STATISTICS IN TRANSITION new series, March 2020
Vol. 21, No. 1 pp. 13–35, DOI 10.21307/stattrans-2020-002
Submitted – 09.07.2019; accepted for publishing – 04.02.2020

Parametric prediction of finite population total under Informative sampling and nonignorable nonresponse

Abdulhakeem Eideh[1]

ABSTRACT

In this paper, we combine two methodologies used in the model-based survey sampling, namely the prediction of the finite population total, named T, under informative sampling and full response, see Sverchkov and Pfeffermann (2004), and the prediction of T with a noninformative sampling design and the nonignorable nonresponse mechanism, see Eideh (2012). The former approach involves the dependence of the first order inclusion probabilities on the study variable, while the latter involves the dependence of the probability of nonresponse on unobserved or missing observations. The main aim of the paper is to consider how to account for the joint effects of informative sampling designs and not-missing-at-random response mechanism in statistical models for complex survey data. For this purpose, theoretically, we use the response distribution and relationships between the moments of the superpopoulation, the sample, sample-complement, response, and nonresponse distributions for the prediction of finite population totals, see Eideh (2016). The derived parametric predictors of T use the observation for the response set of the study variable or variable of interest, values of auxiliary variables and their population totals, sampling weights, and propensity scores. An interesting outcome of the T study is that most predictors known from model-based survey sampling can be derived as a special case from this general theory, see Chambers and Clark (2012).

Key words: response distribution, nonignorable nonresponse, informative sampling design.

1. Introduction

Data collected by sample surveys are used extensively to make inferences on assumed population models. Often, survey design features (clustering, stratification, unequal probability selection, etc.) are ignored and the sample data are then analysed using classical methods based on simple random sampling. This approach can, however, lead to erroneous inference because of sample selection bias implied by informative sampling - the sample selection probabilities depend on the values of the

[1] Department of Mathematics,Al-Quds University, Abu-Dees Campus, Al-Quds, Palestine.
E-mail: msabdul@staff.alquds.edu.

model outcome variable (or the model outcome variable is correlated with design variables not included in the model). See Pfeffermann et al. (1998) and Eideh and Nathan (2006). In addition to the effect of complex sample design, one of the major problems in the analysis of survey data is that of missing values or nonresponse. Little and Rubin (2002) consider three types of the nonresponse mechanism or the missing data mechanism:

(a) Missing completely at random (MCAR): if the response probability does not depend on the study variable, or the auxiliary population variable, the missing data are MCAR.

(b) Missing at random (MAR) given auxiliary population variable: if the response probability depends on the auxiliary population variable but not on the study variable, the missing data are MAR.

(c) Not missing at random (NMAR): if the response probability depends on the value of a missing study variable, the missing data are NMAR.

So, the cross-classification of the sampling design and the response mechanism is summarized in the following table:

Table 1.

Sampling Design	Response Mechanism		
	MCAR	MAR	NMAR
Informative – I	IMCAR	IMAR	INMAR
Noninformative – N	NMCAR	NMAR	NNMAR

For inference problem, Little (1982) classifies the nonresponse mechanism as ignorable (MAR and MCAR) and nonignorable (NMAR). In this sense, the cross classification of the sampling design and the nonresponse mechanism is:

Table 2.

Sampling Design	Nonresponse Mechanism	
	Ignorable – i	Nonignorable – n
Informative – i	ii	in
Noninformative – n	ni	nn

For more information about prediction modelling approach for potentially nonignorable nonresponse in complex surveys, see Little (1983, 2003).

Pfeffermann and Sikov (2011), and Eideh (2012) consider estimation of superpopulation parameters and prediction of finite population parameters (census parameters) under nonignorable nonresponse via response and nonresponse distributions when the sampling design in noninformative.

Eideh (2016) treated the estimation of finite population mean and superpopulation parameters when the sampling design is informative and the nonresponse mechanism is nonignorable. Sverchkov and Pfeffermann (2004) use of the sample and complement-sample distributions for the semiparametric prediction of finite population totals under single-stage sampling. None of the above studies consider simultaneously the problem of informative sampling and the problem of nonignorable nonresponse in the prediction of finite population total. In this paper, we study, within a modelling framework, the parametric prediction of finite population total, by specifying the probability distribution of the observed measurements under informative sampling and nonignorable nonresponse. This is the most general situation in surveys and other combinations of sampling informativeness and response mechanisms can be considered as special cases.

It should be pointed here that, according to Sarndal (2011), "Nonresponse causes both bias and increased variance. Its square is typically the dominant portion of the Mean Squared Error (MSE). We address primarily surveys on individuals and households with quite large sample sizes, as is typical for Journal of Official Statistics for government surveys; consequently, the variance contribution to MSE is low by comparison. Increased variance due to nonresponse is nevertheless an issue; striking a balance between variance increase and bias reduction is considered, for example, in Little and Vartivarian (2005)." Furthermore, Brick (2013) mentioned that "Model assumptions and adjustments are made in an attempt to compensate for missing data. Because the mechanisms that cause unit nonresponse are almost never adequately reflected in the model assumptions, survey estimates may be biased even after the model based adjustments. Nonresponse also causes a loss in the precision of survey estimates, primarily due to reduced sample size and secondarily as the result of increased variation of the survey weights. However, bias is the dominant component of the nonresponse-related error in the estimates, and nonresponse bias generally does not decrease as the sample size increases. Thus, bias is often the largest component of mean square error of the estimates even for subdomains when the sample size is large". Here, we focus on the bias, variance and MSE.

The paper is structured as follows. In Section 2 we review the definition of sample, sample-complement, response, and nonresponse distributions, and derive new relationships between their moments. In Section 3, we derive a parametric predictions and their biases of finite population total under informative sampling and not missing at random the nonresponse mechanism. Also, we apply the results for three models, namely: common mean population model, simple ratio population model, and simple regression population model. Finally, Section 4 provides the conclusions.

2. Response, nonresponse distributions and relationships between their moments

Let $U = \{1,..., N\}$ denote a finite population consisting of N units. Let y be the study variable of interest and let y_i be the value of y for the ith population unit. A probability sample s is drawn from U according to a specified sampling design. The sample size is denoted by n. Let $\mathbf{x}_i = x_i = (x_{i1},..., x_{ip})'$, $i \in U$ be the values of a vector of auxiliary variables, $x_1,...,x_p$, and $\mathbf{z} = z = \{z_1,..., z_N\}$ be the values of known design variables, used for the sample selection process not included in the model under consideration. In what follows, we consider a sampling design with selection probabilities $\pi_i = \Pr(i \in s) > 0$, and sampling weight $w_i = 1/\pi_i$; $i = 1,..., N$. In practice, the π_i's may depend on the population values $(\mathbf{x}, \mathbf{y}, \mathbf{z})$. We express this dependence by writing: $\pi_i = \Pr(i \in s \mid \mathbf{x}, \mathbf{y}, \mathbf{z})$ for all units $i \in U$. Denote by $\mathbf{I} = (I_1,..., I_N)'$ the N by 1 sample indicator (vector) variable, such that $I_i = 1$ if unit $i \in U$ is selected to the sample and $I_i = 0$ if otherwise, so that $s = \{i \mid i \in U, I_i = 1\}$ and its complement is $\bar{s} = c = \{i \mid i \in U, I_i = 0\}$. We consider the population values $y_1,..., y_N$ as random variables, which are independent realizations from a distribution with probability density functions (pdf) $f_p(y_i \mid \mathbf{x}_i; \theta)$, indexed by a vector of parameters θ.

In addition to the effect of complex sample design, one of the major problems in the analysis of survey data is that of missing values. In recent articles by Eideh (2009), Pfeffermann and Sikov (2011), and Eideh (2012), the authors defined and studied the problem of nonignorable nonresponse using the response and nonresponse distributions where the sampling design is noninformative. Denote by $R = (R_1,..., R_N)'$ the N by 1 response indicator (vector) variable such that $R_i = 1$ if unit $i \in s$ is observed and $R_i = 0$ if otherwise. We assume that these random variables are independent of one another and of the sample selection mechanism (Oh and Scheuren 1983). The response set is defined accordingly as $r = \{i \in s \mid R_i = 1\}$ and the nonresponse set by $\bar{r} = \{i \in s \mid R_i = 0\}$. We assume probability sampling, so that $\pi_i = \Pr(i \in s) > 0$ for all units $i \in U$. Let the response probability $\psi_i = \Pr(i \in r \mid \mathbf{x}, \mathbf{y}, \mathbf{z}) > 0$ for all units $i \in s$ and $\phi_i = 1/\psi_i$ be the response weight for $i \in s$. Let $O = \{(\mathbf{x}_i, I_i), i \in U\}, \{\pi_i, R_i, i \in s\} \cup \{(y_i, \mathbf{x}_i), i \in r\}$ and $N, n,$ and m be the available information from the sample and response sets.

According to Eideh (2007, 2009), the (marginal) response pdf of y_i is given by:

$$f_r(y_i \mid \mathbf{x}_i, \theta, \eta, \gamma) = \frac{E_s(\psi_i \mid \mathbf{x}_i, y_i, \gamma) f_s(y_i \mid \mathbf{x}_i, \theta, \eta)}{E_s(\psi_i \mid \mathbf{x}_i, \theta, \eta, \gamma)} \tag{1}$$

where the sample pdf of y_i, see Pfeffermann *et al.* (1998), is:

$$\begin{aligned} f_s(y_i \mid \mathbf{x}_i, \theta, \gamma) &= \frac{\Pr(i \in s | \mathbf{x}_i, y_i, \gamma)}{\Pr(i \in s | \mathbf{x}_i, \theta, \gamma)} f_p(y_i \mid \mathbf{x}_i, \theta) \\ &= \frac{E_p(\pi_i \mid \mathbf{x}_i, y_i, \gamma) f_p(y_i \mid \mathbf{x}_i, \theta)}{E_p(\pi_i \mid \mathbf{x}_i, \theta, \gamma)} \end{aligned} \tag{2a}$$

According to Pfeffermann and Sverchkov (1999), we have

$$E_p(y_i \mid \mathbf{x}_i) = \frac{E_s(w_i y_i \mid \mathbf{x}_i)}{E_s(w_i \mid \mathbf{x}_i)} \tag{2b}$$

Combining (1) and (2) we get:

$$f_r(y_i \mid \mathbf{x}_i, \theta, \eta, \gamma) = \frac{E_s(\psi_i \mid \mathbf{x}_i, y_i, \gamma)}{E_s(\psi_i \mid \mathbf{x}_i, \theta, \eta, \gamma)} \frac{E_p(\pi_i \mid \mathbf{x}_i, y_i, \gamma)}{E_p(\pi_i \mid \mathbf{x}_i, \theta, \gamma)} f_p(y_i \mid \mathbf{x}_i, \theta)$$

Furthermore, Sverchkov and Pfeffermann (2004) define the sample-complement pdf of y_i as:

$$f_{\bar{s}}(y_i \mid \mathbf{x}_i, \theta, \gamma) = \frac{E_p(1 - \pi_i \mid \mathbf{x}_i, y_i, \gamma) f_p(y_i \mid \mathbf{x}_i, \theta)}{E_p(1 - \pi_i \mid \mathbf{x}_i, \theta, \gamma)} \tag{3a}$$

and

$$E_{\bar{s}}(y_i \mid \mathbf{x}_i) = \frac{E_s\{(w_i - 1) y_i \mid \mathbf{x}_i\}}{E_s\{(w_i - 1) \mid \mathbf{x}_i\}} \tag{3b}$$

According to Eideh (2007, 2009), the (marginal) nonresponse pdf of y_i is given by:

$$f_{\bar{r}}(y_i \mid \mathbf{x}_i,\theta,\eta,\gamma)=\frac{E_s(1-\psi_i \mid \mathbf{x}_i,y_i,\gamma)f_s(y_i \mid \mathbf{x}_i,\theta,\eta)}{E_s(1-\psi_i \mid \mathbf{x}_i,\theta,\eta,\gamma)}$$
$$=\frac{E_s(1-\psi_i \mid \mathbf{x}_i,y_i,\gamma)}{E_s(1-\psi_i \mid \mathbf{x}_i,\theta,\eta,\gamma)}\frac{E_p(\pi_i \mid \mathbf{x}_i,y_i,\gamma)}{E_p(\pi_i \mid \mathbf{x}_i,\theta,\gamma)}f_p(y_i \mid \mathbf{x}_i,\theta) \tag{4}$$

Furthermore, for vector of random variables $(y_i,\mathbf{x}_i)$, using Eideh (2007,2009, 2016), we have:

$$E_p(y_i \mid x_i)=\frac{E_r(\phi_i w_i y_i \mid x_i)}{E_r(\phi_i w_i \mid x_i)} \tag{5}$$
$$E_s(y_i \mid x_i)=\frac{E_r(\phi_i y_i \mid x_i)}{E_r(\phi_i \mid x_i)} \tag{6}$$
$$E_{\bar{s}}(y_i \mid x_i)=\frac{E_r\{\phi_i(w_i-1)y_i \mid x_i\}}{E_r\{\phi_i(w_i-1)\mid x_i\}} \tag{7}$$
$$E_{\bar{r}}(y_i \mid x_i)=\frac{E_r\{(\phi_i-1)y_i \mid x_i\}}{E_r\{(\phi_i-1)\mid x_i\}} \tag{8}$$

Remark 1. The important feature of the formulas (5-8) is that, given $\{x_i,y_i,\emptyset_i,w_i; i\in r\}$, we can identify $E_p(y_i \mid x_i)$, $E_s(y_i \mid x_i)$, $E_{\bar{s}}(y_i \mid x_i)$, $E_r(y_i \mid x_i)$ and $E_{\bar{r}}(y_i \mid x_i)$.

Remark 2. Note that

$$E_r(y_i \mid \mathbf{x}_i)=E_p\left\{\frac{E_s(\psi_i \mid \mathbf{x}_i,y_i,\gamma)}{E_s(\psi_i \mid \mathbf{x}_i,\theta,\eta,\gamma)}\frac{E_p(\pi_i \mid \mathbf{x}_i,y_i,\gamma)}{E_p(\pi_i \mid \mathbf{x}_i,\theta,\gamma)}y_i \middle| \mathbf{x}_i\right\}\neq E_p(y_i \mid \mathbf{x}_i)$$

Thus, estimating $E_p(y_i \mid \mathbf{x}_i)$ is often the main target of inference, which shows that ignoring an informative sampling scheme or NMAR nonresponse and thus estimating implicitly $E_r(y_i \mid \mathbf{x}_i)$ can bias the inference.

Using the equations (1-8), we can prove the following:

$$f_r(y_i \mid x_i) = \frac{E_r(\phi_i w_i \mid x_i)}{E_r(\phi_i w_i \mid x_i, y_i)} f_p(y_i \mid x_i) \tag{9}$$

$$f_{\bar{r}}(y_i \mid x_i) = \frac{E_r\{(\phi_i - 1) \mid x_i, y_i\}}{E_r\{(\phi_i - 1) \mid x_i\}} f_r(y_i \mid x_i) \tag{10}$$

$$f_{\bar{s}}(y_i \mid x_i) = \frac{E_r\{(\phi_i(w_i - 1)) \mid x_i, y_i\}}{E_r\{(\phi_i(w_i - 1)) \mid x_i\}} f_r(y_i \mid x_i) \tag{11}$$

Furthermore, using (1) and (6), we have

$$f_s(y_i \mid \mathbf{x}_i) = \frac{E_r(\phi_i \mid \mathbf{x}_i, y_i) f_r(y_i \mid \mathbf{x}_i)}{E_r(\phi_i \mid \mathbf{x}_i)} \tag{1*}$$

Remark 2. Once we identify $f_r(y_i \mid x_i)$, we can completely determine the nonresponse distribution $f_{\bar{r}}(y_i \mid x_i)$ and the nonsampled distribution $f_{\bar{s}}(y_i \mid x_i)$. So, instead of specifying $f_p(y_i \mid x_i)$, we can specify $f_r(y_i \mid x_i)$ based on the study variable and the auxiliary variables for the response set.

Estimation of response probabilities ψ_i for all $i \in s$:

If the nonresponse mechanism is not missing at random, then the classical methods for estimating the response probabilities using auxiliary variables, available for respondents and nonrespondents, is logistic or profit models. If we use the logistic model, then

$$\psi_i = \Pr(R_i = 1 \mid i \in s, x_i) = \frac{\exp(\gamma_0 + \gamma_1 x_i)}{1 + \exp(\gamma_0 + \gamma_1 x_i)}$$

We can fit this model using maximum likelihood approach. Thus, the estimate of ψ_i is:

$$\hat{\psi}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)}$$

If the nonresponse mechanism is NMAR, then values of y_i for $i \in r$ are available, but for $i \notin r$ are not available, so we cannot fit the following model:

$$\psi_i = \Pr(R_i = 1 | i \in s, x_i, y_i) = \frac{\exp(\gamma_0 + \gamma_1 x_i + \gamma_2 y_i)}{1 + \exp \exp(\gamma_0 + \gamma_1 x_i + \gamma_2 y_i)} \tag{12}$$

directly using maximum likelihood method. A recent approach of estimation ψ_i under nonignorable nonresponse is discussed by Sverchkov (2008) and Reddles *et al.* (2016).

3. Parametric prediction of finite population parameter under informative sampling and NMAR nonresponse mechanism

Eideh (2012) uses response and nonresponse distribution to derive new predictors of the finite population total, under common mean or homogeneous population model, simple ratio population model, and simple regression population model. These new predictors take into account the nonignorable nonresponse. In this section we extend the prediction problem when, in addition, the sampling design is informative.

3.1 General theory

Assume single-stage population model. Let

$$T = \sum_{i=1}^{N} y_i = \sum_{i \in s} y_i + \sum_{i \in \bar{s}} y_i = \sum_{i \in r} y_i + \sum_{i \in \bar{r}} y_i + \sum_{i \in \bar{s}} y_i \tag{15}$$

be the finite population total that we want to predict using the data from the response set and possibly values of auxiliary variables. Notice that T can be decomposed into three components: the first component represents the total for observed units in the sample – response set, $\sum_{i \in r} y_i$, the second component represents the total for unobserved units in sample – nonresponse set, $\sum_{i \in \bar{r}} y_i$, and the third component represents the total for non-sample units, $\sum_{i \in \bar{s}} y_i$ - nonsample set.

Let $\hat{T} = \hat{T}(O)$ define the predictor of T based on the available information, from the sample and response set $O = \{(I_i), i \in U\}, \{\pi_i, R_i, i \in s\} \cup \{(y_i), i \in r\}$ and

$N, n,$ and m. The mean square error (MSE) of $\hat{T}$ given O with respect to the population pdf is defined by:

$$MSE_p\left(\hat{T}\right) = E_p\left\{\left(\hat{T} - T\right)^2 \mid O\right\} = \left\{\hat{T} - E_p\left(T \mid O\right)\right\}^2 + Var_p\left(T \mid O\right) \tag{16}$$

It is obvious that (16) is minimized when $\hat{T} = E\left(T \mid O\right)$. Hence, the minimum mean squared error best linear unbiased predictor (BLUP) of $T = \sum_{i=1}^{N} y_i$ is given by:

$$\begin{aligned} T^* &= E_p\left(T \mid O\right) = E_p\left\{\left(\sum_{i \in r} y_i + \sum_{i \in \bar{r}} y_i + \sum_{i \in \bar{s}} y_i\right)\middle| O\right\} \\ &= \sum_{i \in r} y_i + \sum_{i \in \bar{r}} E_{\bar{r}}\left(y_i \middle| O\right) + \sum_{i \in \bar{s}} E_{\bar{s}}\left(y_i \middle| O\right) \end{aligned} \tag{17}$$

We know the values $\{y_i\text{'s}, i \in r\}$, so the sum $\sum_{i \in r} y_i$ is known, thus we need to predict the total for unobserved units in the sample – nonresponse set, $\sum_{i \in \bar{r}} y_i$, and the total for non-sample units, $\sum_{i \in \bar{s}} y_i$. That is, to predict T we need to predict values for $\{y_i\text{'s}, i \in \bar{r}\}$ and $\{y_i\text{'s}, i \in \bar{s}\}$. Thus, our aim is to identify an optimal predictor of $\sum_{i \in \bar{s}} y_i$ and $\sum_{i \in \bar{r}} y_i$ based on the given observed data O.

The predictor given in (17) represents the prediction of T for single-stage sampling when the sampling mechanism in informative and missing value mechanism is NMAR. The analysis that follows assumes known model parameters. In practice, the unknown model parameters are replaced under the frequentist approach by sample estimates, yielding the corresponding "empirical predictors." In the present case, maximum likelihood estimation of the model parameters must be based on the response distribution of the observed units in the sample – response set, see Eideh (2016).

Using (7) and (8), equation (17) can be written as:

$$T^* = \sum_{i \in r} y_i + \sum_{i \in \bar{r}} \frac{E_r\left\{\left(\phi_i - 1\right) y_i\right\}}{E_r\left\{\left(\phi_i - 1\right)\right\}} + \sum_{i \in \bar{s}} \frac{E_r\left\{\phi_i\left(w_i - 1\right) y_i\right\}}{E_r\left\{\phi_i\left(w_i - 1\right)\right\}} \tag{18}$$

Hence, T^* can be estimated based only on the data in the response set, $\{y_i, \phi_i, w_i : i \in r\}$. Using method of moments estimates technique, that is, replace the moment under the response distribution by the average over the response set, for example $\hat{E}_r(a_i) = m^{-1}\sum_{i\in r} a_i$, where m is size of the response set r.

From now on, we use the following notations for the predictor of $T = \sum_{i=1}^{N} y_i$:

- T_{in}^* - Best linear unbiased predictor of T when the sampling design is informative and the nonresponse mechanism is NMAR (nonignorable).
- T_{ii}^* - Best linear unbiased predictor of T when the sampling design is informative and the nonresponse mechanism is ignorable.
- T_{nn}^* - Best linear unbiased predictor of T when the sampling design is noninformative and the nonresponse mechanism is nonignorable.
- T_{ni}^* - Best linear unbiased predictor of T when the sampling design is noninformative and the nonresponse mechanism is ignorable.

According to (18) and using the method of moments estimator, we can show that the best linear unbiased predictor for T is:

$$\begin{aligned}\hat{T}_{in}^* &= \sum_{i\in r} y_i + (n-m)\frac{\sum_{i\in r}(\phi_i - 1)y_i}{\sum_{i\in r}(\phi_i - 1)} + (N-n)\frac{\sum_{i\in r}\phi_i(w_i - 1)y_i}{\sum_{i\in r}\phi_i(w_i - 1)} \\ &= \sum_{i\in r} w_i^{in} y_i\end{aligned} \tag{19}$$

where

$$w_i^{in} = 1 + (n-m)\frac{(\phi_i - 1)}{\sum_{i\in r}(\phi_i - 1)} + (N-n)\frac{\phi_i(w_i - 1)}{\sum_{i\in r}\phi_i(w_i - 1)} \tag{20}$$

Note that

(a) $\sum_{i\in r}(\phi_i - 1)y_i$ is the Horvitz-Thompson estimator of $\sum_{i\in \bar{r}} y_i$.

(b) $\sum_{i\in r}\phi_i(w_i - 1)y_i$ is the Horvitz-Thompson estimator of $\sum_{i\in \bar{s}} y_i$.

(c) $\frac{n-m}{\sum_{i\in r}(\phi_i - 1)}$ is the "Hajek type correction" for controlling the variability of the response weights.

(d) $\frac{(N-n)}{\sum_{i\in r}\phi_i(w_i - 1)}$ is the "Hajek type correction" for controlling the variability of the product of the response weights and the sampling weights.

It is easy to verify tha

a) Under noninformative sampling design and nonignorable nonresponse:

$$w_i^{nn} = 1 + (n-m)\frac{(\phi_i - 1)}{\sum_{i\in r}(\phi_i - 1)} + (N-n)\frac{\phi_i}{\sum_{i\in r}\phi_i}$$

(b) Under noninformative sampling design and ignorable nonresponse:

$$w_i^{ni} = 1 + \frac{(n-m)}{m} + \frac{(N-n)}{m} = \frac{N}{m}$$

(c) Under informative sampling design and ignorable nonresponse:

$$w_i^{ii} = 1 + \frac{(n-m)}{m} + (N-n)\frac{w_i - 1}{\sum_{i\in r}(w_i - 1)}$$

According to (1-8), we can write (17) as:

$$\begin{aligned} T_{in}^{*} &= \sum_{i\in r} y_i + \sum_{i\in \bar{r}}\left\{E_s(y_i|O) - \frac{Cov_s[(\psi_i, y_i)|O]}{1-E_s(\psi_i|O)}\right\} + \sum_{i\in \bar{s}} E_p(y_i|O) - \frac{Cov_p[(\pi_i, y_i)|O]}{1-E_p(\pi_i|O)} \\ &= \sum_{i\in r} y_i + \sum_{i\in \bar{r}} E_s(y_i|O) + \sum_{i\in \bar{s}} E_p(y_i|O) - \left\{\sum_{i\in \bar{r}}\frac{Cov_s[(\psi_i, y_i)|O]}{1-E_s(\psi_i|O)} + \sum_{i\in \bar{s}}\frac{Cov_p[(\pi_i, y_i)|O]}{1-E_p(\pi_i|O)}\right\} \end{aligned} \quad (21)$$

Using (1-8), we can show that the prediction nonresponse bias of T_{in} is:

$$B\left(T_{in}^{*}\right)=E_{p}\left(T_{in}^{*}-T\right)=-\left\{\sum_{i\in\bar{r}}E_{p}\left[y_{i}-E_{\bar{r}}\left(y_{i}\right)\right]+\sum_{i\in\bar{s}}E_{p}\left[y_{i}-E_{\bar{s}}\left(y_{i}\right)\right]\right\}$$
$$=-\left\{\sum_{i\in\bar{r}}\left[E_{p}\left(y_{i}\right)-E_{\bar{r}}\left(y_{i}\right)\right]+\sum_{i\in\bar{s}}\left[E_{p}\left(y_{i}\right)-E_{\bar{s}}\left(y_{i}\right)\right]\right\}$$
$$=-\left\{\sum_{i\in\bar{r}}\left[\left(E_{p}\left(y_{i}\right)-E_{s}\left(y_{i}\right)\right)+\frac{Cov_{s}\left(\psi_{i},y_{i}\right)}{E_{s}\left(1-\psi_{i}\right)}\right]+\sum_{i\in\bar{s}}\frac{Cov_{p}\left(\pi_{i},y_{i}\right)}{1-E_{p}\left(\pi_{i}\right)}\right\} \tag{22}$$

Therefore, the predictor T_{in}^{*} in (22) is unbiased T if:

(a) $Cov_{s}\left(\psi_{i},y_{i}\right)=0$, (or $Cov_{r}\left(\phi_{i},y_{i}\right)=0$), that is, there is no correlation between the study variable and the response probabilities ψ_{i}, consequently, the nonresponse mechanism is ignorable, and

(b) $Cov_{p}\left(\pi_{i},y_{i}\right)=0$, (or $E_{r}\left(\phi_{i}\right)E_{r}\left(\phi_{i}w_{i}y_{i}\right)=E_{r}\left(\phi_{i}y_{i}\right)E_{r}\left(\phi_{i}w_{i}\right)$), that is, there is no correlation between the study variable and the first order inclusion probabilities π_{i}, so the sampling design is noninformative.

If (a) is satisfied then (b) becomes $Cov_{r}\left(\phi_{i}w_{i},y_{i}\right)=0$. In other words, if the sampling design is noninformative and the response mechanism is ignorable, so that $E_{p}\left(y_{i}\right)=E_{s}\left(y_{i}\right)=E_{\bar{s}}\left(y_{i}\right)=E_{\bar{r}}\left(y_{i}\right)=E_{r}\left(y_{i}\right)$, then T_{in}^{*} in unbiased of T .

Note that the stronger the relationship between the study variable and the response probability, and the study variable and the first order inclusion probabilities, the larger the bias.

According to (10) and (11), we can show that (22) has the following form:

$$B\left(T_{in}^{*}\right)=-\left\{\begin{array}{l}\sum_{i\in\bar{r}}\left\{-\frac{Cov_{r}\left(\phi_{i}w_{i},y_{i}\right)}{E_{r}\left(\phi_{i}w_{i}\right)E_{r}\left\{\left(\phi_{i}-1\right)\right\}}+\frac{E_{r}\left(\phi_{i}w_{i}y_{i}\right)E_{r}\left(\phi_{i}\right)-E_{r}\left(\phi_{i}y_{i}\right)E_{r}\left(\phi_{i}w_{i}\right)}{E_{r}\left(\phi_{i}w_{i}\right)E_{r}\left\{\left(\phi_{i}-1\right)\right\}}\right\}\\ -\sum_{i\in\bar{s}}\frac{E_{r}\left(\phi_{i}\right)E_{r}\left(\phi_{i}w_{i}y_{i}\right)-E_{r}\left(\phi_{i}y_{i}\right)E_{r}\left(\phi_{i}w_{i}\right)}{E_{r}\left(\phi_{i}w_{i}\right)\left[E_{r}\left(\phi_{i}w_{i}\right)-E_{r}\left(\phi_{i}\right)\right]}\end{array}\right\} \tag{23}$$

Hence, the bias $B\left(T_{in}^{*}\right)$ can be estimated based only on the data in the response set, $\left\{y_i, \phi_i, w_i : i \in r\right\}$, using method of moments estimates technique, that is, replace the moment under the response distribution by the average over the response set, for example $\hat{E}_r\left(a_i\right) = m^{-1} \sum_{i \in r} a_i$.

To test the informativeness of the sampling design, see Pfeffermann and Sverchkov (1999) and Eideh and Nathan (2006). Moreover, for testing the ignorability of the nonresponse mechanism, see Eideh (2012).

Particular cases:

Case 1: The sampling design is noninformative and the nonresponse process is nonignorable, so that:

$$E_s\left(y_i \middle| O\right) = E_p\left(y_i \middle| O\right) \text{ and } Cov_p\left[\left(\pi_i, y_i\right)\middle| O\right] = 0$$

Therefore,

$$T_{nn}^{*} = \sum_{i \in r} y_i + \sum_{i \in \bar{r}} E_p\left(y_i \middle| O\right) + \sum_{i \in \bar{s}} E_p\left(y_i \middle| O\right) - \sum_{i \in \bar{r}} \frac{Cov_p\left[\left(\psi_i, y_i\right)\middle| O\right]}{1 - E_p\left(\psi_i \middle| O\right)} \tag{24}$$

$$B\left(T_{nn}^{*}\right) = -\sum_{i \in \bar{r}} \left\{ -\frac{Cov_r\left(\phi_i w_i, y_i\right)}{E_r\left(\phi_i w_i\right) E_r\left\{\left(\phi_i - 1\right)\right\}} + \frac{E_r\left(\phi_i w_i y_i\right) E_r\left(\phi_i\right) - E_r\left(\phi_i y_i\right) E_r\left(\phi_i w_i\right)}{E_r\left(\phi_i w_i\right) E_r\left\{\left(\phi_i - 1\right)\right\}} \right\} \tag{25}$$

Case 2: The sampling design is noninformative and the nonresponse process is ignorable, so that:

$$E_r\left(y_i \middle| O\right) = E_s\left(y_i \middle| O\right) = E_p\left(y_i \middle| O\right),\ Cov_p\left[\left(\pi_i, y_i\right)\middle| O\right] = 0 \text{ and } Cov_s\left[\left(\psi_i, y_i\right)\middle| O\right] = 0$$

Therefore,

$$T_{ni}^{*} = \sum_{i \in r} y_i + \sum_{i \in \bar{r}} E_p\left(y_i \middle| O\right) + \sum_{i \in \bar{s}} E_p\left(y_i \middle| O\right) \tag{26}$$

$$B\left(T_{ni}^{*}\right) = -\left\{ \sum_{i \in \bar{r}} \left[E_p\left(y_i\right) - E_{\bar{r}}\left(y_i\right)\right] + \sum_{i \in \bar{s}} \left[E_p\left(y_i\right) - E_{\bar{s}}\left(y_i\right)\right] \right\} = 0 \tag{27}$$

Case 3: The sampling design is informative and the nonresponse process is ignorable, so that:

$$E_r(y_i|O) = E_s(y_i|O) \text{ and } Cov_s[(\psi_i, y_i)|O] = 0$$

Therefore,

$$T_{ii}^* = \sum_{i \in r} y_i + \sum_{i \in \bar{r}} E_s(y_i|O) + \sum_{i \in \bar{s}} E_p(y_i|O) - \sum_{i \in \bar{s}} \frac{Cov_p[(\pi_i, y_i)|O]}{1 - E_p(\pi_i|O)} \tag{28}$$

$$B(T_{ii}^*) = -\left\{ \begin{array}{l} \sum_{i \in \bar{r}} \frac{E_r(\phi_i)E_r(\phi_i w_i y_i) - E_r(\phi_i y_i)E_r(\phi_i w_i)}{E_r(\phi_i)E_r(\phi_i w_i)} - \\ \sum_{i \in \bar{s}} \frac{E_r(\phi_i)E_r(\phi_i w_i y_i) - E_r(\phi_i y_i)E_r(\phi_i w_i)}{E_r(\phi_i w_i)[E_r(\phi_i w_i) - E_r(\phi_i)]} \end{array} \right\} \tag{29}$$

3.2. Common mean or homogeneous population model

Chambers and Clark (2012) studied the homogeneous population model under noninformative sampling design. In this section, we will treat in details the homogeneous population model under informative sampling design and nonignorable nonresponse or informative nonresponse mechanism.

Assume that $y_i \underset{p}{\sim} N(\mu, \sigma^2)$, $i = 1, \ldots, N$ are independent normal random variables, with mean $E_p(y_i) = \mu$ and variance $V_p(y_i) = \sigma^2$. According to equation (17), the best linear unbiased predictor T_{in} of T requires the computation of $E_{\bar{r}}(y_i|O)$ and $E_{\bar{s}}(y_i|O)$, and this based on the specification of $E_p(\pi_i|y_i)$ and $E_s(\psi_i|y_i)$. Different models can be considered for $E_p(\pi_i|y_i)$ and $E_s(\psi_i|y_i)$, see Eideh (2003, 2012). In this paper, for illustration, we consider the following models:

(a) Exponential inclusion probability model:

$$E_p(\pi_i|y_i) = \exp(\eta y_i) \tag{30}$$

(b) Exponential response probability model:

$$E_s(\psi_i|y_i) = \exp(\gamma y_i) \tag{31}$$

According to equations (1) and (2), we can show that the sample distribution of y_i is $y_i \underset{s}{\sim} N(\mu+\eta\sigma^2,\sigma^2)$, $i=1,\ldots,n$, and the response distribution of y_i is: $y_i \underset{r}{\sim} N(\mu+(\eta+\gamma)\sigma^2,\sigma^2)$, $i=1,\ldots,m$. For estimation of μ,σ^2,η and γ; see Eideh (2016).

Computation of $E_{\bar{s}}(y_i)$. After some algebra we can show that

$$E_p(\pi_i)=E_p(E_p(\pi_i|y_i))=E_p\{\exp(\eta y_i)\}=M_p(\eta)=\exp\left(\eta\mu+\frac{\eta^2\sigma^2}{2}\right) \tag{32}$$

$$\begin{aligned} E_p(y_i\pi_i)&=E_p(E_p(\pi_i y_i|y_i))=E_p(y_iE_p(\pi_i|y_i)) \\ &=E_p\{y_i\exp(\eta y_i)\}=\frac{d}{d\eta}M_p(\eta) \\ &=(\mu+\eta\sigma^2)\exp\left(\mu\eta+\frac{\eta^2\sigma^2}{2}\right)=E_s(y_i)M_p(\eta) \end{aligned} \tag{33}$$

where $M_p(\eta)$ is the moment generation function of y_i,

$M_p(\eta)=\exp\left(\mu\eta+\frac{\eta^2\sigma^2}{2}\right)$. So that,

$$Cov_p(\pi_i,y_i)=M_p(\eta)\{E_s(y_i)-E_p(y_i)\}=(\eta\sigma^2)M_p(\eta) \tag{34}$$

$$\frac{Cov_p(\pi_i,y_i)}{E_p(1-\pi_i)}=(\eta\sigma^2)\frac{M_p(\eta)}{1-M_p(\eta)} \tag{35}$$

Hence, according to (3b), we have:

$$E_{\bar{s}}(y_i)=\mu-(\eta\sigma^2)\frac{M_p(\eta)}{1-M_p(\eta)} \tag{36}$$

Computation of $E_{\bar{r}}(y_i)$. Similarly, according to (8), we can show that:

$$E_{\bar{r}}(y_i)=\mu+\eta\sigma^2-(\gamma\sigma^2)\frac{M_s(\gamma)}{1-M_s(\gamma)} \tag{37}$$

where, $M_s(\gamma)=\exp\left(\gamma(\mu+\eta\sigma^2)+\frac{\gamma^2\sigma^2}{2}\right)$.

Thus, using (17), (36) and (37), the BLUP for T under informative sampling and nonignorable nonresponse is:

$$T^*_{in,C} = T^*_{ni,C} + \left\{ (n-m)\left(\eta\sigma^2 - \left(\gamma\sigma^2\right)\frac{M_s(\gamma)}{1-M_s(\gamma)} \right) + (N-n)\left(-\left(\eta\sigma^2\right)\frac{M_p(\eta)}{1-M_p(\eta)} \right) \right\} \tag{38}$$

where

$$T^*_{ni,C} = \sum_{i\in r} y_i + (n-m)\mu + (N-n)\mu = \sum_{i\in r} y_i + (N-m)\mu \tag{39}$$

And the nonresponse bias of $T_{in,C}$ is:

$$\begin{aligned} B\left(T^*_{in,C}\right) &= E_p\left(T^*_{in,C} - T\right) = -\left\{ \sum_{i\in\bar{r}} E_p\left[y_i - E_{\bar{r}}(y_i)\right] + \sum_{i\in\bar{s}} E_p\left[y_i - E_{\bar{s}}(y_i)\right] \right\} \\ &= -\left\{ (n-m)\left[-\eta\sigma^2 + \left(\gamma\sigma^2\right)\frac{M_s(\gamma)}{1-M_s(\gamma)} \right] + (N-n)\left[\left(\eta\sigma^2\right)\frac{M_p(\eta)}{1-M_p(\eta)} \right] \right\} \end{aligned} \tag{40}$$

Particular cases:

Case 1: The sampling design is noninformative, that is ($\eta = 0$) and the nonresponse process is nonignorable:

$$T^*_{nn,C} = \sum_{i\in r} y_i + (n-m)\mu + (N-n)\mu + (n-m)\left\{ -\left(\gamma\sigma^2\right)\frac{M_p(\gamma)}{1-M_p(\gamma)} \right\} \tag{41}$$

$$B\left(T^*_{nn,C}\right) = -\left\{ (n-m)\left(\gamma\sigma^2\right)\frac{M_p(\gamma)}{1-M_p(\gamma)} \right\} \tag{42}$$

Case 2: The sampling design is noninformative, that is ($\eta = 0$) and the nonresponse process is ignorable, that is ($\gamma = 0$):

$$T^*_{ni,C} = \sum_{i\in r} y_i + (n-m)\mu + (N-n)\mu = \sum_{i\in r} y_i + (N-m)\mu \tag{43}$$

$$B\left(T^*_{ni,C}\right) = 0 \tag{44}$$

Case 3: The sampling design is informative and the nonresponse process is ignorable, that is ($\gamma = 0$):

$$T_{ii,C}^{*} = \sum_{i \in r} y_i + (n-m)\mu + (N-n)\mu + \left\{ (n-m)\eta\sigma^2 + (N-n)\left(-\left(\eta\sigma^2\right)\frac{M_p(\eta)}{1-M_p(\eta)} \right) \right\} \tag{45}$$

$$B\left(T_{ii,C}^{*}\right) = -\left\{ (n-m)\left[-\eta\sigma^2\right] + (N-n)\left[\left(\eta\sigma^2\right)\frac{M_p(\eta)}{1-M_p(\eta)} \right] \right\} \tag{46}$$

3.3. Simple ratio population model

The simple ratio population model ® stating that: $y_i | x_i \underset{p}{\sim} N\left(\beta x_i, \sigma^2 x_i\right)$, $i = 1,\ldots,N$ are independent normal random variables, with mean $E_p\left(y_i | x_i\right) = \beta x_i$ and variance $Var_p\left(y_i | x_i\right) = \sigma^2 x_i$.

Under the exponential inclusion probability model:

$$E_p\left(\pi_i | y_i, x_i\right) = \exp\left(\eta_0 x_i + \eta_1 y_i\right) \tag{47}$$

And the exponential response probability model:

$$E_s\left(\psi_i | y_i, x_i\right) = \exp\left(\gamma_0 x_i + \gamma_1 y_i\right) \tag{48}$$

Similarly to the previous section, we can show the following: $y_i \mid x_i \underset{s}{\sim} N\left(\left(\eta_1\sigma^2 + \beta\right)x_i, \sigma^2 x_i\right)$, $i = 1,\ldots,n$ and $y_i \mid x_i \underset{r}{\sim} N\left(\left(\eta_1\sigma^2 + \gamma_1\sigma^2 + \beta\right)x_i, \sigma^2 x_i\right)$, $i = 1,\ldots,m$.

$$E_{\bar{s}}\left(y_i\right) = \beta x_i - \left(\eta_1\sigma^2 x_i\right)\frac{\exp\left(\eta_0 x_i\right)M_p\left(\eta_1\right)}{1-\exp\left(\eta_0 x_i\right)M_p\left(\eta_1\right)} \tag{49}$$

where

$$M_p\left(\eta_1\right) = \exp\left(\eta_1\left(\beta x_i\right) + \frac{\sigma^2 x_i \eta_1^2}{2} \right)$$

$$E_{\bar{r}}\left(y_i\right) = \left(\beta + \eta_1\sigma^2\right)x_i - \left(\gamma_1\sigma^2 x_i\right)\frac{\exp\left(\gamma_0 x_i\right)M_s\left(\gamma_1\right)}{1-\exp\left(\gamma_0 x_i\right)M_s\left(\gamma_1\right)} \tag{50}$$

where

$$M_s(\gamma_1) = \exp\left(\gamma_1(\beta + \eta_1\sigma^2)x_i + \frac{\sigma^2 x_i \eta_1^2}{2}\right)$$

Hence, the BLUP for T under informative sampling and nonignorable nonresponse is:

$$T_{in,R}^* = \sum_{i\in r} y_i + (\sum_{i\in U} x_i - \sum_{i\in r} x_i)\beta + \sum_{i\in\bar{r}}\left\{\eta_1\sigma^2 x_i - (\gamma_1\sigma^2 x_i)\frac{\exp(\gamma_0 x_i)M_s(\gamma_1)}{1-\exp(\gamma_0 x_i)M_s(\gamma_1)}\right\} + \sum_{i\in\bar{s}}\left\{-(\eta_1\sigma^2 x_i)\frac{\exp(\eta_0 x_i)M_p(\eta_1)}{1-\exp(\eta_0 x_i)M_p(\eta_1)}\right\} \tag{51}$$

And the nonresponse bias of $T_{in,R}^*$ is:

$$B(T_{in,R}^*) = -\left\{\begin{array}{l}\sum_{i\in\bar{r}}\left[-(\eta_1\sigma^2 x_i) + (\gamma_1\sigma^2 x_i)\frac{\exp(\gamma_0 x_i)M_s(\gamma_1)}{1-\exp(\gamma_0 x_i)M_s(\gamma_1)}\right] + \\ \sum_{i\in\bar{s}}\left[(\eta_1\sigma^2 x_i)\frac{\exp(\eta_0 x_i)M_p(\eta_1)}{1-\exp(\eta_0 x_i)M_p(\eta_1)}\right]\end{array}\right\} \tag{52}$$

Particular cases:

Case 1: The sampling design is noninformative, that is ($\eta = 0$) and the nonresponse process is nonignorable:

$$T_{nn,R}^* = \sum_{i\in r} y_i + (\sum_{i\in U} x_i - \sum_{i\in r} x_i)\beta + \sum_{i\in\bar{r}}\left\{-(\gamma_1\sigma^2 x_i)\frac{\exp(\gamma_0 x_i)M_p(\gamma_1)}{1-\exp(\gamma_0 x_i)M_p(\gamma_1)}\right\} \tag{53}$$

$$B(T_{nn,R}^*) = -\sum_{i\in\bar{r}}\left[(\gamma_1\sigma^2 x_i)\frac{\exp(\gamma_0 x_i)M_p(\gamma_1)}{1-\exp(\gamma_0 x_i)M_p(\gamma_1)}\right] \tag{54}$$

Case 2: The sampling design is noninformative, that is ($\eta = 0$) and the nonresponse process is ignorable, that is ($\gamma = 0$):

$$T^{*}_{ni,R} = \sum_{i \in r} y_i + (\sum_{i \in U} x_i - \sum_{i \in r} x_i)\beta \tag{55}$$

$$B(T_{ni,R}) = 0 \tag{56}$$

Case 3: The sampling design is informative and the nonresponse process is ignorable, that is ($\gamma = 0$):

$$T^{*}_{ii,R} == \sum_{i \in r} y_i + (\sum_{i \in U} x_i - \sum_{i \in r} x_i)\beta + \sum_{i \in \bar{s}} \left\{ -(\eta_1 \sigma^2 x_i) \frac{\exp(\eta_0 x_i) M_p(\eta_1)}{1 - \exp(\eta_0 x_i) M_p(\eta_1)} \right\} \tag{57}$$

$$B(T^{*}_{ii,R}) = -\left\{ \sum_{i \in \bar{r}} \left[-(\eta_1 \sigma^2 x_i) \right] + \sum_{i \in \bar{s}} \left[(\eta_1 \sigma^2 x_i) \frac{\exp(\eta_0 x_i) M_p(\eta_1)}{1 - \exp(\eta_0 x_i) M_p(\eta_1)} \right] \right\} \tag{58}$$

3.4. Simple regression population model

The simple regression population model (L) stating that: $y_i | x_i \underset{p}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = 1, \ldots, N$ are independent normal random variables, with mean $E_p(y_i | x_i) = \beta_0 + \beta_1 x_i$ and variance $Var_p(y_i | x_i) = \sigma^2$. Assuming the models given in (48) and (49), we can show the following:

$$y_i \mid x_i \underset{s}{\sim} N(\beta_0 + \eta_1 \sigma^2 + \beta_1 x_i, \sigma^2),\ i = 1, \ldots, n \tag{59}$$

$$y_i \mid x_i \underset{r}{\sim} N(\beta_0 + \eta_1 \sigma^2 + \gamma_1 \sigma^2 + \beta_1 x_i, \sigma^2),\ i = 1, \ldots, m \tag{60}$$

$$E_{\bar{s}}(y_i) = \beta_0 + \beta_1 x_i - (\eta_1 \sigma^2) \frac{\exp(\eta_0 x_i) M_p(\eta_1)}{1 - \exp(\eta_0 x_i) M_p(\eta_1)} \tag{61}$$

where

$$M_p(\eta_1) = \exp\left(\eta_1 (\beta_0 + \beta_1 x_i) + \frac{\eta_1^2 \sigma^2}{2} \right)$$

$$E_{\bar{r}}(y_i) = \beta_0 + \beta_1 x_i + \eta_1 \sigma^2 - (\gamma_1 \sigma^2) \frac{\exp(\gamma_0 x_i) M_s(\gamma_1)}{1 - \exp(\gamma_0 x_i) M_s(\gamma_1)} \tag{62}$$

where

$$M_s(\gamma_1) = \exp\left(\gamma_1\left(\beta_0 + \beta_1 x_i + \eta_1\sigma^2\right) + \frac{\eta_1^2\sigma^2}{2}\right)$$

Therefore, the BLUP for T under informative sampling and nonignorable nonresponse is:

$$T_{in,L}^* = \sum_{i\in r} y_i + (N-m)\beta_0 + \beta_1\left(\sum_{i\in U} x_i - \sum_{i\in r} x_i\right) + \\ \sum\nolimits_{i\in\bar{r}}\left\{\eta_1\sigma^2 - \left(\gamma_1\sigma^2\right)\frac{\exp(\gamma_0 x_i)M_s(\gamma_1)}{1-\exp(\gamma_0 x_i)M_s(\gamma_1)}\right\} + \sum\nolimits_{i\in\bar{s}}\left\{-\left(\eta_1\sigma^2\right)\frac{\exp(\eta_0 x_i)M_p(\eta_1)}{1-\exp(\eta_0 x_i)M_p(\eta_1)}\right\} \tag{63}$$

And the nonresponse bias of $\hat{T}_{in,L}$ is:

$$B\left(T_{in,L}^*\right) = -\left\{\begin{array}{l}\sum_{i\in\bar{r}}\left[-\eta_1\sigma^2 + \left(\gamma_1\sigma^2\right)\frac{\exp(\gamma_0 x_i)M_s(\gamma_1)}{1-\exp(\gamma_0 x_i)M_s(\gamma_1)}\right] + \\ \sum_{i\in\bar{s}}\left[\left(\eta_1\sigma^2\right)\frac{\exp(\eta_0 x_i)M_p(\eta_1)}{1-\exp(\eta_0 x_i)M_p(\eta_1)}\right]\end{array}\right\} \tag{64}$$

Particular cases:

Case 1: The sampling design is noninformative, that is ($\eta_1 = 0$) and the nonresponse process is nonignorable:

$$T_{nn,L}^* = \sum_{i\in r} y_i + (N-m)\beta_0 + \beta_1\left(\sum_{i\in U} x_i - \sum_{i\in r} x_i\right) + \sum_{i\in\bar{r}}\left\{-\left(\gamma_1\sigma^2\right)\frac{\exp(\gamma_0 x_i)M_p(\gamma_1)}{1-\exp(\gamma_0 x_i)M_p(\gamma_1)}\right\} \tag{65}$$

$$B\left(T_{nn,L}^*\right) = -\sum_{i\in\bar{r}}\left[\left(\gamma_1\sigma^2\right)\frac{\exp(\gamma_0 x_i)M_p(\gamma_1)}{1-\exp(\gamma_0 x_i)M_p(\gamma_1)}\right] \tag{66}$$

Case 2: The sampling design is noninformative, that is ($\eta_1 = 0$) and the nonresponse process is ignorable, that is ($\gamma_1 = 0$):

$$T^*_{ni,L} = \sum_{i\in r} y_i + (N-m)\beta_0 + \beta_1\left(\sum_{i\in U} x_i - \sum_{i\in r} x_i\right) \tag{67}$$

$$B\left(T^*_{ni,L}\right) = 0$$

Case 3: The sampling design is informative and the nonresponse process is ignorable, that is ($\gamma_1 = 0$):

$$\begin{aligned} T^*_{ii,L} &= \sum_{i\in r} y_i + (N-m)\beta_0 + \beta_1\left(\sum_{i\in U} x_i - \sum_{i\in r} x_i\right) + \eta_1\sigma^2(n-m) + \\ &\sum_{i\in\bar{s}}\left\{-\left(\eta_1\sigma^2\right)\frac{\exp(\eta_0 x_i)M_p(\eta_1)}{1-\exp(\eta_0 x_i)M_p(\eta_1)}\right\} \end{aligned} \tag{68}$$

$$B\left(T^*_{ii,L}\right) = -\left\{\sum_{i\in\bar{r}} -\eta_1\sigma^2 + \sum_{i\in\bar{s}}\left[\left(\eta_1\sigma^2\right)\frac{\exp(\eta_0 x_i)M_p(\eta_1)}{1-\exp(\eta_0 x_i)M_p(\eta_1)}\right]\right\} \tag{69}$$

Remark: Empirical BLUP of T . In practice the parameters are unknown, so in order to obtain the empirical best unbiased predictors, we replace the unknown parameters by their estimates, see Eideh (2016).

4. Conclusions

In this paper we combine two methodologies used in the model-based survey sampling: the prediction of finite population total T, under informative sampling, and full response, and the prediction of T when the sampling design is noninformative and the nonresponse mechanism is nonignorable. One incorporates the dependence of the first order inclusion probabilities on the study variable, while the other incorporates the dependence of the probability of nonresponse on unobserved or missing observations. For this aim, we use the response distribution and relationships between moments of the superpopulation, sample, sample-complement, response, and non-response distributions, for the prediction of finite population totals. Common best linear unbiased predictors derived under model-based survey sampling are shown to be special cases of the present theory. The general theory was applied for a homogeneous population model, simple ratio population model, simple linear regression population model, and multiple regression population model. The paper is purely mathematical and focuses on the role of informativeness of the sampling design and informativeness of nonresponse in adjusting various predictors for bias reduction. Further experimentation (simulation and real data problem) with this kind of predictors is therefore highly recommended. We hope that the new mathematical results obtained will encourage further theoretical, empirical and practical research in these directions.

Acknowledgements

The research was partially supported by a grant from DAAD (Deutscher Akademischer Austauschdienst German Academic Exchange Service) - Research Stays for University Academics and Scientists, 2018. Also, the author would like to thank Professor Timo Schmid for hosting the author during his visit to Free University Berlin.

REFERENCES

NONRESPONSE and WEIGHTING ADJUSTMENTS: A Critical Review, Journal of Official Statistics 29, pp. 329–353.

CHAMBERS, R. L., CLARK, R. G., (2012). An Introduction to Model-Based Survey Sampling with Applications, Oxford Statistical Science Series.

EIDEH, A. H., (2016). Estimation of Finite Population Mean and Superpopulation Parameters when the Sampling Design is Informative and Nonresponse Mechanism is Nonignorable. Pakistan Journal of Statistics and Operation Research (PJSOR), Pak.j.stat.oper.res. Vol. XII No. 3, 2016, pp. 467–489.

EIDEH A. H., (2012). Estimation and Prediction under Nonignorable Nonresponse via Response and Nonresponse Distributions, Journal of Indian Society of Agriculture Statistics., 66(3) 2012, pp 359–380.

EIDEH A. H., (2009). On the use of the sample distribution and sample likelihood for inference under informative probability sampling. DIRASAT (Natural Science), Volume 36 (2009), Number 1, pp18–29.

EIDEH, A. H., (2007). Method of Moments Estimators of Finite Population Parameters in Case of Full Response and Nonresponse, Contributed Paper for the 56th Biennial Session of the International Statistical Institute, August 22–29, 2007, Lisboa, Portugal, ISI 2007 Book of Abstracts, p 430.

EIDEH, A. H., NATHAN, G., (2006). Fitting time series models for longitudinal survey data under informative sampling, Journal of Statistical Planning and Inference 136, 9, pp 3052–306, [Corrigendum, 137 (2007), p 628].

LITTLE, R. J. A., (1983). Superpopulation models for nonresponse. In Incomplete Data in Sample Surveys, Vol. 2, Theory and bibliographies, part VI, pp. 337–413. New York, Academic Press.

LITTLE, R. J. A., (2003). Bayesian methods for unit and item nonresponse, In Analysis of survey data, Chambers R.L. and Skinner C.L., pp. 289–306, Wiley, New York.

LITTLE, R. J. A., RUBIN, D. B., (2002). Statistical analysis with missing data. New York: Wiley.

LITTLE, R.J.A., VARTIVARIAN, S., (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means?, Survey Methodology, Vol. 31, No. 2, pp. 161–168.

PFEFFERMANN, D., KRIEGER, A. M., RINOTT, Y., (1998). Parametric distributions of complex survey data under informative probability sampling'. Statistica Sinica, 8, pp 1087–1114.

PFEFFERMANN, D., SVERCHKOV, M., (1999). Parametric and semi-parametric estimation of regression models fitted to survey data, Sankhya, 61, B, pp 166–186.

PFEFFERMANN, D., SIKOV, A., (2011). Imputation and Estimation under Nonignorable Nonresponse in Household Surveys with Missing Covariate Information, Journal of Official Statistics, Vol. 27, No. 2, 2011, pp. 181–209.

RIDDLES, M. K., KIM, J. K., IM, J., (2016). A propensity-score-adjustment method for nonignorable nonresponse, J. Surv. Stat. Methodol., 4, pp. 215–245.

SÄRNDAL, C .E., (2011). The 2010 Morris Hansen lecture dealing with survey nonresponse in data collection, in estimation, Journal of Official Statistics, Vol. 27, No. 1, 2011, pp. 1–21.

SVERCHKOV, M., PFEFFERMANN, D., (2004). Prediction of finite population totals based on the sample distribution, Survey Methodology, 30, pp 79–92.

SVERCHKOV, M., (2008), A New Approach to Estimation of Response Probabilities When Missing Data Are Not Missing at Random, Proceedings of the Survey Research Methods Section, pp. 867–874.