the survey questionnaire with an unknown probability of participation $p_{np,i}$ and 0, otherwise.
- Next, $s_{np} = \{I_i = 1, i \in U\}$ and the variable y is measured for each $i \in s_{np}$.

Two main assumptions are usually made :

1. $p_{np,i}$ does not depend on $y_i, i \in U$,
2. $p_{np,i} > 0, \forall i \in U$.

Other assumptions are added depending on the method used.
Few papers consider the first assumption not fulfilled: Beaumont (2020); Kim and Morikawa (2023); Pfeffermann and Sverchkov (2025); Burakauskaitė and Čiginas (2025).

Undercoverage

- If Assumption 2 ($p_{np,i} > 0, \forall i \in U$) is not fulfilled.
- Let $U_{np} \subseteq U$.
- Wu (2022): U_{np} can be a deterministic (units with certain features will never be included in the non-probability sample; for instance, internet access) or stochastic (a random sample from U defined by units i with $p_{np,i} > 0$).
- Chen et al. (2023) - use a 'measure of accessibility' depending on the participation prob. (based only on x_i) to define a such U_{np} (deterministic case).

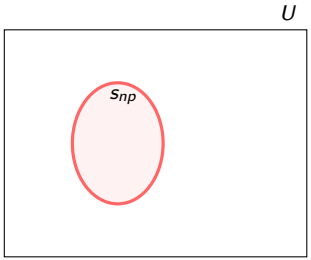
Undercoverage and latent random mechanism

- Savitsky et al. (2023) - assume that $s_{np} \subseteq U_{np}$ is governed by an underlying, latent random mechanism with unknown participation probabilities depending only on x; application in official statistics - quota sample of government employment collected in the Current Employment Statistics, U.S. Bureau of Labor Statistics.
- The definition of U_{np} may be related to the existence of a latent variable 'willing to respond to non-probability samples', depending both on x and y.
- For instance, willingness to participate is often driven by interest in topic, survey mode (web, app, phone), past experience with surveys, incentives to participate in surveys, and personal traits (e.g. personality, civic engagement), but also on x=age.

Generalised calibration - extension to non-probability samples

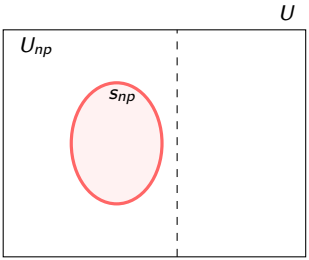
s_{np} = non-probability sample

Case 1: non-probability sample framework



$s_{np} \subseteq U$

Case 2: non-probability sample framework

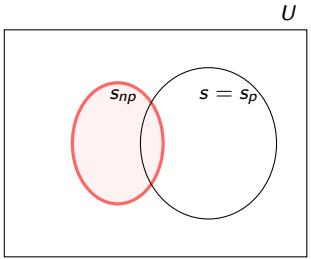


$s_{np} \subseteq U_{np} \subseteq U$

Generalised calibration - extension to non-probability samples

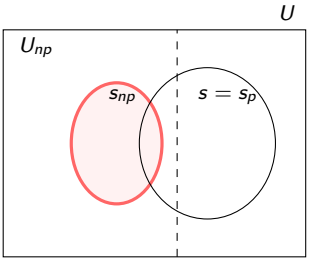
s_{np} = non-probability sample, s_p = probability sample

Case 3: non-probability sample framework



$s_{np} \subseteq U$

Case 4: non-probability sample framework



$s_{np} \subseteq U_{np} \subseteq U$

Context - non-probability sample

- **Case 3**: Beaumont (2020) advocates the use of the generalised calibration, when $p_{np,i}$ depends on $y_i, i \in U$, and discusses conditions of its application, such as the independence of x and the participation indicator after conditioning on y.
- **Case 4**: U_{np} is a domain of U, which is usually unknown and $s_{np} \subseteq U_{np}$.

Assumptions

For Case 4: we propose using the generalised calibration to estimate $p_{np,i}$ under the following assumptions

- A1 The probability $p_{np,i}$ of $I_i = 1$ depends on $y_i, i \in U_{np}$.
- A2 All the units $i \in U_{np}$ have $p_{np,i} > 0$.
- A3 The indicator variables $I_i, i \in U_{np}$ are independent given y_i.
- A4 There is a probability sample $s_p \subseteq U$, with unknown $y_i, i \in s_p$.
- A5 A set of auxiliary variables $\mathbf{x}_i$ is available and a linear regression model m with $E_m(y_i|x_i) = \mathbf{x}_i^T \beta$ holds for $i \in U$.
- A6 $\mathbf{x}$ and the participation indicator I are independent conditional on y.

For Case 3: $U_{np} = U$.

Generalised calibration

The following calibration equations given by Beaumont (2020); Chen et al. (2020) are used to estimate $p_{np,i}$:

$$\sum_{i \in s_{np}} \frac{\mathbf{x}_i}{p_{np,i}} = \sum_{j \in s_p} \frac{\mathbf{x}_j}{\pi_j}. \tag{5}$$

The generalised calibration method provides $\hat{p}_{np,i}$ using (5). Next, the generalised calibration of the total $Y = \sum_{k \in U} y_k$ is

$$\widehat{Y}_{gcal} = \sum_{i \in s_{np}} w_i^g y_i \tag{6}$$

and of the mean $\overline{Y} = \sum_{k \in U} y_k / N$

$$\widehat{\overline{Y}}_{gcal} = \sum_{i \in s_{np}} w_i^g y_i / \sum_{i \in s_{np}} w_i^g, \quad w_i^g = 1/\hat{p}_{np,i}. \tag{7}$$

Remarks

- **Problems**: selection bias and undercoverage.
- While the two problems are not eliminated, we try to reduce their impact using the generalised calibration compared to those of the naive and one-step calibration estimators under certain conditions similar to the non-response case and a relatively large size of U_{np}.

Some results

- $$p_{np,i} = \frac{exp(\lambda^T \mathbf{z}_i)}{1 + exp(\lambda^T \mathbf{z}_i)}, \text{ with } \mathbf{z}_i = (1, y_i)^T. \quad (8)$$
- We estimate $\overline{Y} = \sum_{k \in U} y_k / N$ using
$$\widehat{\overline{Y}}_{gcal} = \sum_{i \in s_{np}} w_i^g y_i / \sum_{i \in s_{np}} w_i^g, \quad w_i^g = 1/\hat{p}_{np,i}. \quad (9)$$
- $\widehat{\overline{Y}}_{gencal}$ is compared to the naive estimator $\sum_{i \in s_{np}} y_i / n_{np}$, with $n_{np} = |s_{np}|$ and the one-step calibration estimator of $\overline{Y}$.
$$\widehat{\overline{Y}}_{cal} = \sum_{i \in s_{np}} w_i^c y_i / \sum_{i \in s_{np}} w_i^c. \quad (10)$$

Settings

- A total of 50,000 runs were used, $N = 1000, \boldsymbol{\lambda} = (0, 5, 0.5)^T$.
- Two simulation settings.
- **Setting 1**: $x_i \sim N(2, 4^2), y_i = 0.3 + 0.3x_i + \varepsilon_i, \varepsilon_i \sim N(0, 1)$ iid, $i = 1, 2, \ldots, N$, with $cor(x, y) = 0.75$ and $cor(p_{np}, y) = 0.98$.
- **Setting 2**: x_i is the same as in Setting 1, $y_i = 0.3 + 0.1x_i + \varepsilon_i, \varepsilon_i \sim N(0, 1), i = 1, 2, \ldots, n$ iid with a $cor(x, y) = 0.35$ and $cor(p_{np}, y) = 0.68$.
- **Case 3**: $U_{np} = U$.
- **Case 4**: in each run, $U_{np} \subset U$ is stochastic, with two options.
  Option 1: srswor
  Option 2: it is seen as the basis of a latent process for selecting a non-probability sample: a systematic sample with inclusion probabilities proportional to the first principal component of x and y, and size $N_{np} = 500$.
- In each run, s_p is a systematic sample with inclusion probabilities proportional to $|x|$ and size 400 selected from U, while s_{np} is selected from U_{np} using Poisson sampling with probabilities $p_{np,i}$.

Setting 1: $cor(x, y) = 0.75, cor(p_{np}, y) = 0.98, \overline{p} = 0.70$

Estimator	|RB| (%)	RMSE
$U_{np} = U$		
Naive	37.083	0.3260
Calibration	17.785	0.1647
Gen. calibration	1.849	0.0957
Option I, $N_{np} = 500$		
Naive	47.356	0.4167
Calibration	23.986	0.2192
Gen. calibration	4.418	0.1156
Option II, $N_{np} = 500$		
Naive	85.92	0.7529
Calibration	34.45	0.3074
Gen. calibration	17.89	0.2105

Setting 2: $cor(x, y) = 0.35, cor(p_{np}, y) = 0.68, \overline{p} = 0.90$

Estimator	|RB| (%)	RMSE
$U_{np} = U$		
Naive	24.240	0.1205
Calibration	21.290	0.1067
Gen. calibration	3.520	0.1023
Option I, $N_{np} = 500$		
Naive	34.370	0.1654
Calibration	31.040	0.1499
Gen. calibration	7.060	0.1514
Option II, $N_{np} = 500$		
Naive	74.810	0.3592
Calibration	61.490	0.2956
Gen. calibration	59.120	0.3142

Conclusions

- The generalised calibration method is useful for dealing with non-ignorable non-response at the unit level, when the response probabilities depend on the variable of interest or variables known only for respondents.
- Useful for surveys with sensitive topics: income, household wealth, etc.
- Conditions if $z = y$: $cor(x, y)$ sufficiently high and independence of x and R conditional on y.
- If the first condition is not met, the variance increases. If the second condition is not met, the bias increases. For a discussion, see also Haziza and Beaumont (2017); Lesage et al. (2019).
- Caution should be exercised when applying the generalised calibration. It should always be compared with the one-step calibration estimator.

Conclusions

- Extension to handle non-probability samples in combination with a probability sample, when a linear model between y and x holds in U.
- Different assumptions can be used, which differ from those usually employed.
- The generalised calibration estimator of the population mean may outperform the calibration estimator and naive estimators in terms of the absolute value of the relative bias and mean square error, provided the following conditions are met:
  1. the participation probabilities dependent on y;
  2. a good level of correlation between x and y;
  3. independence of x and I conditional on y;
  4. a sufficient size of the subpopulation U_{np} (which should not be too small compared to the population size);
  5. the mechanism by which U_{np} arises from U is not solely dependent on x. This is reasonable given our assumption that participation probabilities depend on y.
- The generalised calibration deserves to be studied in greater detail because it offers interesting perspectives for non-probability samples.

Andersson, P. G. and Särndal, C.-E. (2016). Calibration for nonresponse treatment: In one or two steps? *Statistical Journal of the IAOS*, 32:375–381.

Beaumont, J. F. (2020). Are probability surveys bound to disappear for the production of official statistics? *Survey Methodology*, 46:1–28.

Burakauskaitė, I. and Čiginas, A. (2025). Non-probability and probability sample integration through combined estimators. In *Proceedings of the Baltic Nordic Ukrainian Workshop on Survey Statistics Theory and Methodology 2025, Vilnius, Lithuania*.

Chang, T. and Kott, P. S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika*, 95:555–571.

Chen, Y., Li, P., and Wu, C. (2020). Doubly robust inference with nonprobability survey samples. *Journal of the American Statistical Association*, 115:2011–2021.

Chen, Y., Li, P., and Wu, C. (2023). Dealing with undercoverage for non-probability survey samples. *Survey Methodology*, 49(2):499–513.

De Menezes, L. M. and Bartholomew, D. J. (1996). New developments in latent structure analysis applied to social attitudes. *Journal of Royal Statistical Society A*, 159:213–224.

Deville, J.-C. (2000). Generalized calibration and application to weighting for non-response. In *Compstat - Proceedings in Computational Statistics: 14th Symposium Held in Utrecht, The Netherlands*, pages 65–76, New York. Springer.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382.

Fay, R. E. (1991). A design-based perspective on missing data variance. In *Proceedings of the 1991 Annual Research Conference*, pages 429–440. US Bureau of the Census.

Folsom, R. E. and Singh, A. C. (2000). The generalized exponential model for design weight calibration for extreme values, nonresponse and poststratification. In *Section on Survey Research Methods, American Statistical Association, 598–603*.

Groves, R. M., Couper, M., Presser, S., Singer, E., Tourangeau, R., Acosta, G. P., and Nelson, L. (2006). Experiments in producing nonresponse bias. *Public Opinion Quarterly*, 70(5):720–736.

Haziza, D. and Beaumont, J. F. (2017). Construction of weights in surveys: A review. *Statistical Science*, 32(2):206–226.

Haziza, D. and Lesage, E. (2016). A discussion of weighting procedures for unit nonresponse. *Journal of Official Statistics*, 32(1):129–145.

Kim, J. and Morikawa, K. (2023). An empirical likelihood approach to reduce selection bias in voluntary samples. *Calcutta Statistical Association Bulletin*, 75:8–27.

Kott, P. S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, 32:133–142.

Kott, P. S. (2014). Calibration weighting when model and calibration variables can differ. In Mecatti, F., Conti, P. L., and Ranalli, M. G., editors, *Contributions to Sampling Statistics*, pages 1–18. Springer, Berlin.

Kott, P. S. and Chang, T. (2010). Using calibration weighting to adjust for nonignorable unit nonresponse. *Journal of the American Statistical Association*, 105:1265–1275.

Kott, P. S. and Liao, D. (2017). Calibration weighting for nonresponse that is not missing at random: allowing more calibration than response-model variables. *Journal of Survey Statistics and Methodology*, 5:159–174.

Lesage, E., Haziza, D., and D'Haultfoeuille, X. (2019). A cautionary tale on instrumental calibration for the treatment of nonignorable unit nonresponse in surveys. *Journal of the American Statistical Association*, 114(526):906–915.

Lundström, S. and Särndal, C.-E. (2009). Calibration of weights in surveys with nonresponse and frame imperfections, course presented at Eustat, Bilbao.

Matei, A. and Ranalli, M. G. (2015). Dealing with non-ignorable nonresponse in survey sampling: a latent modeling approach. *Survey Methodology*, 41 (1):1–19.

Moustaki, I. and Knott, M. (2000). Weighting for item non-response in attitude scales using latent variable models with covariates. *Journal of Royal Statistical Society, Series A*, 163:445–459.

Pfeffermann, D. and Sverchkov, M. (2025). Use of nonprobability samples for official statistics, state of the art. *Survey Methodology*, 51(1):169–196.

Ranalli, M. G., Matei, A., and Neri, A. (2023). Generalised calibration with latent variables for the treatment of unit nonresponse in sample surveys. *Statistical Methods & Applications*, 32(1):169–195.

Särndal, C.-E. and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. Wiley, New York.

Savitsky, T. D., Williams, M. R., Gershunskaya, J., and Beresovsky, V. (2023). Methods for combining probability and nonprobability samples under unknown overlaps. *Statistics in transition*, 24(5):1–34.

Shao, J. and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonneglible sampling fractions. *Journal of the American Statistical Association*, 94:254–265.

Tillé, Y. and Matei, A. (2025). *sampling: Survey Sampling*. R package version 2.11.

Wu, C. (2022). Statistical inference with non-probability survey samples. *Survey Methodology*, 48(2):283–311.