

Generalised calibration with latent variables for the treatment of unit nonresponse in sample surveys

M. Giovanna Ranalli* Alina Matei† Andrea Neri‡

(accepted in Statistical Methods and Applications)

Abstract

Sample surveys may suffer from nonignorable unit nonresponse. This happens when the decision of whether or not to participate in the survey is correlated with variables of interest; in such a case, nonresponse produces biased estimates for parameters related to those variables, even after adjustments that account for auxiliary information. This paper presents a method to deal with nonignorable unit nonresponse that uses generalised calibration and latent variable modelling. Generalised calibration enables to model unit nonresponse using a set of auxiliary variables (instrumental or model variables), that can be different from those used in the calibration constraints (calibration variables). We propose to use latent variables to estimate the probability to participate in the survey and to construct a reweighting system incorporating such latent variables. The proposed methodology is illustrated, its properties discussed and tested on two simulation studies. Finally, it is applied to adjust estimates of the finite population mean wealth from the Italian Survey of Household Income and Wealth.

Keywords: auxiliary information, finite population, latent class models, latent trait models, nonignorable nonresponse.

1 Introduction

Unit nonresponse occurs when sample selected units do not participate in a survey. When the decision whether or not to participate depends on one or more variables

*Dept. of Political Science, University of Perugia, Italy giovanna.ranalli@unipg.it

†Institute of Statistics, University of Neuchâtel and Institute of pedagogical research and documentation, Neuchâtel, Switzerland, alina.matei@unine.ch

‡Central Bank of Italy, Rome, Italy, andrea.neri@bancaditalia.it

of interest, the missing data are said to be missing not at random. It is also referred to as nonignorable nonresponse because the missing data cannot be ignored when estimating finite population totals or means of survey variables in order to get consistent estimators (Deville, 2000; Kott and Chang, 2010).

In order to adjust for nonresponse, it would be essential to have access to granular auxiliary information on nonrespondents. Unfortunately, this is rarely the case. Generally, auxiliary information takes the form of a set of auxiliary variables whose value is known for the respondents and whose population totals are known or can be unbiasedly estimated. In this context, calibration weighting methods are often used to include auxiliary information at the estimation stage (Deville and Särndal, 1992; Särndal, 2007) and have been shown to be very useful in treating unit nonresponse as well (see e.g. Särndal and Lundström, 2005). Their goal is the construction of a single set of weights to be used for all variables of interest modifying specified initial weights (usually the basic design weights), while satisfying benchmark constraints on known auxiliary information. No explicit model is specified for the treatment of the nonresponse mechanism; it is implicitly given by the calibration procedure.

Extending the calibration procedure, instrument vector method or generalised calibration allows distinction to be drawn among auxiliary variables between those that are useful to model unit nonresponse (instrumental or model variables) and those that are used in the calibration constraints (calibration variables). See Kott (2006), Chang and Kott (2008), and see Kott (2014) for a review on calibration weighting when model and calibration variables differ. It is important to note that in generalised calibration model variables need only be known for the respondents. This offers a particularly useful tool when the response probability of the units depends on one or more survey variables of interest and/or on other variables that we observe only on the respondents. Within this framework, the only proposal available in the literature to address unit nonresponse consists in using directly the variable of interest as an instrument in generalised calibration. Lesage et al. (2019) discuss the properties of the final estimator and the conditions under which it is successful in reducing nonresponse bias. This approach has at least two main pitfalls. First, nonresponse is likely to depend not just upon the variable of interest. This is particularly the case with multipurpose surveys. A second pitfall is due to the fact that observed variables may be affected by measurement error. Failing to consider such an issue may result in biased estimators.

In this paper we propose to use latent variables as instruments in generalised calibration for the treatment of nonresponse. We believe that response to a survey

is the outcome of a complex process that involves several aspects which cannot be captured by a single variable. Latent variable models can be used to extract either continuous constructs (latent trait models) or categorical ones (latent class models) from a set of manifest variables collected from respondents. Latent variable modelling has already been studied in the context of unit nonresponse. Matei and Ranalli (2015) discuss a setting in which a latent variable to be used as a covariate in a classical logistic model for response indicators is estimated from the set of response indicators in the presence of item nonresponse. Here, we present a more general framework to use latent variables in the context of calibration. In addition, we are able to test our approach on a real survey, the Survey of Household Income and Wealth conducted by the National Central Bank of Italy (see for instance Banca d'Italia, 2018). This application is relevant since the survey collects very sensitive information and it is affected, like other similar surveys, by nonignorable unit nonresponse. In particular, rich households are known to be difficult to enroll in a sample survey. We use the 2008 wave for which we have also auxiliary information that enables us to enhance the performance of our estimator. To the best of our knowledge, this is the first application to real survey data of generalised calibration to deal with the issue of nonignorable nonresponse.

Notation and a review of generalised calibration as a treatment of unit nonresponse is provided in Section 2. A brief overview of the latent variable models used in the paper is provided in Section 3. The proposed methodology is introduced in Section 4. We test the methodology using two simulation studies (Section 5) and finally apply it to adjust estimates from the Italian Survey of Household Income and Wealth (Section 6). Some concluding remarks are provided in Section 7.

2 Calibration and generalised calibration for unit nonresponse

We are interested in estimating finite population quantities of interest (like totals, means and functions thereof). The finite population is considered fixed, and the source of randomness (other than the nonresponse mechanism) is the sampling design. In this sense, our approach is mainly design based. Below we introduce the framework of this approach.

2.1 Framework

Let $U = \{1, \dots, k, \dots, N\}$ be the set of labels identifying the units of a finite population of interest and let $\mathbf{y}_k = (y_{1k}, \dots, y_{mk}, \dots, y_{Mk})$ be the value taken on unit $k \in U$ by an M -vector of variables of interest \mathbf{y} . We are interested in estimating its population total $\mathbf{t}_y = \sum_U \mathbf{y}_k$. We will use the shorthand \sum_A for $\sum_{k \in A}$, with $A \subseteq U$ an arbitrary set. To this end, a sample s of fixed size n is selected from U using a sampling design $p(s)$ with first and second order inclusion probabilities $\pi_k = P(k \in s)$ and $\pi_{kl} = P(k, l \in s)$, respectively, for k and $l \in U$. Let $I_k = 1$ be the indicator variable for unit k selected in the sample, so that $P(I_k = 1) = \pi_k = E\{I_k | \mathcal{F}\}$, where $\mathcal{F} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ is the finite population. Let $d_k = \pi_k^{-1}$ and $d_{kl} = \pi_{kl}^{-1}$, for k and $l \in U$. Assuming $\pi_k > 0$, for all $k \in U$, the Horvitz-Thompson estimator (Horvitz and Thompson, 1952) of \mathbf{t}_y is

$$\hat{\mathbf{t}}_y^{\text{HT}} = \sum_s d_k \mathbf{y}_k$$

and has variance

$$\mathbf{V}\{\hat{\mathbf{t}}_y^{\text{HT}} | \mathcal{F}\} = \sum_U \sum_U \Delta_{kl} \mathbf{y}_k^T \mathbf{y}_l,$$

with $\Delta_{kl} = \pi_k^{-1} \pi_l^{-1} (\pi_{kl} - \pi_k \pi_l) = d_{kl}^{-1} (d_k d_l - d_{kl})$.

We denote by $r \subseteq s$ the set of survey respondents and let the response indicator $R_k = 1$ if unit $k \in r$ and 0 otherwise. Thus $r = \{k \in s | R_k = 1\}$ and the response mechanism is given by the probability distribution $q(r|s)$. The random indicator variables R_k are assumed independent of one another and of the sample selection mechanism. Then, the probability of responding of unit k to the survey is given by

$$p_k = P(R_k = 1 | k \in s).$$

This is the two-phase approach where the sampling design is the first phase and the response mechanism is the second phase; it is based on the quasi-randomisation model of Oh and Scheuren (1983) (see also Särndal et al., 1992, Chapter 9). If the p_k 's were known, the double expansion estimator

$$\hat{\mathbf{t}}_y^{\text{2E}} = \sum_r d_k p_k^{-1} \mathbf{y}_k$$

would be unbiased for \mathbf{t}_y . In practice, a two-step procedure is applied in which the response probabilities p_k are estimated assuming a response model and using data from the sample, and then applied to modify the basic design weight, $d_k, k \in s$. The prefix ‘‘quasi’’ is added to emphasise that inference depends not only on the design, but also on the assumed response model.

2.2 Calibration

Let \mathbf{x} be a set of auxiliary variables whose value $\mathbf{x}_k = (x_{1k}, \dots, x_{qk})$ is known for $k \in r$ and whose population total \mathbf{t}_x is known or can be unbiasedly estimated with the Horvitz-Thompson estimator $\widehat{\mathbf{t}}_x^{\text{HT}} = \sum_s d_k \mathbf{x}_k$. Let \mathbf{t}_x^* denote the population total of \mathbf{x} or its Horvitz-Thompson estimator accordingly. As opposed to the two-step approach, calibration pursues the construction of a set of weights w_k^c to be used for all variables of interest using a single-step approach. In particular, it modifies specified initial weights (usually the d_k 's), while satisfying benchmark constraints on known auxiliary information, i.e.

$$\sum_r w_k^c \mathbf{x}_k = \mathbf{t}_x^*. \quad (1)$$

No explicit model is specified for the treatment of the nonresponse mechanism; it is implicitly given by the calibration procedure. In fact, in the calibration approach the initial weights d_k are adjusted by a factor $F(\mathbf{x}_k \boldsymbol{\gamma}_c)$ that depends on the auxiliary information and using a differentiable known function $F(\cdot)$. The final set of weights is given by

$$w_k^c = d_k F(\mathbf{x}_k \widehat{\boldsymbol{\gamma}}_c),$$

where $\boldsymbol{\gamma}_c$ is estimated using the constraints given in (1). In particular, $\widehat{\boldsymbol{\gamma}}_c$ is the value of $\boldsymbol{\gamma}_c$ for which

$$\sum_r d_k F(\mathbf{x}_k \boldsymbol{\gamma}_c) \mathbf{x}_k - \mathbf{t}_x^* = \mathbf{0}.$$

Newton-Raphson methods can be used to this end. See e.g. Kott (2006, Section 4). Then, the calibration estimator of \mathbf{t}_y is

$$\widehat{\mathbf{t}}_y^{\text{CAL}} = \sum_r w_k^c y_k = \sum_r d_k F(\mathbf{x}_k \widehat{\boldsymbol{\gamma}}_c) y_k.$$

Fuller et al. (1994) and Särndal and Lundström (2005) consider “linear” calibration, i.e. $F(\mathbf{x}_k \boldsymbol{\gamma}_c) = 1 + \mathbf{x}_k \boldsymbol{\gamma}_c$, while Folsom (1991), Folsom and Singh (2000) and Kott (2006) consider “nonlinear” calibration, in which $F(\cdot)$ is a nonlinear function as, e.g., for raking ratio $F(\mathbf{x}_k \boldsymbol{\gamma}_c) = \exp(\mathbf{x}_k \boldsymbol{\gamma}_c)$.

The function $F(\cdot)$ plays an important role in calibration for nonresponse. Different specifications of the function $F(\cdot)$ have been suggested since the introduction of calibration in the paper by Deville and Särndal (1992). Note that, in the case in which nonresponse is not present, the effect on the final estimate of these alternative choices is limited. In fact, in that case $\widehat{\boldsymbol{\gamma}}_c$ converges in probability to 0, when $n \rightarrow \infty$, $N \rightarrow \infty$, and $F(0) = F'(0) = 1$, where $F'(\cdot)$ is the first derivative of $F(\cdot)$,

and $F(u) \approx 1 + u$, i.e. all choices of $F(\cdot)$ provide estimators that are asymptotically equivalent to the one based on linear calibration (Deville and Särndal, 1992). In the case of nonresponse this is not the case anymore, since $\hat{\gamma}_c$ does not converge to 0 (see also Haziza and Lesage, 2016).

The choice of $F(\cdot)$ is very relevant in the presence of nonresponse also because it influences significantly the properties of the final estimator. In fact, the calibration weight adjustment implicitly estimates the inverse of the probability of response, as p_k is implicitly assumed to be the inverse of $F(\cdot)$. Calibration provides *double protection* against nonresponse bias in the sense of Kim and Park (2006). That is, if either a linear prediction model or the implicit selection model holds, then the resulting estimator has a negligible bias. Fuller et al. (1994) noted this property for the case of linear calibration: the estimator is nearly unbiased under the joint response-sampling mechanism that treats response as a second phase of random sampling and as long as the probability of response is such that $p_k = (1 + \mathbf{x}_k \boldsymbol{\gamma}_c)^{-1}$. Kott (1994) who introduces the double protection in a survey setting (see also Haziza and Rao, 2006), also notes this equivalence for the regression estimator. Therefore, linear calibration can provide unrealistic values for such implicitly estimated probabilities, since weighting adjustment allows the implicit estimated selection probability to be less than 0 or greater than 1. For this reason nonlinear calibration based on generalisations of raking is usually employed. See Kott and Liao (2012) for a discussion of the properties of alternative functions $F(\cdot)$. In addition, Haziza and Lesage (2016) provide a thorough discussion and simulation results of alternative choices of such functions when using calibration in nonresponse.

2.3 Generalised calibration

As discussed so far, no discrimination is made within the set of auxiliary variables available: a single set of variables is employed at the same time for nonresponse treatment, sampling error reduction and coherence among estimates. There are cases, however, in which one has a good reason to believe that nonresponse depends on a set of *p model* variables \mathbf{z} , whose components need not coincide with the components of the *calibration* vector \mathbf{x} . This can be accommodated in the calibration framework using “instrument vector” or “generalised” calibration introduced by Deville (2000) (see also Kott, 2006) in which weights are given by

$$w_k = d_k F(\mathbf{z}_k \hat{\boldsymbol{\gamma}}),$$

where $\hat{\boldsymbol{\gamma}}$ is an estimator of a set of parameters $\boldsymbol{\gamma}$ such that, again, $\sum_r w_k \mathbf{x}_k = \mathbf{t}_x^*$.

The model variables \mathbf{z} are assumed to be related to the nonresponse propensity and are often called “instrumental variables”. Note that we use this term following the survey sampling literature on the topic, rather than the econometric literature on sample selection models, under which \mathbf{x} would be called instrumental variable, and model variables \mathbf{z} are more closely related to those variables that satisfy “exclusion restrictions” (see Lesage et al., 2019, for more details on this topic).

A solution for $\boldsymbol{\gamma}$ is most easily found when the number of model variables \mathbf{z} (p) coincides with that of the calibration variables \mathbf{x} (q). Then, given $F(\cdot)$ and $\widehat{\boldsymbol{\gamma}}$, the generalised calibration estimator of \mathbf{t}_y is

$$\widehat{\mathbf{t}}_y^{\text{GCAL}} = \sum_r d_k F(\mathbf{z}_k \widehat{\boldsymbol{\gamma}}) \mathbf{y}_k.$$

In particular, in the case of linear calibration,

$$w_k = d_k \left(1 + (\mathbf{t}_x^* - \sum_r d_k \mathbf{x}_k) (\sum_r d_k \mathbf{z}_k^T \mathbf{x}_k)^{-1} \mathbf{z}_k^T \right).$$

Then, the estimate of the total reduces to a sort of regression estimator such that

$$\widehat{\mathbf{t}}_y^{\text{GCAL-lin}} = \sum_r d_k \mathbf{y}_k + (\mathbf{t}_x^* - \sum_r d_k \mathbf{x}_k) \widehat{\boldsymbol{\beta}}, \quad (2)$$

where $\widehat{\boldsymbol{\beta}} = (\sum_r d_k \mathbf{z}_k^T \mathbf{x}_k)^{-1} \sum_r d_k \mathbf{z}_k^T \mathbf{y}_k$. See Chang and Kott (2008) and Kott and Liao (2017) for methods to find an estimate also when $p < q$, and Kott (2014) for a review on calibration weighting when model and calibration variables differ.

Note that the value of the model variables needs to be known only for $k \in r$ in order to compute the estimates $\widehat{\boldsymbol{\gamma}}$ and the final set of weights w_k . For this reason, generalised calibration is particularly useful for nonresponse treatment, because unlike other reweighting methods, the problem can be addressed even when the variables that cause nonresponse are known only for the respondents. This is particularly relevant when nonresponse is nonignorable, i.e. when the topic of the survey and, therefore, the variables of interest and/or other variables influence the response probability of a unit, and we observe such variables only on the respondents. In fact, the possibility of introducing the variables of interest (known only for the respondents) as instrumental variables as a means of correcting this type of nonresponse has been investigated in the literature. See Deville (2000); Kott and Chang (2010). This aspect is being further addressed to investigate the properties of the generalised calibration estimator that uses the variable of interest in the instrument vector (Lesage et al., 2019).

2.4 Variance estimation for generalised calibration

Let $\hat{t}_{ym}^{\text{GCAL}} = \sum_r d_k F(\mathbf{z}_k \hat{\boldsymbol{\gamma}}) y_{mk}$ denote the estimator of the total t_{ym} of a particular variable of interest y_m and let $\tilde{t}_{ym}^{\text{GCAL}} = \sum_r d_k F(\mathbf{z}_k \boldsymbol{\gamma}) y_{mk}$. Following Kott (2006, p. 138), an estimator of the variance of $\hat{t}_{ym}^{\text{GCAL}}$ can be derived from the form of the asymptotic variance of $\tilde{t}_{ym}^{\text{GCAL}}$. Let

$$\boldsymbol{\alpha}_m = \left(\sum_U F'(\mathbf{z}_k \boldsymbol{\gamma}) p_k \mathbf{z}_k^T \mathbf{x}_k \right)^{-1} \sum_U F'(\mathbf{z}_k \boldsymbol{\gamma}) p_k \mathbf{z}_k^T y_{mk}. \quad (3)$$

Then, under the two-phase approach, the variance of $\tilde{t}_{ym}^{\text{GCAL}}$ is the sum of two variance components – the sampling and the nonresponse variance component. That is,

$$V\{\tilde{t}_{ym}^{\text{GCAL}} | \mathcal{F}\} = V_{\text{sam}} + V_{\text{nr}},$$

where $V_{\text{sam}} = V\{\sum_s d_k e_{mk}^* | \mathcal{F}\}$ and

$$e_{mk}^* = \begin{cases} y_{mk} & \text{if } \mathbf{t}_x^* = \hat{\mathbf{t}}_x^{\text{HT}} \\ e_{mk} = y_{mk} - \mathbf{x}_k \boldsymbol{\alpha}_m & \text{if } \mathbf{t}_x^* = \mathbf{t}_x \end{cases},$$

and $V_{\text{nr}} = E\{\sum_s d_k^2 p_k^{-1} (1 - p_k) e_{mk}^2 | \mathcal{F}\}$. Then, a variance estimator of $\hat{t}_{ym}^{\text{GCAL}}$ is given by

$$\begin{aligned} \hat{V}_{2p}(\hat{t}_{ym}^{\text{GCAL}}) &= \hat{V}_{\text{sam}} + \hat{V}_{\text{nr}} \\ &= \sum_{k \neq l} \sum_{k, l \in r} d_{kl} \Delta_{kl} \frac{\hat{e}_{mk}^*}{\hat{p}_k} \frac{\hat{e}_{lk}^*}{\hat{p}_l} + \sum_r d_k \Delta_{kk} \frac{(\hat{e}_{mk}^*)^2}{\hat{p}_k} \\ &\quad + \sum_r \frac{(1 - \hat{p}_k)}{\hat{p}_k^2} (d_k \hat{e}_{mk})^2, \end{aligned}$$

where $\hat{e}_{mk}^* = y_{mk}$ if $\mathbf{t}_x^* = \hat{\mathbf{t}}_x^{\text{HT}}$ or $\hat{e}_{mk}^* = \hat{e}_{mk} = y_{mk} - \mathbf{x}_k \hat{\boldsymbol{\alpha}}_m$ with

$$\hat{\boldsymbol{\alpha}}_m = \left(\sum_r d_k F'(\mathbf{z}_k \hat{\boldsymbol{\gamma}}) \mathbf{z}_k^T \mathbf{x}_k \right)^{-1} \sum_r d_k F'(\mathbf{z}_k \hat{\boldsymbol{\gamma}}) \mathbf{z}_k^T y_{mk} \quad (4)$$

if $\mathbf{t}_x^* = \mathbf{t}_x$, and $\hat{p}_k = F(\mathbf{z}_k \hat{\boldsymbol{\gamma}})^{-1}$.

Obtaining \hat{V}_{sam} for a particular design may be cumbersome because one has to start from the respondents set. The ‘‘reverse’’ approach proposed by Shao and Steel (1999) considerably simplifies the task. In this framework, the finite population is first divided into two strata – the respondents and the nonrespondents – using the response mechanism (first phase). Then, within the population of respondents, the sample is selected using the sampling design (second phase). Since we assume that

the response mechanism is independent of the sampling process, the two terms can be exchanged in the decomposition of the variance. Therefore, the variance of a generic estimator \hat{t}_{ym} can be written as

$$\begin{aligned} V\{\hat{t}_{ym}|\mathbf{R}_N, \mathcal{F}\} &= E\{V\{\hat{t}_{ym}|\mathbf{R}_N, \mathcal{F}\}|\mathcal{F}\} + V\{E\{\hat{t}_{ym}|\mathbf{R}_N, \mathcal{F}\}|\mathcal{F}\} \\ &= V_1 + V_2, \end{aligned}$$

where $\mathbf{R}_N = (R_1, \dots, R_N)$, and the conditional distribution given \mathbf{R}_N is the distribution over the sampling mechanism treating the response indicators $R_k, k \in U$ as fixed.

Following Kim and Riddles (2012, Section 4) we can estimate the variance of $\hat{t}_{ym}^{\text{GCAL}}$ with that of its linear approximation. This can be written as

$$\tilde{t}_{ym}^{\text{GCAL}} = \mathbf{t}_x^* \tilde{\boldsymbol{\alpha}}_m + \sum_s d_k R_k F(\mathbf{z}_k \boldsymbol{\gamma})(y_{mk} - \mathbf{x}_k \tilde{\boldsymbol{\alpha}}_m),$$

where

$$\tilde{\boldsymbol{\alpha}}_m = \left(\sum_U R_k F'(\mathbf{z}_k \boldsymbol{\gamma}) \mathbf{z}_k^T \mathbf{x}_k \right)^{-1} \sum_U R_k F'(\mathbf{z}_k \boldsymbol{\gamma}) \mathbf{z}_k^T y_{mk}. \quad (5)$$

Then we have $V\{\tilde{t}_{ym}^{\text{GCAL}}|\mathbf{R}_N, \mathcal{F}\} = V\{\sum_s d_k \eta_{mk}^*|\mathbf{R}_N, \mathcal{F}\}$, where

$$\eta_{mk}^* = \begin{cases} \mathbf{x}_k \tilde{\boldsymbol{\alpha}}_m + R_k F(\mathbf{z}_k \boldsymbol{\gamma})(y_{mk} - \mathbf{x}_k \tilde{\boldsymbol{\alpha}}_m) & \text{if } \mathbf{t}_x^* = \hat{\mathbf{t}}_x^{\text{HT}} \\ R_k F(\mathbf{z}_k \boldsymbol{\gamma})(y_{mk} - \mathbf{x}_k \tilde{\boldsymbol{\alpha}}_m) & \text{if } \mathbf{t}_x^* = \mathbf{t}_x \end{cases}. \quad (6)$$

Since $V_1 = \sum_U \sum_U \Delta_{kl} \eta_{mk}^* \eta_{ml}^*$, a plug-in estimator for V_1 is

$$\hat{V}_1 = \sum_s \sum_s d_{kl} \Delta_{kl} \hat{\eta}_{mk}^* \hat{\eta}_{ml}^*,$$

where $\hat{\eta}_{mk}^*$ is such that $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\alpha}}_m$ are substituted with $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\alpha}}_m$, respectively, in η_{mk}^* . Now, to estimate V_2 , note that

$$V_2 = V\left\{ \sum_U \eta_{mk}^* | \mathcal{F} \right\} = \sum_U \frac{1 - p_k}{p_k} (y_{mk} - \mathbf{x}_k \tilde{\boldsymbol{\alpha}}_m)^2$$

can be estimated by

$$\hat{V}_2 = \sum_r d_k \frac{1 - \hat{p}_k}{\hat{p}_k^2} \hat{e}_{mk}^2. \quad (7)$$

Finally, we obtain

$$\hat{V}_{\text{rev}}(\hat{t}_{ym}^{\text{GCAL}}) = \hat{V}_1 + \hat{V}_2.$$

Note that the second term in the reverse approach is negligible compared to the first term if n/N is negligible.

The two variance estimators considered above require computation of coefficients and residuals for each variable of interest. Since we are usually interested in estimating the total of a whole set of variables of interest, a replication based variance estimator will also be considered. In particular, a jackknife variance estimator of $\widehat{t}_{ym}^{\text{GCAL}}$ is

$$\widehat{V}_{\text{jack}}(\widehat{t}_{ym}^{\text{GCAL}}) = \frac{n_r - 1}{n_r} \sum_{l \in r} ({}^{(l)}\widehat{t}_{ym}^{\text{GCAL}} - \widehat{t}_{ym}^{\text{GCAL}})^2,$$

where n_r is the size of r , ${}^{(l)}\widehat{t}_{ym}^{\text{GCAL}} = \sum_{k \in r} {}^{(l)}w_k y_{mk}$, the jackknife replicate weights are

$${}^{(l)}w_k = w_k \frac{{}^{(l)}d_k}{d_k} + \left(\mathbf{t}_x^* - \sum_{k' \in r} w_{k'} \frac{{}^{(l)}d_{k'}}{d_{k'}} \mathbf{x}_{k'} \right) \left(\sum_{k' \in r} {}^{(l)}d_{k'} F'(\mathbf{z}_{k'} \widehat{\boldsymbol{\gamma}}) \mathbf{z}_{k'}^T \mathbf{x}_{k'} \right)^{-1} {}^{(l)}d_k F'(\mathbf{z}_k \widehat{\boldsymbol{\gamma}}) \mathbf{z}_k^T,$$

${}^{(l)}d_k = 0$ when $k = l$ and ${}^{(l)}d_k = n_r d_k / (n_r - 1)$ otherwise (see Kott, 2006).

Note that the jackknife variance estimator is an estimator of the first term of the reverse approach V_1 assuming with replacement sampling. In addition, $\widehat{V}_{\text{jack}}(\widehat{t}_{ym}^{\text{GCAL}})$ is consistent for V_1 even if the nonresponse model is incorrectly specified.

3 Latent variable models

Our proposal suggests to use latent variable models to obtain a suitable model variable to be used in generalised calibration, in surveys where such latent variables can be computed. Latent variable models are multivariate regression models that link continuous or categorical manifest response variables to unobserved latent variables (see Skrondal and Rabe-Hesketh, 2007, for a survey). Several methods exist to determine such latent variables depending on the nature of the response variables. According to the nature of the latent phenomenon, we can distinguish between *latent class* and *latent trait* models. Latent classes refer to the categories of a discrete latent variable; such categories may, but need not, be ordered along a continuum. Latent traits, on the other hand, refer to a latent continuum which all individuals, based on their pattern of responses to a set of observed variables, are mapped on.

For example, we could use latent class models as a method of grouping survey respondents with respect to some underlying, unobservable variables, using responses to categorical manifest variables, such as “Do you have enough money to keep your home in a decent state of repair?”, with responses coded as either “have or don’t want or need” or “can’t afford”, included into a survey on below average income households. Alternatively, we could use continuous latent variables which are unmeasured

variables that are believed to influence responses to a survey and are related to the variable of interest or the survey topic, such as “willing to respond to the survey” or “reluctance to provide confidential information”. Other examples of applications of these ideas are discussed in Section 4.2. We now briefly present below the two types of latent variable models considered in this paper: latent class models and latent trait models.

3.1 Latent class models

In latent class models, the latent variable is supposed to be discrete (Lazarsfeld and Henry, 1968; Goodman, 1974). Latent class models allow manifest variables to be polytomous (ordered or unordered). Chapter 6 in Bartholomew et al. (2011) gives an overview of latent class models.

Let the latent class variable be denoted by ϑ_k , a particular latent class by c and the number of latent classes by C . Let the vector of manifest variables observed for $k \in r$ be denoted by $\boldsymbol{\omega}_k = (\omega_{1k}, \dots, \omega_{\ell k}, \dots, \omega_{Lk})$, whereas $\mathbf{h} = (h_1, \dots, h_L)$ refers to a possible response pattern. Components of the manifest vector may include components of the response variable vector \mathbf{y}_k and of the model vector \mathbf{z}_k . Then the latent class model can be expressed as

$$P(\boldsymbol{\omega}_k = \mathbf{h}) = \sum_{c=1}^C P(\vartheta_k = c)P(\boldsymbol{\omega}_k = \mathbf{h}|\vartheta_k = c), \quad (8)$$

in which the probability of observing a response pattern \mathbf{h} is a weighted average of class-specific probabilities. In fact, the term $P(\boldsymbol{\omega}_k = \mathbf{h}|\vartheta_k = c)$ is the conditional response probability of observing pattern \mathbf{h} given that unit k belongs to class c , and the weight is the probability that such unit belongs to the latent class c .

Conditional to the latent variable, the observed variables are assumed to be independent within latent classes, and the conditional probability of unit k takes the form

$$P(\boldsymbol{\omega}_k = \mathbf{h}|\vartheta_k = c) = \prod_{\ell=1}^L P(\omega_{\ell k} = h_{\ell}|\vartheta_k = c).$$

The model can be extended to allow for the presence of covariates that influence the probabilities $P(\vartheta_k = c)$.

The parameters of a latent class model are usually estimated by means of the EM algorithm. The number of latent classes may be chosen using model selection criteria like AIC, BIC or cAIC. The final step in a traditional latent class analysis is to use the estimated parameters of the model to classify cases into the appropriate

latent classes. The classification problem of assigning respondents to latent classes may be approached from a Bayesian point of view. In fact, posterior probabilities $P(\vartheta_k = c | \boldsymbol{\omega}_k = \mathbf{h})$ can be estimated and the final estimate of ϑ_k is given by the class with the largest posterior probability given an observed response pattern.

Note that in the context of treatment of unit nonresponse in sample surveys, classification of units in latent classes provides an alternative way of building response homogeneity groups. In this case our latent variable $\boldsymbol{\theta}_k = (\theta_{1k}, \dots, \theta_{Ck})$ will be an indicator variable vector such that $\theta_{ck} = 1$ if $\vartheta_k = c$, and 0, otherwise (see the application in Section 6). When classes can be ordered, ϑ_k can also be used as is as a latent variable (see the first simulation study in Section 5.1).

3.2 Latent trait models

For ease of notation, we consider L binary manifest variables and let $\boldsymbol{\omega}_k$ be, again, the vector of these manifest variables observed for $k \in r$. Denote by $q_{k\ell} = Pr(\omega_{\ell k} = 1 | \boldsymbol{\theta}_k)$, where $\boldsymbol{\theta}_k = (\theta_{1k}, \dots, \theta_{jk}, \dots, \theta_{Jk})$ is the value taken on unit k by the vector of $J < L$ continuous latent variables. A latent trait model (see e.g. Bartholomew et al., 2002) is defined as

$$\text{logit}(q_{k\ell}) = \beta_{\ell 0} + \sum_{j=1}^J \beta_{\ell j} \theta_{jk},$$

for $\ell = 1, \dots, L$, where $\beta_{\ell 0}, \dots, \beta_{\ell J}$ are model parameters. An important special case is obtained by taking $J = 1$, i.e. a unidimensional latent trait model

$$\text{logit}(q_{k\ell}) = \beta_{\ell 0} + \beta_{\ell 1} \theta_k, \tag{9}$$

where, for model identifiability, the location and scale parameters of θ_k are set to pre-specified values, usually 0 and 1, respectively. Model (9) is also referred to as a two parameter logistic Rasch model, and is essentially a logistic regression except that the θ_k 's are not observed. Covariates may be introduced in the process of estimating the latent variable θ_k . Extension to ordered categorical manifest variables can be accommodated using Polytomous Rasch Models and Partial Credit Models (Fischer and Molenaar, 1995).

When $J = 1$, it is assumed that the L manifest variables measure a single latent variable. The probability of success $q_{k\ell}$ is assumed to be a monotonic increasing function of the latent trait θ_k (*monotonicity* assumption). In addition, under the assumption of *conditional independence*, the values of the L manifest variables are assumed to be independent for each k given θ_k . Different goodness-of-fit measures of

Model (9) are available in the literature to check whether these assumptions hold (see e.g. Bartholomew et al., 2002). Note that, however, these assumptions are particularly relevant when the objective of the study is the estimation and interpretation of the model parameters. In our perspective, the main objective is prediction and the ability to obtain a good reduction of the information from the manifest variables, to be used as members of the instrumental vector. The parameters of Model (9) can be estimated using different methods: maximum likelihood, generalised least squares, EM algorithm or Markov Chain Monte Carlo. The book of Bartholomew et al. (2011) provides a good description of these methods (Section 4.5 for an EM algorithm, Section 4.8 for the generalised least squares, and Section 4.11 for Markov Chain Monte Carlo methods).

4 Generalised calibration with latent variables

4.1 The proposed methodology

Once an estimate of the latent variable is obtained, either using latent class or latent trait models, it can be used in a generalised calibration framework as the instrumental variable \mathbf{z} or as a component of the vector of instrumental variables. The choice between latent trait models over latent class models is essentially dictated by the application at hand. Other types of latent models may also be used, according to the type of manifest variables we have available.

In this section we wish to discuss the properties of $\hat{\mathbf{t}}_y^{\text{GCAL}} = \sum_r w_k \mathbf{y}_k$, when a latent variable is used in the construction of w_k . First note that, since the components of \mathbf{z}_k come from the survey and may include information on the survey variables of interest, properties that involve a superpopulation linear model between a particular survey variable, say y_m , and the auxiliary variables \mathbf{x} are more problematic to establish, rather than in the simpler calibration case (see Kott, 2006; Kott and Chang, 2010). Then, we will focus on the response model. In a multipurpose setting, in addition, this is a sensible choice also because a good model for one survey variable may not be as good for another survey variable.

A necessary condition, therefore, under which $\hat{\mathbf{t}}_y^{\text{GCAL}}$ is a consistent estimator for \mathbf{t}_y , is that $p_k = F(\mathbf{z}_k \boldsymbol{\gamma})^{-1}$. This condition has two main ingredients: the function $F(\cdot)$, that we have discussed already, and the choice of \mathbf{z} . On this regard, Lesage et al. (2019) show that in order to achieve consistency, it is also required that the response mechanism is independent of \mathbf{x} given \mathbf{z} and, similarly, the response mechanism is independent of each element of \mathbf{y} given \mathbf{z} . In other words, the instrumental

variables should extract all the relevant information that links the response process with \mathbf{x} and with the response variables. Note that, in practice, it is usually not possible to validate the choice of $F(\cdot)$ and/or the choice of the model variables in \mathbf{z} , and their relationship with the survey and the auxiliary variables. In fact, model variables are only available for the respondents. In our opinion, external data on the nonrespondents and/or on the population should be available in order to validate such assumptions. See, for example, the application in Section 6.

As of variance estimation, we will consider the estimators proposed in the literature and reviewed in Section 2.4. We tested these estimators in the simulation studies in Section 5.

4.2 Motivation of using latent variables

There are several reasons why we find it is useful to introduce latent variables in generalised calibration. The first, and more obvious, is to reduce nonresponse bias compared to the classical calibration estimator, because it introduces information on the nonresponse behaviour when we believe that nonresponse depends on unobservable variables that can be estimated from a set of manifest variables using latent variable models. Of course this can be achieved also using generalised calibration using all the manifest variables as model variables. However, we will show in the simulation studies in Section 5, that using latent variables reduces considerably the variance of the generalised calibration estimator as opposed to including the variable(s) of interest. Note that this is true also when nonresponse depends on the variable of interest. In fact, it is known that in order to reduce the bias and the variance of the generalised calibration estimator, the model and the calibration variables should be highly correlated. Lesage et al. (2019) show that the generalised calibration estimator has a problem of variance amplification if the calibration variables and the model variables are not correlated, as this implies that the vector γ is poorly estimated.

Using information from the survey variables of interest in the instrumental vector is shown to provide a trade-off between bias reduction and variance increase. The use of latent variables provides a very useful tool to achieve a compromise since we use an instrumental variable that is well correlated with both the variable of interest and the calibration variables. Note that calibration variables can be used in the estimation of the latent variables, by this increasing correlation and decreasing variance (see the first simulation study in Section 5.1).

In addition, depending on the context, a latent variable may be included as single instrumental variable, or together with other instrumental variables in generalised calibration. Consider the following examples. In attitude and behavioural surveys, nonresponse may be related to the survey topic itself (see Groves et al., 2006). It is the case of surveys on attitudes towards politics or abortion. In this context, a latent variable called “attitude towards politics” or “attitude towards abortion” can better explain the response probability of a unit to answer the survey questionnaire than any observed variable. Thus, it is possible to model the response probability as a function of this latent variable.

A similar example concerns a latent variable called “will to respond to the survey” which can be computed from item response indicators to the observed variables (e.g. Moustaki and Knott, 2000; Matei and Ranalli, 2015). The “will to respond to the survey” latent variable is well measured when nonresponse is nonignorable; e.g. in attitude surveys or in income surveys. Since this latent variable is a measure of the will to answer the survey, it can be assumed that response probabilities depend on it. As in the previous example, the “will to respond to the survey” latent variable may be included as an instrumental variable in the generalised calibration method.

On the other hand, most surveys are multipurpose. It is often difficult to make the assumption that nonresponse depends upon a single variable of interest. Again, the use of few latent variables which capture the behavior of variables of interest is useful because it reduces the dimensionality of the problem and allows for bias adjustment, without increasing the variance as much as when several variables of interest are used. In addition, note that the use of several variables of interest as model variables largely increases the variability of the final set of calibration weights. Finally, latent variables are a useful means of dealing with the presence of measurement error in the manifest variables. Indeed, the latent variable approach is widely used to deal with such issues; see e.g. Biemer (2009).

Section 6 provides evidence with regard to both these issues using data from the Survey on Household Income and Wealth (SHIW). It is conducted by the National Central Bank of Italy every two years and collects detailed information on household income, wealth and consumption. In these instances, it is not sensible to assume that unit nonresponse depends only on a single surveyed item. In the SHIW survey, some households may refuse to participate because they don’t want to report their income from self-employment. They may fear the consequences of disclosure of the answers to tax authorities. Others may not report all the houses they own because of social desirability bias or because they fear they could be robbed. Others may not

be willing to participate because they don't want to report their financial wealth. Indeed, the survey collects information on two topics that are generally considered highly sensitive: income and household wealth. Therefore, it is reasonable to assume that each of these factors could contribute to nonresponse. A second issue that arises in the case of SHIW, is that of measurement error, since previous research has shown that both income and wealth are affected by underreporting (Neri and Zizza, 2010; Neri and Ranalli, 2011). Failing to consider such an issue may result in biased estimates. In this context, it is possible to create response homogeneity groups which are here in fact latent classes. The latter are constructed using survey responses as proxies of true wealth, and used in the generalised calibration method as instrumental variables.

5 Simulation studies

We present two simulation studies. The first one uses a categorical latent variable, while the second one a continuous one. In both studies, for a generic estimator \hat{t}_y the following measures are computed:

- the Monte Carlo Bias

$$B = E_{sim}(\hat{t}_y) - t_y,$$

where

$$E_{sim}(\hat{t}_y) = \sum_{i=1}^{nsim} \hat{t}_{yi} / nsim,$$

\hat{t}_{yi} is the value of the estimator \hat{t}_y at the i -th simulation run and $nsim$ is the total number of simulation runs;

- the Monte Carlo Variance

$$VAR = \frac{1}{nsim - 1} \sum_{i=1}^{nsim} (\hat{t}_{yi} - E_{sim}(\hat{t}_y))^2;$$

- the Monte Carlo Mean Squared Error

$$MSE = B^2 + VAR.$$

5.1 Simulation using a categorical latent variable

The first simulation study is based on the 'election' data available in the R package `poLCA` (Linzer and Lewis, 2014). This data set reports information gathered

during the 2000 American National Election Study public opinion poll. Following Linzer and Lewis (2011), three latent classes are constructed based on two sets of six items concerning a series of traits (moral, caring – really cares about people like you, knowledgeable, strong leadership, dishonest, and intelligent) potentially describing the candidates Al Gore and George W. Bush: ‘a reasonable theoretical approach might suppose that there are three latent classes of survey respondents: Gore supporters, Bush supporters, and those who are more or less neutral’.

Each question has four possible choices: (1) extremely well; (2) quite well; (3) not too well; and (4) not well at all. The party respondent’s identification (seven different parties, from strong Democratic to strong Republican, with Independents at 3-4-5 on the scale) is used as covariate x in the construction of the three latent classes. After deleting observations with missing values, the population size is 880. The model predicts that 26% of the units are Gore supporters, 35% of the units are Bush supporters, and 39% of the units are neutral. Eight observations considered as badly classified into the three classes have been removed from the population. Thus, $N = 872$. The variable of interest y is the fourth item in this data set (the question is ‘In your opinion, does the phrase **he is knowledgeable** describe Al Gore extremely well, quite well, not too well, or not well at all?’).

We focus on estimation of the population total of people answering ‘extremely well’ or ‘quite well’ to the fourth item. The variable of interest y is, thus, dichotomised, taking value 1 if the answer is ‘extremely well’ or ‘quite well’ and 0 otherwise. The population total is $t_y = 476$.

A number of 10,000 samples of size $n = 300$ have been drawn from the population using simple random sample without replacement. From the selected samples, the respondents sets were generated using Poisson sampling with probabilities $p_k = F(\mathbf{z}_k \boldsymbol{\gamma})^{-1}$. The response probabilities p_k were computed using four response models, respectively, where z is replaced by ϑ or y :

$$\text{Model 1: } p_k = 1/(1.1 + 0.9\vartheta_k);$$

$$\text{Model 2: } p_k = 1/\exp(0.2 + 0.5\vartheta_k);$$

$$\text{Model 3: } p_k = 1/(1.5 - 0.3y_k);$$

$$\text{Model 4: } p_k = 1/\exp(0.4 - 0.2y_k).$$

For each response model, the population average for p_k is around 0.7. The population correlation coefficient between the auxiliary information x (here ‘Party’) and y is

-0.56, while between x and ϑ is 0.75; the correlation coefficient between y and ϑ is -0.63.

Corresponding to Models 1, 2, 3, and 4, there are respectively four link functions to use in generalised calibration:

$$\mathbf{Lin}-\vartheta : F(\vartheta_k\gamma) = 1.1 + 0.9\vartheta_k;$$

$$\mathbf{Rak}-\vartheta : F(\vartheta_k\gamma) = \exp(0.2 + 0.5\vartheta_k);$$

$$\mathbf{Lin}-y : F(y_k\gamma) = 1.5 - 0.3y_k;$$

$$\mathbf{Rak}-y : F(y_k\gamma) = \exp(0.4 - 0.2y_k).$$

We used two different forms of the link function (linear and exponential) since the choice of $F(\cdot)$ is important as underlined in Section 2. To test the performances of the different weighting systems, three estimators were computed: a generalised calibration, where z is the latent variable ϑ (denoted below by $\mathbf{GCAL}-\vartheta$), a generalised calibration, where z is the variable of interest y ($\mathbf{GCAL}-y$), and a simple calibration estimator, where z is the auxiliary variable x ($\mathbf{CAL}-x$). Table 1 reports the simulation results considering the previous Monte Carlo measures. The (generalised) calibration weights have been computed using the appropriate link function in the R package `sampling` (Tillé and Matei, 2021) and then applied to estimate t_y . Finally, variance estimators have been considered as well. Note that in this particular setting in which we use simple random sampling without replacement to select the sample, and we use population level auxiliary information, we have found out that \widehat{V}_{2p} and \widehat{V}_{rev} coincide. Appendix A provides a proof of this equivalence. We will denote simply by \widehat{V} in Table 1 the result from these estimators.

Model 1, Model 2, Model 3 and Model 4, corresponding to $\mathbf{Lin}-\vartheta$, $\mathbf{Rak}-\vartheta$, $\mathbf{Lin}-y$ and $\mathbf{Rak}-y$, respectively are the *true response models*. Note that a true response model is not necessarily the same with the underlying model used in the computed estimators, because the model variable may be different. The models underlying the computed estimators are called here *working models*. The link functions corresponding to them share the same form of the link functions (linear or exponential) corresponding to the true response models. In Table 1 the three estimators are given in the column ‘Working model’, below each $F(\cdot)$ corresponding to a true response model.

In Model 1 and Model 2, z is taken to be ϑ , following thus theoretical approaches where political partisanship is strongly correlated with attitudes and behavior and

Table 1: First simulation study. Monte Carlo Bias, Variance and MSE for each estimator in each simulation setting. Monte Carlo mean and 95% confidence interval coverage for each variance estimator. Election data set, $n = 300, N = 872$.

	Working model	B	VAR	MSE	\widehat{V}	95%	$\widehat{V}_{\text{jack}}$	95%jack
Lin-ϑ								
\hat{t}_y	GCAL- y	-18.9	1371.7	1728.0	1372.9	92.4%	1850.1	96.4%
	CAL- x	5.7	424.6	457.4	420.0	93.8%	572.2	97.0%
	GCAL- ϑ	0.1	480.7	480.7	476.1	94.5%	645.2	97.5%
Rak-ϑ								
\hat{t}_y	GCAL- y	21.9	1075.3	1557.0	1107.6	91.7%	1626.0	96.6%
	CAL- x	-4.9	311.8	335.9	309.1	94.0%	462.6	97.8%
	GCAL- ϑ	0.9	347.0	348.0	344.3	94.7%	514.6	97.9%
Lin-y								
\hat{t}_y	GCAL- y	-0.1	1410.1	1410.2	1427.1	95.7%	1915.3	97.8%
	CAL- x	33.4	431.2	1547.9	429.8	64.1%	582.1	75.2%
	GCAL- ϑ	26.1	483.9	1164.9	488.2	77.9%	657.4	86.0%
Rak-y								
\hat{t}_y	GCAL- y	0.03	1419.9	1419.9	1439.9	95.8%	1928.3	97.8%
	CAL- x	30.0	432.3	1334.3	432.8	69.9%	585.2	80.0%
	GCAL- ϑ	23.3	486.1	1030.3	492.0	81.3%	661.7	88.5%

thus can influence the probability to answer a political survey. In Model 3 and Model 4, z is the variable of interest y . The last two models represent a tool to test the robustness of GCAL- ϑ , since p_k depends on y_k , and not on ϑ_k .

Following the results given in Table 1, we note that when the response model is based on ϑ (**Lin- ϑ** and **Rak- ϑ**), GCAL- ϑ has, as expected, a negligible bias and a smaller variance than GCAL- y . It also reduces the bias compared to CAL- x , since ϑ is more correlated with the variable of interest y than x . On the other hand, as expected, its variance is larger than CAL- x . It also shows a larger mean square error than CAL- x , but of similar size. GCAL- ϑ provides thus a good compromise: it shows smaller bias than the classical calibration estimator CAL- x and smaller variance than GCAL- y .

When GCAL- y is based on the true response model (**Lin- y** and **Rak- y**) its bias is negligible, but its variance is relatively larger than that of the other estimators, including the classical calibration estimator. The classical calibration estimator CAL- x has the opposite behavior, with a large bias and a small variance that places concern on the coverage rate. This behavior of the generalised calibration estimator versus the classical calibration one is also verified in other simulation studies given in the literature (see e.g. Osier, 2013). When the response model is based

on y (**Lin**- y and **Rak**- y), GCAL- ϑ also provides the previous good compromise: it shows smaller bias than CAL- x and smaller variance than GCAL- y . But also its mean square error is smaller than the mean square errors of GCAL- y and GCAL- x .

Table 1 also reports the performance of the variance estimators. For each variance estimator, its Monte Carlo expectation is computed, together with the empirical coverage rate of a 95% confidence interval. We note that \widehat{V} has better behavior than \widehat{V}_{jack} in terms of Monte Carlo expectation. However, better coverage rates are provided by \widehat{V}_{jack} when the working model is misspecified. In fact, \widehat{V} shows lower coverage rate when the response model depends on y and it is not the true one for CAL- x and GCAL- ϑ ; this is due to the presence of a larger bias displayed by the estimators CAL- x and GCAL- ϑ . In some instances, the coverage rate of \widehat{V}_{jack} is above 95%. This may be due to a relatively high sampling fraction, $300/872 = 34.4\%$. Since \widehat{V}_{jack} does not incorporate the finite population correction, the first term V_1 is overestimated and this compensates the underestimation of the second term V_2 of the reverse approach that is not negligible in this case.

5.2 Simulation using a continuous latent variable

We consider the ‘Abortion’ data set formed by four binary variables extracted from the 1986 British Social Attitudes Survey and concerning attitudes to abortion. $N = 379$ individuals answered the following questions after being asked if the law should allow abortion under the circumstances presented under each item:

1. The woman decides on her own that she does not want to keep the baby.
2. The couple agrees that they do not wish to have a child.
3. The woman is not married and does not wish to marry the man.
4. The couple cannot afford any more children.

The data are analyzed in Bartholomew et al. (2002) and found to hide a latent continuous variable, interpretable as the “attitude to abortion”. It is plausible in this example to assume that the response probability of a unit k to the survey depends on its attitude to abortion θ_k , rather than on a single variable of interest enumerated before. In our study, we will consider both cases.

We focus on estimation of the total of the second variable in this data set (y), having in the population $t_y = 225$. At the population level, the latent variable θ_k has been estimated for each unit using Model (9) and the R package `1tm` (Rizopoulos,

2006). The method used for estimation is marginal maximum likelihood. An auxiliary variable x_k has been generated as $x_k = 1 + \sum_{m=1}^4 y_{mk} + \epsilon_k$, with $\epsilon_k \sim N(0, 1)$. Its population total is assumed known. The simulation data frame consists of $N = 379$ observations with binary answers to the previous four items, associated latent variable θ_k and auxiliary information x_k . The correlation coefficient between x and θ , and between y and θ is around 0.8 in the population, respectively.

We draw 10,000 simple random samples without replacement of size $n = 100$ (similar results not shown here have been obtained using the same settings, and a sample size $n = 50$). For each selected sample, nonresponse was simulated using Poisson sampling with probabilities defined for each unit according to two different models:

$$\text{Model 5: } p_k = 1 / \exp(0.1 + 0.5\theta_k);$$

$$\text{Model 6: } p_k = 1 / \exp(0.1 + 0.5y_k).$$

The corresponding link functions in generalised calibration are respectively:

$$\mathbf{Rak}-\theta : F(\theta_k\gamma) = \exp(0.1 + 0.5\theta_k);$$

$$\mathbf{Rak}-y : F(y_k\gamma) = \exp(0.1 + 0.5y_k).$$

In both models, the population average for p_k is around 0.7. As in Section 5.2, three estimators were computed: a generalised calibration, where z is the latent variable θ (denoted below by GCAL- θ), a generalised calibration, where z is the variable of interest y (GCAL- y), and a simple calibration estimator, where z is the auxiliary variable x (CAL- x). In particular, by using GCAL- θ , we wish to investigate the ability of the proposed approach to reduce bias when Model 5 is used to create the set of the respondents in the population, and its robustness in terms of bias and variance in the case when Model 6 is employed. In this example, only the exponential link function was used. Similar results (not shown here) have been obtained using the linear case.

Table 2 shows the results for ‘abortion’ data. Estimators behave similarly to the first simulation study. The generalised calibration estimator GCAL- θ has, as expected, a very good performance in the **Rak**- θ setting, but is also a very good compromise in the other case. We also note that when the response model depends on y (**Rak**- y) the use of θ_k as model variable shows good behavior in terms of both bias and variance compared to GCAL- y and CAL- x . In particular, we can see, again, that the proposed approach is robust also if nonresponse depends only one

variable of interest. In addition, it is always much less variable than the one that uses the variable of interest. Note that the estimator provided by our approach shows, in these settings, smaller values of the mean square error than those given by the calibration estimator that uses only classical auxiliary information. The variance estimators behave similarly to the the first simulation.

Table 2: Second simulation study. Monte Carlo Bias, Variance and MSE for each estimator in each simulation setting. Monte Carlo mean and 95% confidence interval coverage for each variance estimator. Abortion data set, $n = 100$, $N = 379$.

	Working model	B	VAR	MSE	\widehat{V}	95%	$\widehat{V}_{\text{jack}}$	95%jack
Rak-θ								
\hat{t}_y	GCAL- y	12.1	403.1	549.0	414.5	93.1%	508.8	95.8%
	CAL- x	-8.8	195.5	272.9	192.3	90.4%	242.3	94.0%
	GCAL- θ	0.4	236.9	237.0	234.8	94.3%	292.5	96.6%
Rak-y								
\hat{t}_y	GCAL- y	0.9	405.2	406.1	417.4	95.4%	512.1	97.3%
	CAL- x	-21.3	206.7	661.6	204.6	68.3%	256.8	76.4%
	GCAL- θ	-11.3	246.2	372.8	247.1	87.6%	306.5	91.7%

6 Application to the Italian Survey on Household Income and Wealth

The Survey on Household Income and Wealth (SHIW) is conducted by the National Central Bank of Italy every two years to study the economic behaviours of Italian households by collecting detailed information on their income and wealth. Our goal here is to make inferences about the average household net wealth for the Italian population in 2008. We use the 2008 wave since for this survey we have access to auxiliary information that can be used to better explore the properties of our estimator. The sample consists of about 8,000 households selected from population registers using a complex two-stage (municipality-household) sampling design. The survey also collects paradata on the recruitment process of households, such as information on the call attempts and on the characteristics of the dwelling, such as exterior appearance and the overall assessment of the area where the building is located. Because of the sensitiveness of the issues surveyed, measurement error and unit nonresponse are two major components of total survey error. On the contrary, item nonresponse is negligible because of the design of the survey (in case the respondent does not provide information on income or wealth, the overall interview is not accepted).

Previous research based on SHIW data shows that nonresponse is nonignorable and depends on the true wealth (for more details see ?Neri and Ranalli, 2011; ?). It should be noted that the true wealth is not observed either for respondents because of measurement error. Our approach uses survey responses related to wealth as proxies of true wealth and builds latent classes to be used as response homogeneity groups. In particular, we use the following variables to build the vector of manifest variables ω_k : the individual observed 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. Overall, the vector of manifest variables ω is made of $L = 12$ components.

Using latent class analysis, we classify respondents into five latent classes. We used the BIC measure to select the number of classes to use. The value of the BIC for the six class model was slightly smaller than the one for five. However, the sixth class was very small and not very different from another one, and so we decided to use a classification in five latent classes. The latent classes can be interpreted in terms of the household’s economic well-being. The first class is formed by the **very rich** and it represents about 16% of the sample. It is mainly made of individuals with a high level of financial and non-financial wealth, living in a luxurious residence and who are **difficult to contact** or to interview. The second class can be named the **well-off**: it includes households with a medium-high wealth, living in a luxurious residence, with high financial wealth, and who are **easy to contact** or to interview (about 30% of the sample). The third class is made of **average** households (average wealth, owners of some land and non-residential buildings, **easy to contact** or to interview – about 16% of the sample). The fourth class is composed of **below average** households (about 25% of the sample). They have low levels of wealth but they are **easy to contact** or to interview. Finally, the last class is made of the **very poor**. Households in this class have a low wealth, almost zero financial wealth and they are **difficult to contact** or to interview (about 13% of the sample). Note that the latent classes are not strictly ordered. This is another reason why we have decided to use a latent class model, instead of a latent trait model: in this application one latent trait was not fitting the data well, by this implying that these manifest variables hide a multidimensional structure.

The predicted latent class memberships θ_k are used as the model variables z_k . As calibration variables, we use two sources of auxiliary information. The first one is

the Italian National Statistical Office and consists of the distribution of individuals according to some demographic variables: age (5 classes), gender, education (3 levels), nationality (italian/foreigner), job status (employed/unemployed/inactive), geographical area (north/centre/south). The second one is the Italian Department of the Treasury holding administrative records of real estate owners and consists of the distribution of individuals according to the value of the owned dwelling (5 classes). Overall, the vector of population totals \mathbf{t}_x is made of $q = 18$ components.

Six different approaches to estimate the mean household wealth are compared:

- (1) the Hajek estimator $\sum_r d_k y_k / \sum_r d_k$ (without nonresponse adjustment);
- (2) a double expansion estimator in which response probabilities are estimated via a logistic model that uses covariates known also for nonrespondents (see details in Neri and Ranalli, 2011);
- (3) the classical calibration estimator using the aforementioned 18 calibration variables;
- (4) the generalised calibration using the predicted latent class memberships $\hat{\theta}_k$ as model variables;
- (5) the generalised calibration using twelve manifest variables ω_k as model variables;
- (6) the generalised calibration using the individual observed wealth class (classes of y_k) as model variables.

For cases (4), (5) and (6), since $p < q$, we use a two-step generalised calibration approach to obtain final weights $w_k = d_k F(\mathbf{z}_k \hat{\gamma}_1) F(\mathbf{x}_k \hat{\gamma}_2)$. The first step is the adjustment step $F(\mathbf{z}_k \hat{\gamma}_1)$ obtained using the method proposed in Chang and Kott (2008). Since it does not provide calibrated weights, in the second step $F(\mathbf{x}_k \hat{\gamma}_2)$ is obtained calibrating on \mathbf{t}_x using $d_k F(\mathbf{z}_k \hat{\gamma}_1)$ as starting weights. The inverse of the two step adjustment $F(\mathbf{z}_k \hat{\gamma}_1) F(\mathbf{x}_k \hat{\gamma}_2)$ provides a measure of the implicit response probability estimated with generalised calibration. Table 3 reports the mean value of $1/F(\mathbf{z}_k \hat{\gamma}_1) F(\mathbf{x}_k \hat{\gamma}_2)$ computed for household wealth and income percentiles. As expected, the higher the household income or wealth, the lower the estimated response probability. Households in the lowest quintile of both distribution have a response propensity of about 0.8. This figure drops to about 0.4 in the upper tail of the distribution. The relation is almost monotonic as far as wealth is concerned.

As to income, the drop in the response probability is mainly due to households with an median-high income.

Table 3: Estimated mean of the implicit response probability estimated using the proposed methodology by household (HH) wealth and income percentile.

Percentile of HH wealth	Percentile of HH income					Total
	less than 20	20–39.9	40–59.9	60–79.9	80–100	
less than 20	0.82	0.83	0.81	0.81	0.72	0.82
20-39.9	0.75	0.74	0.74	0.72	0.73	0.74
40-59.9	0.63	0.69	0.64	0.58	0.47	0.63
60-79.9	0.55	0.54	0.54	0.50	0.50	0.52
80-100	0.44	0.46	0.45	0.45	0.44	0.45
Total	0.73	0.72	0.65	0.53	0.47	

Table 4 reports the results for all the approaches with an estimate of the standard error using jackknife. Nonresponse adjustments via calibration all provide an increase in the estimate. This is reasonable since we expect wealthier households to be less cooperative. It is worth noting that the true mean of the households' wealth is unknown for this application. In fact, there is a macro estimate of total household wealth produced the Bank of Italy using different sources that do not include the SHIW survey. For 2008 this estimate is about 8,500 billions of euro, i.e. 368,840 euros mean estimate. We do not use this information in calibration because of the differences in definitions between the two sources. Moreover, the macro estimate is constructed using also statistical models and assumptions. Nonetheless, it can provide some insights on the goodness of our approach. In Table 4 the ratio of the estimates to this value is also provided. It can be noted that all estimates based on generalised calibration get closer to the macro one.

The estimated variance of the generalised calibration estimators (4) – (6) is larger than that of classical calibration (3), because of the increased complexity in the first step nonresponse adjustment. As already noted in the simulations, generalised calibration that uses the variables of interest as model variables (cases (5) and (6)) provide more variable estimators than the one that uses latent variables. By comparing (5) with (4), it seems that the reduction in dimensionality performed by latent class analysis allows for a more stable estimator, without losing too much in terms of information. This is also true when comparing (4) and (6), that have the same number of model variables. Finally, note that (4) shows a far less variable set of weights as opposed to (5) and (6).

Table 4: Estimated mean of the households’ wealth using different estimators, ratio to the macro estimate of the total, 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

7 Conclusions

This paper proposes the use of the latent variables in generalised calibration to handle nonignorable unit nonresponse in the surveys where such latent variables can be computed. The method is tested on a series of simulation studies and applied to adjust estimates from the Italian Survey of Household Income and Wealth.

We find that the method presents some merits compared to the main current proposal available in the literature, which consists in using the variable of interest as an instrument in generalised calibration. This approach has at least two main pitfalls. First, nonresponse is likely to depend not just upon the variable of interest. This is particularly the case with multipurpose surveys. One may also try to add other observed variables, but this usually increases the complexity of the estimator and, therefore, the variability of the estimates and of the final set of weights. A second pitfall is that observed variables may be affected by measurement error. Failing to consider such an issue may result in biased estimators.

We considered in our examples only a single latent variable as instrumental variable. However, several latent variables can also be used, because the vector \mathbf{z} is p -dimensional. For example, in the continuous case, in addition to “will to respond to the survey” presented in Section 4.2, a second latent variable “reluctance to provide confidential information” can also be employed as instrumental variable, since they could explain more accurately dependence among the manifest variables. In the discrete case, more than one class variable can be used. We borrow from Jeon et al. (2017) an example where it is investigated adolescent violent behavior and drug-using behaviors. For this example, nonresponse may be related to the survey topic.

Three class variables are provided using data from 4,957 male high-school students who participated into a “Youth Risk Behavior Surveillance System” in 2015: violent behavior (with four classes), alcohol drinking (with three classes), tobacco cigarette smoking (with three classes), and other drug use (with three classes). These class variables can be introduced as instrumental variables in our approach. On the other hand, Jeon et al. (2017) propose a joint latent class analysis, which seeks to explain the joint patterns of two or more latent class variables. For the example provided, three joint classes are identified based on the previous class variables, reducing finally again the approach to the use of a single latent variable.

We find that our approach is a parsimonious method to relax the assumption that nonresponse depends on the single variable of interest. Moreover, the latent variable framework is widely used in the literature to deal with the issue of measurement error. Therefore it allows us to relax the assumption that manifest variables used as instruments are error-free. Our simulations show that the latent variable approach reduces the variance compared to the generalised calibration even when the response probabilities depend on the variable of interest and the latter is used as an instrumental variable. It is also more efficient than classical calibration that uses only auxiliary information as model variables.

We have considered latent trait models, which provide a continuous measure of the response propensity, and latent class models, which provide a categorical measure thereof. The choice between the two is essentially dictated by the application at hand. However, some guidelines can be provided. In general, weights based on a small number of classes tend to be more stable than the weights obtained using an estimated propensity. Latent classes provide a model based alternative to the construction of post-strata that are, by construction, built using similar units. In addition, note that, very often in surveys, even when a continuous response propensity is estimated, it is not used as is, but it is often categorised in classes (see e.g. Haziza and Lesage, 2016). We do believe that using latent classes provides a much better and data-driven classification of units, than subjectively setting thresholds of values for response propensities.

We also present an application of the method to survey data relating to household wealth and compare it with other methods. To the best of our knowledge, this is the first application to real data of generalised calibration with variables of interest used as instrumental variables. To avoid the issue of variance amplification as highlighted in Lesage et al. (2019), it is important to use auxiliary variables that are well correlated with the latent variables and the non-response behavior. In this way

it is possible to estimate better the calibration parameters. In our application, this role has been played by auxiliary information coming from administrative records of real estate owners and consists of the distribution of individuals according to the value of the owned dwelling. More research is still needed on tools to select auxiliary and model variables in generalised calibration.

Of course, this approach suffers from bias when the assumed nonresponse model is misspecified in terms of the model variables or in terms of the link function. The robustness of the generalised calibration estimator in this case is still an open problem. We can never safely assume one thing or the other unless we have external independent data (on nonrespondents) that can help verify these. As an alternative, it is important to have external validation information that we can use as a sort of gold standard to evaluate results. In our application, we have used to this end a macro estimate of the total household wealth produced using different sources that do not include the SHIW survey. We find that the use of latent variables seems to provide a good compromise between bias reduction and variance increase, and helps to make generalised calibration more usable.

We end the discussion with some cautionary considerations in the application of the proposed methodology. As we underlined, our proposal is suitable for estimation using data issued from surveys where it makes sense to compute latent variables (where the assumptions and usual measures of goodness-of-fit are fulfilled for such models). Sometimes, however, even in these surveys, possibly poor latent variables for generalised calibration can occur, showing a weak correlation with the calibration variables, and/or being not necessarily related to the nonresponse structure. We provide below some practical guidelines to decide between the use a (possibly poor) latent variable or the direct use of the manifest variables:

- compute some measure of association and/or correlation between the latent variable(s) and the calibration variables; if their values are low, a simple calibration approach is suitable instead of generalised calibration, and the proposed method should be avoided;
- if the measures of association and/or correlation are important, compute two estimates: using the latent variable(s) and using the manifest variables;
- for each estimate, compute: the jackknife standard error estimator proposed in Section 2.4, the estimated coefficient of variation, and the standard deviation of the final weights. Finally, choose the method providing the best results, minimising these three measures.

A Equivalence of $\widehat{V}_{2p}(\hat{t}_{ym}^{\text{GCAL}})$ and $\widehat{V}_{\text{rev}}(\hat{t}_{ym}^{\text{GCAL}})$ under simple random sampling without replacement and $\mathbf{t}_x^* = \mathbf{t}_x$

Under simple random sampling without replacement and $\mathbf{t}_x^* = \mathbf{t}_x$,

$$\begin{aligned}
\widehat{V}_{2p}(\hat{t}_{ym}^{\text{GCAL}}) &= \widehat{V}_{\text{sam}} + \widehat{V}_{\text{nr}} \\
&= \sum_{k \neq l} \sum_{k, l \in r} d_{kl} \Delta_{kl} \frac{\hat{e}_{mk}}{\hat{p}_k} \frac{\hat{e}_{ml}}{\hat{p}_l} + \sum_r d_k \Delta_{kk} \frac{\hat{e}_{mk}^2}{\hat{p}_k} \\
&+ \sum_r \frac{(1 - \hat{p}_k)}{\hat{p}_k^2} (d_k \hat{e}_{mk})^2 \\
&= \frac{N^2}{n^2} \frac{N-1}{n-1} \left[\frac{n-1}{N-1} - \frac{n}{N} \right] \sum_{k \neq l} \sum_{k, l \in r} \frac{\hat{e}_{mk}}{\hat{p}_k} \frac{\hat{e}_{ml}}{\hat{p}_l} + \frac{N}{n} \left(\frac{N}{n} - 1 \right) \sum_r \frac{\hat{e}_{mk}^2}{\hat{p}_k} \\
&+ \frac{N^2}{n^2} \sum_r \frac{(1 - \hat{p}_k)}{\hat{p}_k^2} \hat{e}_{mk}^2,
\end{aligned}$$

and

$$\begin{aligned}
\widehat{V}_{\text{rev}}(\hat{t}_{ym}^{\text{GCAL}}) &= \widehat{V}_1 + \widehat{V}_2 \\
&= \sum_s \sum_s d_{kl} \Delta_{kl} R_k F(\mathbf{z}_k \hat{\gamma}) \hat{e}_{mk} R_l F(\mathbf{z}_l \hat{\gamma}) \hat{e}_{ml} + \sum_r d_k \frac{1 - \hat{p}_k}{\hat{p}_k^2} \hat{e}_{mk}^2 \\
&= \frac{N}{n} \frac{N-n}{n-1} \left[\sum_r \frac{\hat{e}_{mk}^2}{\hat{p}_k^2} - \frac{1}{n} \left(\sum_r \frac{\hat{e}_{mk}}{\hat{p}_k} \right)^2 \right] + \frac{N}{n} \sum_r \frac{1 - \hat{p}_k}{\hat{p}_k^2} \hat{e}_{mk}^2 \\
&= \left[\frac{N}{n} \frac{N-n}{n-1} + \frac{N}{n} \right] \sum_r \frac{\hat{e}_{mk}^2}{\hat{p}_k^2} - \frac{N}{n} \frac{N-n}{n-1} \frac{1}{n} \left(\sum_r \frac{\hat{e}_{mk}}{\hat{p}_k} \right)^2 + \\
&- \frac{N}{n} \sum_r \frac{\hat{e}_{mk}^2}{\hat{p}_k}. \tag{A.1}
\end{aligned}$$

Since

$$\sum_{k \neq l} \sum_{k, l \in r} \frac{\hat{e}_{mk}}{\hat{p}_k} \frac{\hat{e}_{ml}}{\hat{p}_l} = \left(\sum_r \frac{\hat{e}_{mk}}{\hat{p}_k} \right)^2 - \sum_r \frac{\hat{e}_{mk}^2}{\hat{p}_k^2},$$

$\widehat{V}_{2p}(\hat{t}_{ym}^{\text{GCAL}})$ can be written after some algebra as

$$\widehat{V}_{2p}(\hat{t}_{ym}^{\text{GCAL}}) = \frac{N^2}{n^2} \frac{N-1}{n-1} \left[\frac{n-1}{N-1} - \frac{n}{N} \right] \left(\sum_r \frac{\hat{e}_{mk}}{\hat{p}_k} \right)^2 + \frac{N}{n} \frac{N-1}{n-1} \frac{\hat{e}_{mk}^2}{\hat{p}_k^2} - \frac{N}{n} \frac{\hat{e}_{mk}^2}{\hat{p}_k},$$

that is equivalent to $\widehat{V}_{\text{rev}}(\hat{t}_{ym}^{\text{GCAL}})$ as in equation A.1.

References

- Banca d'Italia, . (2018). Italian household income and wealth in 2016. *Survey on Italian Household Income and Wealth*.
- Bartholomew, D. J., Knott, M., and Moustaki, I. (2011). *Latent Variable Models and Factor Analysis. A unified Approach. 3rd Edition*. Wiley.
- Bartholomew, D. J., Steele, F., Moustaki, I., and Galbraith, J. I. (2002). *The Analysis and Interpretation of Multivariate Data for Social Scientists*. Chapman and Hall/CRC.
- Biemer, P. (2009). Measurement errors in surveys. In Pfeffermann, D. and Rao, C. R., editors, *Handbook of Statistics 29A: Sample Surveys: Design, Methods and Applications*, pages 277–312. Oxford, UK:Elsevier.
- Chang, T. and Kott, P. S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika*, 95:555–571.
- Deville, J.-C. (2000). Generalized calibration and application to weighting for nonresponse. 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.
- Fischer, G. and Molenaar, I. (1995). *Rasch Models: Foundations, Recent Developments and Applications*. Springer and Verlag, New York.
- Folsom, R. E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. In *ASA Proceedings of the Social Statistics Section*, pages 197–202.
- 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.
- Fuller, W. A., Loughin, M. M., and Baker, H. D. (1994). Regression weighting in the presence of nonresponse with application to the 1987-1988 Nationwide Food Consumption Survey. *Survey Methodology*, 20:75–85.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61:215–231.
- 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 Lesage, É. (2016). A discussion of weighting procedures for unit nonresponse. *Journal of Official Statistics*, 32(1):129–145.
- Haziza, D. and Rao, J. N. K. (2006). A nonresponse model approach to inference under imputation for missing survey data. *Survey Methodology*, 32:53–64.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685.
- Jeon, S., Lee, J., Anthony, J. C., and Chungb, H. (2017). Latent class analysis for multiple discrete latent variables: A study on the association between violent behavior drug-using behaviors. *Structural Equation Modeling*, 24(6):911–925.
- Kim, J. K. and Park, H. (2006). Imputation using response probability. *Canadian Journal of Statistics*, 34:171–182.
- Kim, J. K. and Riddles, M. K. (2012). Some theory for propensity-score-adjustment estimators in survey sampling. *Survey Methodology*, 38(2):157–165.
- Kott, P. S. (1994). A note on handling nonresponse in sample surveys. *Journal of the American Statistical Association*, 89(426):693–696.
- 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.
- 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. (2012). Providing double protection for unit nonresponse with a nonlinear calibration-weighting routine. *Survey Research Methods*, 6(2):105–111.
- 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.
- Lazarsfeld, P. and Henry, N. (1968). *Latent structure analysis*. Boston, Houghton Mifflin.
- Lesage, É., Haziza, D., and D’Haultfoeulle, 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.
- Linzer, D. A. and Lewis, J. B. (2011). poLCA: An R package for polytomous variable

- latent class analysis. *Journal of Statistical Software*, 42(10):1–29.
- Linzer, D. A. and Lewis, J. B. (2014). *poLCA: Polytomous variable Latent Class Analysis*. R package version 1.4.1.
- Matei, A. and Ranalli, M. G. (2015). Dealing with non-ignorable nonresponse in survey sampling: A latent modeling approach. *Survey Methodology*, 41(1):145–164.
- 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.
- Neri, A. and Ranalli, M. G. (2011). To misreport or not to report? The measurement of household financial wealth. *Statistics in Transition*, 12-2:281–300.
- Neri, A. and Zizza, R. (2010). Income reporting behaviour in sample surveys. Technical Report Temi di discussione (Economic working papers) 777, Banca d’Italia.
- Oh, H. L. and Scheuren, F. (1983). Weighted adjustment for nonresponse. In Madow, W. G., Nisselson, H., and Olkin, I., editors, *Incomplete Data in Sample Survey*, volume 2, pages 143–184, New York. Academic Press.
- Osier, G. (2013). Dealing with non-ignorable non-response using generalised calibration: a simulation study based on the Luxembourgish Household Budget Survey. Technical report, STATEC Luxembourg, <http://www.statistiques.public.lu/catalogue-publications/economie-statistiques/2013/65-2013.pdf>.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17 (5):1–25.
- Särndal, C.-E. (2007). The calibration approach un survey theory and practice. *Survey Methodology*, 33:99–119.
- Särndal, C.-E. and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. Wiley, New York.
- Särndal, C.-E., Swensson, B., and Wretman, J. H. (1992). *Model Assisted Survey Sampling*. Springer, New York.
- Shao, J. and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association*, 94:254–265.
- Skrondal, A. and Rabe-Hesketh, S. (2007). Latent variable modelling: A survey. *Scandinavian Journal of Statistics*, 34:712–745.
- Tillé, Y. and Matei, A. (2021). *sampling: Survey Sampling*. R package version 2.9.