Life Cycle Reliability and Safety Engineering
https://doi.org/10.1007/s41872-024-00268-4

ORIGINAL RESEARCH

Design based study for adjusting non-response while estimating population mean

Ajeet Kumar Singh[1] · Ashutosh Ashutosh[2] · Jayendra Kumar Singh[3]

Received: 6 June 2024 / Accepted: 2 August 2024

Abstract
In the present study, we proposed a class of estimators for estimating a finite population mean in the presence of non-response to the study variable. We set out to investigate their properties under the polynomial regression model (PRM) modelling approach. Some of the special cases of the class were discussed separately to show how some non-response versions of the existing estimators can be generated and studied from the general class. The comparison of the model-based mean square errors of the estimators under different settings of the considered model was illustrated with the help of some empirical data.

Keywords Auxiliary variable · Modelling approach · Non-response · PRM · MSE

1 Introduction

1.1 Significance of non-response

In sampling theory, it is generally assumed that the true value of each unit in the population can be determined without error. In practice, this assumption may be violated for various reasons and because of practical constraints that exist at the time of the survey. The problem of nonresponse in sample surveys is usually due to a lack of interest on the part of respondents in answering, persons not present at home, lack of knowledge about the survey or the questions asked ethical issues, and refusal of respondents to answer the given questionnaire. In postal surveys, questionnaires are mailed to units selected in the sample with a request that they be returned within a specified period of time. In general, however, many respondents do not answer the questions or do not return the completed questionnaire within the specified time. In such cases, one can use the information obtained from the available questionnaires, but this may result in a loss of estimator efficiency due to the smaller sample size. In general, the postal questionnaire method is very often used in sample surveys to reduce the cost of the survey. Due to the high non-response rate that occurs for the reasons mentioned above. The component of bias creeps into the estimation procedure, leading to results as inaccurate as would not be expected without the presence of non-response. For mail surveys, Hansen and Hurwitz (Hansen and Hurwitz 1946) proposed a method of subsampling from nonresponding units and provided an estimate based on values obtained from responding units and subsamples obtained by the personal interview method from some nonresponding units. The literature on the various sources of non-response and methods of eliminating the problem of non-response can be found in Kish (Kish 1965), Kumar et al. (Kumar et al. 2019), Singh et al. (Singh and Singh 2022). Cochran (Cochran 1977) referred to the failure to measure some of the units in the selected sample, while Zarkovich (Zarkovich 1966) and Ford (Ford 1976) referred to the same problem as missing data. Sudman (Sudman 1976) addressed the problem of bias due to non-cooperation. Sukhatme and Sukhatme (Sukhatme and Sukhatme 1970) described the effects of incomplete samples. All of these approaches refer to the phenomenon of non-response. Sarndal et al. (Sarndal et al. 2003), Groves (Groves 2004), and Dillman et al. (Dillman et al. 2002) and

✉ Ajeet Kumar Singh
ajeetvns.singh@gmail.com

Ashutosh Ashutosh
kumarashubhustat@gmail.com

Jayendra Kumar Singh
jayendra88.singh@gmail.com

1 Department of Statistics, Faculty of Science, University of Rajasthan, Jaipur 302004, India

2 Department of Statistics, Allahabad Degree College, University of Allahabad, Prayagraj, India

3 Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi 221005, India

Published online: 03 September 2024

Chaudhary et al. (Chaudhary and Kumar 2016) have also defined various aspects and reasons for non-response.

1.2 Model-based approach

To make inference about the finite population, on the basis of a probability sample, by measuring the sample elements, two types of problems may arise Sarndal et al. (1992):

(a) Inference about the finite population itself.
(b) Inference about a model or a super-population from which the given population thought to have generated.

Case (a) can be dealt with design-based inference as it is usual in most of the survey sampling problems, but case (b) creates a different situation where the population is considered to be a random sample from a super-population which can be designated by a model D. Among others the approach was advocated by Brewer (Brewer 1963), Royall (Royall et al. 1971), Cassel et al. (Cassel et al. 1976), Singh et al. (Singh et al. 2009), Basu (Basu 1958), Singh et al. (Singh et al. 2017) and Singh et al. (Sarndal et al. 2003) Recently, Ahmed and Shabbir (Shakeel Ahmed and Javid Shabbir 2019) have discussed the utility of the estimator for paradigm of the model-based and in the presence of non-ignorable non-response.

Royall and Herson (Royall and Herson 1973a) (Royall and Herson 1973b) suggested a particular type of super population model, termed as PRM, which is described as

$$Y_t = \delta_0\beta_0 + \delta_1\beta_1 x_t + \delta_2\beta_2 x_t^2 + \cdots + \delta_J\beta_J x_t^J + \varepsilon_t\left[v(x_t)\right]^{1/2}$$

$$= \sum_{j=0}^{J} \delta_j\beta_j x_t^j + \varepsilon_t\left[v(x_t)\right]^{1/2} for\ t = 1, 2, \ldots, N \tag{1}$$

With $E_D(Y_t) = h(x_t) = \sum_{j=0}^{J} \delta_j\beta_j x_t^j$;

$Var\left[Y_t\right] = \sigma^2 v(x_t); Cov(Y_s, Y_t) = 0, s \neq t$

where Y_t is the study variable associated with the tth unit of the universe of size N, x_t is the value of the tth unit on the auxiliary variable X of the universe and it is non-negative. $\varepsilon_t; t = 1, 2, \ldots, N$ are independent random error with mean zero and variance σ^2, δ_j(j = 0, 1,.., J) is zero or one according as the term x_t^j is absent or present respectively. In the model (1), $v(x_t)$ are known function of x-values and $\beta_j; j = 1, 2, \ldots, J$ are unknown model parameters. It is denoted this model as $D\left[\delta_0, \delta_1, \delta_2, \ldots, \delta_J : v(x)\right]$. Chambers (Chambers 1986) has described that in both sample survey theory and practice, mean of Y_t is proportional to x_t. The variance of Y_t is proportional to $v(x_t)$.

2 Initiation of the problem

2.1 Let a finite population of size N, denoted by Ω, consists of N_1 respondents and N_2 non-respondents units. We selected a sample of size n from a universe that consists n_1 respondents and n_2 non-respondents. For efficient estimate, we collect some information the non-respondents. Therefore, we select a sub-sample h_2 from the non-respondents n_2 units. Let the sample of size n and h_2 be denoted by s and s_{h_2} respectively. Further let $\Omega = s \cup \bar{s}$, s and $\bar{s}$ are two disjoint sets, such that s represent the observed part of the universe and $\bar{s}$ represents the missing part of the universe. Further, consider $s = s_1 \cup s_2$(s_1 and s_2 samples of disjoint sets) and $s_2 = s_{h_2} \cup \bar{s}_{h_2}$. The value of where $\bar{s}_{h_2}$ is sub-portion of s_2 sample. We have discuss the notations:

Z = Variable X or Y

and

$\overline{Z} = N^{-1} \sum_{\Omega} z_t$: population mean of Z^{th} variable.

$S_Z^2 = (N-1)^{-1} \sum_{\Omega} (z_t - \overline{Z})^2$: Z^{th} population mean square.

Also sample means:

$$\bar{z}_{s_i} = n_i^{-1} \sum_{s_i} z_t\ (i = 1, 2); \bar{z}_{s_{h_2}} = h_2^{-1} \sum_{s_{h2}} z_t;$$

$$\bar{z}_{\bar{s}_{h_2}} = (n_2 - h_2)^{-1} \sum_{\bar{s}_{h_2}} z_t.$$

Higher order moments:

We considered for X

$$\overline{X}^{(j)} = \frac{1}{N} \sum_{\Omega} x_t^j; \bar{x}^{(j)} = \frac{1}{n} \sum_{s} x_t^j; \bar{x}_{\bar{s}}^{(j)}$$

$$= M^{-1} \sum_{\bar{s}} x_t^j; \bar{x}_{s_i}^{(j)} = n_i^{-1} \sum_{s_i} x_t^j; (i = 1, 2)$$

$$\bar{x}_{s_{h_2}}^{(j)} = (h_2)^{-1} \sum_{s_{h_2}} x_t^j; \bar{x}_{\bar{s}_{h_2}}^{(j)} = \frac{1}{(n_2 - h_2)} \sum_{\bar{s}_{h_2}} x_t^j \quad \text{for} \quad j = 1, 2, 3 \ldots$$

Obviously, we have seen that

$$\overline{X}^{(1)} = \overline{X}; \bar{x}^{(1)} = \bar{x}; \quad \bar{x}_{s_i}^{(1)} = \bar{x}_{s_i} (i = 1, 2);$$

$$\bar{x}_{s_{h_2}}^{(1)} = \bar{x}_{s_{h_2}}; \quad \bar{x}_{\bar{s}}^{(1)} = \bar{x}_{\bar{s}}; \quad \bar{x}_{\bar{s}_{h_2}}^{(1)} = \bar{x}_{\bar{s}_{h_2}}.$$

Hansen and Hurwitz (Hansen and Hurwitz 1946) suggested the problem of non-response of population mean $\overline{Y}$ is given by.

$$\bar{y}_w^* = \frac{n_1\bar{y}_{s_1} + n_2\bar{y}_{s_{h_2}}}{n}. \tag{2}$$

They obtained the variance of estimator as

$$V(\bar{y}_w^*) = \left(\frac{1}{n} - \frac{1}{N}\right)S_Y^2 + \frac{(f_2 - 1)}{n}\frac{N_2}{N}S_{2Y}^2; \tag{3}$$

where $f_2 = \frac{n_2}{h_2}$ and $\mathrm{S}_{2\mathrm{Y}}^2 = (N_2 - 1)^{-1}\sum_{t=1}^{N_2}(y_t - \overline{Y}_{N_2})^2$; $\overline{\mathrm{Y}}_{\mathrm{N}_2}$ shows non-respondent population mean.

3 Proposed family of strategy

Family $T_1^*(\alpha)$: Here, we first write the family of one parameter estimators for population mean in the presence of non-response

$$T_1^*(\alpha) = \bar{y}_w^*\psi^*\left(\alpha, \overline{X}, \bar{x}\right) \tag{4}$$

$$\text{where } \psi^*\left(\alpha, \overline{X}, \bar{x}\right) = \exp\left[\alpha\left(\frac{\overline{X} - \bar{x}_d}{\overline{X} + \bar{x}_d}\right)\right]$$

$$\text{where } \bar{x}_d = d\bar{x} + (1 - d)\overline{X}\,;\, d = \frac{n}{(N + n)}. \tag{5}$$

Remark 1 Instances of estimator $T_1^*(\alpha)$ when $\alpha = 0$.

$$T_1^*(0) = \bar{y}_w^* \tag{6}$$

It is extended form of estimator developed by Hansen and Hurwitz (Groves 2004). They have also developed some other cases for $\alpha = 1$ and -1, we obtained an exponential-type ratio and product estimators.

4 *D*-Bias and *D*-MSE of $T_1^*(\alpha)$

The simple form of the model $D\left[\delta_0, \delta_1, \ldots, \delta_J : v(x)\right]$, D—based estimator bias and MSE. ***Theorem 1:*** Value of D-bias of estimator $T_1^*(\alpha)$ is

$$B_D\left[T_1^*(\alpha)\right] = \sum_{j=0}^{J}\delta_j\beta_j\left[\psi^*\left(\alpha, \overline{X}, \bar{x}\right)\left(\frac{n_1\bar{x}_{s_1}^{(j)} + n_2\bar{x}_{s_{h_2}}^{(j)}}{n}\right) - \overline{X}^{(j)}\right] \tag{7}$$

The appendix A, Section –I mentioned proof of this Eq. (7)

Theorem 2 *The value of D-MSE of the estimator $T_1^*(\alpha)$ is given by.*

$$M_D\left(T_1^*(\alpha)\right) = \left[B_D\left(T_1^*(\alpha)\right)\right]^2 + \left[\frac{\psi^*\left(\alpha, \overline{X}, \bar{x}\right)}{n} - \frac{1}{N}\right]^2\sigma^2\sum_{s_1}v(x_t)$$

$$+\left[\frac{f_2}{n}\psi^*\left(\alpha, \overline{X}, \bar{x}\right) - \frac{1}{N}\right]^2\sigma^2\sum_{s_{h_2}}v(x_k) + \frac{\sigma^2}{N^2}\left[\sum_{\bar{s}_{h_2}}v(x_t) + \sum_{\bar{s}}v(x_t)\right]. \tag{8}$$

The appendix A, Section –I mentioned proof of this Eq. (8).

Remark 2 We have seen from Eq. (8). The D-MSE of $T_1^*(\alpha)$ is consisting two parts one is D -Bias and second is variance of estimator. We know that variance of the estimator does not depend on the function $\sum_{j=0}^{J}\delta_j\beta_j x_t^j$. However, it is depending on error term ε_t and variance $v(x_t)$. In the case spoil selection of polynomial regression there is no change in the variance of estimator.

5 *D*-Bias and *D*-MSE of particular cases of $T_1^*(\alpha)$.

5.1. The PRM instance cases assumed two different forms with a functions $h(x_t) = \sum_{j=0}^{J}\delta_j\beta_j x_t^j$ and function $v(x_t)$. The instances forms of the model is given by

$$Y_t = \beta_1 x_t + \varepsilon_t\left[x_t\right]^{1/2} \tag{9}$$

$$Y_t = \beta_0 + \beta_1 x_t + \varepsilon_t\left[x_t^2\right]^{1/2} \tag{10}$$

$$Y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t\left[x_t^g\right]^{2^{-1}}. \tag{11}$$

Obviously models (9), (10) & (11) may be denoted respectively as $D[0,\ 1\ : x]$, $D\left[1,\ 1 : x^2\right]$ and $D[1,\ 1,\ 1\ : x^g]$. The variance function of model based idea was developed Cochran (Chambers 1986) and Brewer (Brewer 1963) when $v(x_t)$ assumed to be x_t^g with $0 \le g \le 2$. We motive by their contribution, we have provided six different PRMs:

$$\text{Model I } D[0,\ 1 : 1] \Rightarrow h(x_t) = \beta_1 x_t, v(x_t) = x_t^0 \tag{12}$$

$$\text{Model II } D[0, 1 : x] \Rightarrow h(x_t) = \beta_1 x_t, v(x_t) = x_t \tag{13}$$

$$\text{Model III } D\left[0, 1 : x^2\right] \Rightarrow h(x_t) = \beta_1 x_t, v(x_t) = x_t^2 \tag{14}$$

$$\text{Model IV } D[1, 1 : 1] \Rightarrow h(x_t) = \beta_0 + \beta_1 x_t, v(x_t) = x_t^0 \tag{15}$$

$$\text{Model V } D[1, 1 : x] \Rightarrow h(x_t) = \beta_0 + \beta_1 x_t, v(x_t) = x_t \tag{16}$$

$$\text{Model VI } D\left[1, 1 : x^2\right] \Rightarrow h(x_t) = \beta_0 + \beta_1 x_t, v(x_t) = x_t^2 \tag{17}$$

6 D-MSE of $T_1^*(\alpha)$ under Models I-VI

The MSE of the family of estimators $T_1^*(\delta)$ in the models I–VI can be obtained from expression (8). These are presented as follows:

$$M_D\left[T_1^*(\alpha)\right]_I=\left[\beta_1\left\{\psi^*\left(\alpha,\overline{X},\overline{x}\right)\left(\frac{n_1\overline{x}_{s_1}+n_2\overline{x}_{s_{h_2}}}{n}\right)-\overline{X}\right\}\right]^2$$

$$+\left[\frac{\psi^*\left(\alpha,\overline{X},\overline{x}\right)}{n}-\frac{1}{N}\right]^2 n_1\sigma^2+\left[\frac{f_2}{n}\psi^*\left(\alpha,\overline{X},\overline{x}\right)-\frac{1}{N}\right]^2 h_2\sigma^2$$

$$+\frac{\sigma^2}{N^2}\left[M+\left(n_2-h_2\right)\right]. \tag{18}$$

$$M_D\left[T_1^*(\alpha)\right]_{II}=\left[\beta_1\left\{\psi^*\left(\alpha,\overline{X},\overline{x}\right)\left(\frac{n_1\overline{x}_{s_1}+n_2\overline{x}_{s_{h_2}}}{n}\right)-\overline{X}\right\}\right]^2$$

$$+\left[\frac{\psi^*\left(\alpha,\overline{X},\overline{x}\right)}{n}-\frac{1}{N}\right]^2 n_1\sigma^2\overline{x}_{s_1}+\left[\frac{f_2}{n}\psi^*\left(\alpha,\overline{X},\overline{x}\right)-\frac{1}{N}\right]^2 h_2\sigma^2\overline{x}_{s_{h_2}}$$

$$+\frac{\sigma^2}{N^2}\left\{\sum_{\overline{s}_{h_2}}x_1+\sum_{\overline{s}}x_1\right\}. \tag{19}$$

$$M_D\left[T_1^*(\alpha)\right]_{III}=\left[\beta_1\left\{\psi^*\left(\alpha,\overline{X},\overline{x}\right)\left(\frac{n_1\overline{x}_{s_1}+n_2\overline{x}_{s_{h_2}}}{n}\right)-\overline{X}\right\}\right]^2$$

$$+\left[\frac{\psi^*\left(\alpha,\overline{X},\overline{x}\right)}{n}-\frac{1}{N}\right]^2 \sigma^2\sum_{s_1}x_t^2+\left[\frac{f_2}{n}\psi^*\left(\alpha,\overline{X},\overline{x}\right)-\frac{1}{N}\right]^2 \sigma^2\sum_{s_{h_2}}x_t^2$$

$$+\frac{\sigma^2}{N^2}\left[\sum_{\overline{s}_{h_2}}x_t^2+\sum_{\overline{s}}x_t^2\right]. \tag{20}$$

$$M_D\left[T_1^*(\alpha)\right]_{IV}=\left[\beta_0\left\{\psi^*\left(\alpha,\overline{X},\overline{x}\right)-1\right\}+\beta_1\left\{\psi^*\left(\alpha,\overline{X},\overline{x}\right)\left(\frac{n_1\overline{x}_{s_1}+n_2\overline{x}_{s_{h_2}}}{n}\right)-\overline{X}\right\}\right]^2$$

$$+\left[\frac{\psi^*\left(\alpha,\overline{X},\overline{x}\right)}{n}-\frac{1}{N}\right]^2 n_1\sigma^2+\left[\frac{f_2}{n}\psi^*\left(\alpha,\overline{X},\overline{x}\right)-\frac{1}{N}\right]^2 h_2\sigma^2$$

$$+\frac{\sigma^2}{N^2}\left[M+\left(n_2-h_2\right)\right]. \tag{21}$$

$$M_D\left[T_1^*(\alpha)\right]_V=\left[\beta_0\left\{\psi^*\left(\alpha,\overline{X},\overline{x}\right)-1\right\}+\beta_1\left\{\psi^*\left(\alpha,\overline{X},\overline{x}\right)\left(\frac{n_1\overline{x}_{s_1}+n_2\overline{x}_{s_{h_2}}}{n}\right)-\overline{X}\right\}\right]^2$$

$$+\left[\frac{\psi^*\left(\alpha,\overline{X},\overline{x}\right)}{n}-\frac{1}{N}\right]^2 n_1\sigma^2\overline{x}_{s_1}+\left[\frac{f_2}{n}\psi^*\left(\alpha,\overline{X},\overline{x}\right)-\frac{1}{N}\right]^2 h_2\sigma^2\overline{x}_{s_{h_2}}$$

$$+\frac{\sigma^2}{N^2}\left\{\sum_{\overline{s}_{h_2}}x_t+\sum_{\overline{s}}x_t\right\}. \tag{22}$$

$$M_D\left[T_1^*(\alpha)\right]_{VI}=\left[\beta_0\left\{\psi^*\left(\alpha,\overline{X},\overline{x}\right)-1\right\}+\beta_1\left\{\psi^*\left(\alpha,\overline{X},\overline{x}\right)\left(\frac{n_1\overline{x}_{s_1}+n_2\overline{x}_{s_{h_2}}}{n}\right)-\overline{X}\right\}\right]^2$$

$$+\left[\frac{\psi^*\left(\alpha,\overline{X},\overline{x}\right)}{n}-\frac{1}{N}\right]^2 \sigma^2\sum_{s_1}x_t^2+\left[\frac{f_2}{n}\psi^*\left(\alpha,\overline{X},\overline{x}\right)-\frac{1}{N}\right]^2 \sigma^2\sum_{s_{h_2}}x_t^2$$

$$+\frac{\sigma^2}{N^2}\left[\sum_{\overline{s}_{h_2}}x_t^2+\sum_{\overline{s}}x_t^2\right]. \tag{23}$$

7 Some existing strategies with their MSEs

Two families of estimators in the presence of non-response was discussed Singh et al. (Singh et al. 2017) and compared them under different PRMs. The estimators and their MSEs under $D\left[\delta_0,\delta_1,\ldots,\delta_J\,:\,v(x)\right]$ are as follows

$$\text{(i) } T_s^*(\alpha)=\overline{y}_w^*\psi\left(\alpha,\overline{X},\overline{x}\right) \tag{24}$$

where $\psi\left(\alpha,\overline{X},\overline{x}\right)=\exp\left[\alpha\left(\frac{\overline{X}-\overline{x}}{\overline{X}+\overline{x}}\right)\right]$, and

$$M_D\left(T_s^*(\alpha)\right) = \left[B_D\left(T_s^*(\alpha)\right)\right]^2 + \left[\frac{\psi\left(\alpha,\overline{X},\overline{x}\right)}{n} - \frac{1}{N}\right]^2 \sigma^2 \sum_{s_1} v\left(x_t\right)$$

$$+\left[\frac{f_2}{n}\psi\left(\alpha,\overline{X},\overline{x}\right) - \frac{1}{N}\right]^2 \sigma^2 \sum_{s_{h_2}} v\left(x_t\right) + \frac{\sigma^2}{N^2}\left[\sum_{\bar{s}_{h_2}} v\left(x_t\right) + \sum_{\bar{s}} v\left(x_t\right)\right] \tag{25}$$

$$\text{(ii)}\, t_{NR}^*(\alpha) = \overline{y}_w^{**}\frac{\overline{X}}{\overline{\overline{x}}}, \tag{26}$$

and

$$\overline{y}_w^{**} = \frac{n_1\overline{y}_{s_1} + n_2\, t_{s_2}^*(\alpha)}{n} \tag{27}$$

$$t_{s_2}(\alpha) = \overline{y}_{s_{h_2}} \exp\left[\alpha\left(\frac{\overline{x}_{s_2} - \overline{x}_{s_{h_2}}}{\overline{x}_{s_2} + \overline{x}_{s_{h_2}}}\right)\right] \tag{28}$$

$$M_D\left[t_{NR}^*(\alpha)\right] = \left[\sum_{j=0}^{J} \delta_j\beta_j\left(\frac{\bar{X}}{n\bar{x}}\left(n_1\bar{x}_{s_1}^{(j)} + n_2\psi\left(\alpha,\bar{\mathrm{x}}_{\mathrm{s}_2},\bar{x}_{s_{h2}}\right)\bar{x}_{s_{h_2}}^{(j)}\right) - \bar{X}^{(j)}\right)\right]^2$$
$$+\left(\frac{\bar{X}}{n\bar{x}} - \frac{1}{N}\right)^2\sigma^2\sum_{s_1} v(x_t) + \left[\frac{n_2}{nh_2}\psi\left(\alpha,\bar{\mathrm{x}}_{\mathrm{s}_2},\bar{x}_{s_{h2}}\right)\frac{\bar{X}}{\bar{x}} - \frac{1}{N}\right]^2$$
$$\sigma^2\sum_{s_{h_2}} v(x_t) + \frac{1}{N^2}\sigma^2\sum_{\bar{s}_{h_2}} v(x_t) + \frac{1}{N^2}\sigma^2\sum_{\bar{s}} v(x_t). \tag{29}$$

8 Robustness of the estimator $T_1^*(\alpha)$ and its comparison with $T_s^*(\alpha)$ and $t_{NR}^*(\alpha)$

8.1 In remark 2, we have seen, the D-MSE of the strategies is affected by the deviations in $h(x_t)$ and function $v(x_t)$, if it a D-biased, while the bias is affected only by the function $h(x_t)$ and is completely do not dependent in $v(x_t)$. Therefore, it would be desirable to consider change in the amount of MSE with the deviation of the model, either due to misspecification in $h(x_t)$ or in $v(x_t)$ or both.

Royall and Herson [17&18], therefore, considered an estimator 'robust' if there is a nominal change in the amount of D—MSE due to the deviation of the model, that is, if the optimality of an estimator vitiates slightly under the deviation of model, it could be termed as robust one. Thus, it can be stated that the general aim of the model approach is to find out a strategy which performs well in some broad sense allowing for our uncertainty about the assumed model, that is, a strategy which is almost insensitive to errors in the model. Since a large number of theoretical models may be thought of, it is quite impossible to examine the robustness of an estimator theoretically under the deviations of the models. It is, therefore, advisable to examine the robustness of the estimator under some working models with known parameters.

8.2 It is also appropriate to have a study of comparison of MSEs of estimators $T_1^*(\alpha)$**,**$T_s^*(\alpha)$ and $t_{NR}^*(\alpha)$**,** all of which are developed under same set-up. Based upon an empirical data, such a comparison has also been presented in the next section.

9 Empirical data and results

9.1 We have taken a real data Singh et al. (Singh et al. 2017). We have considered two different (15% and 30%) non—response rate. The number of dwelling is x_t and dwelling occupied with y_t. The following values are as:

$N = 90,\ \beta_0 = 0.8787,\ \beta_1 =—4.9157,\ \overline{X} = 41.4556,\ \sigma^2 = 0.7998.$

(i) The values of non-response rate at 15% are:

$n = 20,\ n_1 = 17,\ n_2 = 3,\ h_2 = 2,\ f_2 = 1.5,\ \overline{x} = 39.55,\ \overline{x}_{s_1} = 39.824,\ \overline{x}_{s_2} = 38.0, \overline{x}_{s_{h_2}} = 30.5,\ \sum_{s_1} x_t = 677,\ \sum_{s_{h2}} x_t = 61,\ \sum_{\bar{s}_{h_2}} x_t = 53,\ \sum_{\bar{s}} x_t = 2940\ \sum_{s_1} x_t^2 = 36{,}729,\ \sum_{s_{h_2}} x_t^2 = 2081,\ \sum_{\bar{s}} x_t^2 = 179{,}614, \sum_{\bar{s}_{h_2}} x_t^2 = 2809.$

(ii) The values of non-response rate at 30% are:

$n = 20, n_1 = 14, n_2 = 6, h_2 = 4, f_2 = 1.5, \overline{x} = 39.55, \overline{x}_{s_1} = 36.5,\ \overline{x}_{s_2} = 46.6667,\ \overline{x}_{s_{h_2}} = 50,\ \sum_{s_1} x_t = 511,\ \sum_{s_{h2}} x_t = 200,\ \sum_{\bar{s}_{h_2}} x_t = 80,\ \sum_{\bar{s}} x_t = 2940,\ \sum_{s1} x_t^2 = 25{,}337,\ \sum_{\bar{s}} x_t^2 = 179{,}614, \sum_{\bar{s}h2} x_t^2 = 3328.$

9.2 Tables 1, 2, 3, 4, 5, 6 details the MSEs of stretegies $T_1^*(\alpha)$**,** $T_s^*(\alpha)$ & $t_{NR}^*(\alpha)$ for $\alpha = 0, 1$ and -1 with 15% and 30% non response rates over Models I-VI.

10 Conclusions

We obtained the following point from the tables:

(i) The proposed estimator $T_1^*(\alpha)$ is nearly robust for models I, II, IV and V regardless of the choice of α of a given nonresponse rate. There is a significant change in the MSE value of the estimator for models III and VI compared to models I, II, IV and V, but for a fixed nonresponse rate. The estimator can again be considered robust under the variance function x^2, regardless of the choice of α. Thus, the estimator is not affected under misspecification of $h\left(x_t\right)$ when the variance function is x_t^g for $g = 0$ and 1. Similarly, the conclusion for the variance function x^g for $g = 2$.

Table 1 The D-MSE of the estimators (α =0) for the Models I to VI in presence of 15% of non-response rate

g	Models	$T_s^*(\alpha)$	$t_{NR}^*(\alpha)$	$T_1^*(\alpha)$
0	I	221.970	33.640	221.970
1	II	223.250	35.040	223.250
2	III	291.170	109.300	291.170
0	IV	221.970	34.130	221.970
1	V	223.250	35.530	223.250
2	VI	291.170	109.790	291.170

Table 2 The D- MSE of the estimators (α=0) for the Models I to VI in presence of 30% of non-response rate

g	Models	$T_s^*(\alpha)$	$t_{NR}^*(\alpha)$	$T_1^*(\alpha)$
0	I	19.850	26.590	19.850
1	II	21.380	28.270	21.380
2	III	110.820	126.380	110.820
0	IV	19.850	26.160	19.850
1	V	21.390	27.840	21.390
2	VI	110.820	125.950	110.820

Table 3 The D-MSE of the estimators (α=1) for the Models I to VI in presence of 15% of non-response rate

g	Models	$T_s^*(\alpha)$	$t_{NR}^*(\alpha)$	$T_1^*(\alpha)$
0	I	198.940	9.460	198.940
1	II	200.280	10.920	200.230
2	III	290.650	87.100	268.680
0	IV	199.040	9.820	199.040
1	V	200.390	11.2860	200.330
2	VI	290.760	87.660	268.790

Table 4 The D-MSE of the estimators (α=1) for the Models I to VI in presence of 30% of non-response rate

g	Models	$T_s^*(\alpha)$	$t_{NR}^*(\alpha)$	$T_1^*(\alpha)$
0	I	13.090	6.460	13.090
1	II	14.640	8.080	13.090
2	III	104.810	102.600	104.800
0	IV	13.120	6.290	13.120
1	V	14.670	7.920	14.670
2	VI	104.830	102.430	104.840

Table 5 The D-MSE of the estimators (α=—1) for the Models I to VI in presence of 15% of non-response rate

g	*Models*	$T_s^*(\alpha)$	$t_{NR}^*(\alpha)$	$T_1^*(\alpha)$
0	I	246.160	67.960	246.160
1	II	247.490	69.310	247.430
2	III	336.380	141.700	314.820
0	IV	246.050	68.720	246.050
1	V	247.380	69.770	247.320
2	VI	336.270	142.360	314.700

Table 6 The D-MSE of the estimators (α=—1) for the Models I to VI in presence of 30% of non-response rate

g	Models	$T_s^*(\alpha)$	$t_{NR}^*(\alpha)$	$T_1^*(\alpha)$
0	I	82.590	61.890	27.980
1	II	84.050	63.630	29.500
2	III	169.480	165.620	118.200
0	IV	82.220	61.070	27.940
1	V	83.680	62.810	29.460
2	VI	169.110	164.800	118.160

(ii) $T_1^*(\alpha)$ is an efficient or sometimes better than the estimators $T_s^*(\alpha)$ and $t_{NR}^*(\alpha)$ in terms of estimators accuracy.

(iii) It is interesting to note that for α =0, $T_s^*(\alpha)$ coincides with the estimator $T_1^*(\alpha)$ under all the models and non-response rates. This is because of the reason that for α =0, both estimators reduces to $T_1^*(0) = \overline{y}_w^* = T_s^*(0)$.

However, the conclusions drawn above are based on solely on empirical data and a particular configuration of the sample for different non-response rates. Therefore, a comparison of the estimators at different non-response rates would not be possible due to changing in the sample configuration. Furthermore, the results presented here are limited to the data at hand so no consistent conclusions can be drawn. Clearly, results may change with other data. The presentation made here is only an attempt to get an idea of the nature of the proposed family on the misspecifications of PRMs.

Appendix

Section I*:*

We have

$$B_D\left[T_1^*(\alpha)\right] = E_D\left[T_1^*(\alpha) - \overline{Y}\right]$$

$$= E_D\left[\bar{y}_w^*\psi^*\left(\alpha,\bar{X},\bar{x}\right) - \bar{Y}\right]$$

$$= E_D\left[\psi^*\left(\alpha,\bar{X},\bar{x}\right)\left(\frac{n_1\bar{y}_{s_1} + n_2\bar{y}_{s_{h_2}}}{n}\right) - \frac{1}{N}\sum_{k=1}^{N} y_t\right]$$

$$= E_D\left[\psi^*\left(\alpha,\bar{X},\bar{x}\right)\left(\frac{n_1}{n}\frac{1}{n_1}\sum_{s_1} y_k + \frac{n_2}{n}\frac{1}{h_2}\sum_{s_{h_2}} y_t\right) - \frac{1}{N}\sum_{k=1}^{N} y_t\right]. \tag{A1}$$

Now, using the PRM

$$Y_t = \delta_0\beta_0 + \delta_1\beta_1x_t + \delta_2\beta_2x_t^2 + \ldots + \delta_J\beta_Jx_t^J + \varepsilon_t\left[v\left(x_t\right)\right]^{1/2}$$
$$= \sum_{j=0}^{J}\delta_j\beta_jx_t^j + \varepsilon_t\left[v\left(x_t\right)\right]^{1/2} \text{ for } t = 1, 2, \ldots, N$$

With $E_D\left(Y_t\right) = \sum_{j=0}^{J}\delta_j\beta_jx_t^j$,
$Var\left[Y_t\right] = \sigma^2 v\left(x_t\right), Cov\left(Y_s, Y_t\right) = 0$ for $s \neq t$,
$E_D\left(\varepsilon_t\right) = 0$ for all k, $E_D\left(\varepsilon_t^2\right) = \sigma^2$ for all t.
we can write

$$B_D[T_1^*(\alpha)] = E_D\left[\psi^*\left(\alpha,\bar{X},\bar{x}\right)\frac{1}{n}\sum_{s_1}\left(\sum_{j=0}^{J}\delta_j\beta_jx_t^j + \varepsilon_k\left(v\left(x_t\right)\right)^{1/2}\right)\right.$$

$$+\psi^*\left(\alpha,\bar{X},\bar{x}\right)\frac{f_2}{n}\sum_{s_{h_2}}\left(\sum_{j=0}^{J}\delta_j\beta_jx_t^j + \varepsilon_t\left(v\left(x_t\right)\right)^{1/2}\right)$$

$$\left.-\frac{1}{N}\sum_{\Omega}\left(\sum_{j=0}^{J}\delta_j\beta_jx_t^j + \varepsilon_t\left(v\left(x_t\right)\right)^{1/2}\right)\right] \tag{A2}$$

Since $E_D(\varepsilon_t) = 0$ for all t, we have

$$B_D[T_1^*(\alpha)] = \psi^*\left(\alpha,\bar{X},\bar{x}\right)\left(\sum_{j=0}^{J}\delta_j\beta_j\frac{1}{n}\sum_{s_1}x_t^j + \sum_{j=0}^{J}\delta_j\beta_j\frac{f_2}{n}\sum_{s_{h_2}}x_t^j\right) - \sum_{j=0}^{J}\delta_j\beta_j\frac{1}{N}\sum_{\Omega}x_t^j$$

$$= \sum_{j=0}^{J}\delta_j\beta_j\left[\psi^*\left(\alpha,\bar{X},\bar{x}\right)\left(\frac{n_1\bar{x}_{s_1}^{(j)} + n_2\bar{x}_{s_{h_2}}^{(j)}}{n}\right) - \bar{X}^{(j)}\right] \tag{A3}$$

Thus Eq. (7) follows.

Section II

The D-MSE of the strategy $T_1^*(\alpha)$ for model 1 is derived as follows:

We have

$$M_D\left(T_1^*(\alpha)\right) = E_D\left[T_1^*(\alpha) - \bar{Y}\right]^2$$

$$= E_D\left[\bar{y}_w^*\psi^*\left(\alpha,\bar{X},\bar{x}\right) - \bar{Y}\right]^2$$

$$= E_D\left[\psi^*\left(\alpha,\bar{X},\bar{x}\right)\left(\frac{1}{n}\sum_{s_1}y_t + \frac{f_2}{n}\sum_{s_{h_2}}y_t\right) - \frac{1}{N}\sum_{k=1}^{N}Y_t\right]^2$$

$$= E_D\left[\psi^*\left(\alpha,\bar{X},\bar{x}\right)\frac{1}{n}\sum_{s_1}\left(\sum_{j=0}^{J}\delta_j\beta_jx_t^j + \varepsilon_t\left(v\left(x_t\right)\right)^{1/2}\right)\right.$$

$$+\psi^*\left(\alpha,\bar{X},\bar{x}\right)\frac{f_2}{n}\sum_{s_{h_2}}\left(\sum_{j=0}^{J}\delta_j\beta_jx_t^j + \varepsilon_t\left(v\left(x_t\right)\right)^{1/2}\right)$$

$$\left.-\frac{1}{N}\sum_{\Omega}\left(\sum_{j=0}^{J}\delta_j\beta_jx_t^j + \varepsilon_t\left(v\left(x_t\right)\right)^{2^{-1}}\right)\right]^2. \tag{A4}$$

such that $E_D\left(\varepsilon_t,\varepsilon_s\right) = 0$ for $s \neq t$ and $E_D\left(\varepsilon_t^2\right) = \sigma^2$, we have

$$M_D\left(T_1^*(\alpha)\right) = \left[\sum_{j=0}^{J}\delta_j\beta_j\left(\psi^*\left(\alpha,\bar{X},\bar{x}\right)\left(\frac{n_1\bar{x}_{s_1}^{(j)} + n_2\bar{x}_{s_{h_2}}^{(j)}}{n}\right) - \bar{X}^{(j)}\right)\right]^2 + \left[\frac{\psi^*\left(\alpha,\bar{X},\bar{x}\right)}{n} - \frac{1}{N}\right]^2\sigma^2\sum_{s_1}v\left(x_t\right)$$

$$+\left[\frac{f_2}{n}\psi^*\left(\alpha,\bar{X},\bar{x}\right) - \frac{1}{N}\right]^2\sigma^2\sum_{s_{h_2}}v\left(x_t\right) + \frac{\sigma^2}{N^2}\left[\sum_{\bar{s}_{h_2}}v\left(x_t\right) + \sum_{\bar{s}}v\left(x_t\right)\right] \tag{A5}$$

Expression (A5) can further be written as

$$M_\xi\left(T_1^*(\alpha)\right) = \left[B_\xi\left(T_1^*(\alpha)\right)\right]^2 + \left[\frac{\psi^*\left(\alpha,\bar{X},\bar{x}\right)}{n} - \frac{1}{N}\right]^2\sigma^2\sum_{s_1}v\left(x_t\right)$$

$$+\left[\frac{f_2}{n}\psi^*\left(\alpha,\bar{X},\bar{x}\right) - \frac{1}{N}\right]^2\sigma^2\sum_{s_{h_2}}v\left(x_t\right) + \frac{\sigma^2}{N^2}\left[\sum_{\bar{s}_{h_2}}v\left(x_t\right) + \sum_{\bar{s}}v\left(x_t\right)\right] \tag{A6}$$

Hence the expression (8) follows.

Funding There is no funding.

Data availability We have taken data from paper Singh et al. (2017) which is published in Kish, L. (1965) Survey Sampling.

Declarations

Conflict of interest No potential conflict of interest was reported by the authors.

References

Bahl S, Tuteja RK (1991) Ratio and product type-exponential estimator. Inform Opt Sci XII I:159–163

Basu D (1958) On sampling with and without replacement, Sankhya, 20 A, 287–294.

Brewer KRW (1963) Ratio estimation and finite populations: Some results deducible from the assumption of an underlying stochastic process. Aust J Stat 5:93–105

Cassel CM, Sarndal CE, Wretman JH (1976) Some results on generalized difference estimation and generalized regression estimation for finite populations. Biometrika 63:615–620

Chambers RL (1986) Outlier robust finite population estimation. J Am Stat Assoc 81(396):1063–1069

Chaudhary MK, Kumar A (2016) Combined-type estimators of finite population mean using double sampling scheme under nonresponse. J Adv Res Appl Math Stat 1(3&4):6–18

Cochran WG (1953) *Sampling Techniques,* John Wiley and Sons, Inc., New York, I, Edition.

Cochran WG (1977) Sampling Techniques, Wiley Eastern Limited, New Delhi III Edition.

Dillman D, Eltinge J, Groves RM, Little R (2002) Survey non-response in design, data collection and analysis, In Survey *Non-response*, 3–26. Wiley, New York

Ford BL (1976) Missing data procedures: A comparative study, American Statistical Association Proceedings, Social Statistics Section, 324–329.

Groves R (2004) Survey errors and survey costs. Wiley, New York

Hansen MH, Hurwitz WN (1946) The problem of non-response in sample surveys. J Am Stat Assoc 41:517–529

Kish L (1965) Survey sampling. Wiley and Sons, New York, I Edition

Kish L (1967) Survey Sampling. John Wiley and Sons Inc, New York, II Edition

Kumar A, Singh AK, Singh VK (2019) Investigating the performance of a family of exponential-type estimators in presence of measurement error in Communications in Statistics—Theory and Methods, Vol-49, Issue No-23, PP. 1–20.

Royall RM (1971) Linear regression models in finite population sampling theory. In: Godambe VP, Sprott DA (eds) Foundations of Statistical Inference. Holt, Rinehart and Winston, Toronto, pp 259–274

Royall RM, Herson J (1973a) Robust estimation in finite populations I. J Am Stat Assoc 68(344):880–889

Royall RM, Herson J (1973b) Robust estimation in finite populations II: Stratification on a size variable. J Am Stat Assoc 68(344):890–893

Sarndal CE, Swensson B, Wretman J (1992) Model Assisted Survey Sampling, I. Springer-Verlag, New York Inc

Sarndal CE, Swensson B, Wretman J (2003) Model Assisted Survey Sampling. Edition, Springer-Verlag, New York Inc, II

Shakeel Ahmed and Javid Shabbir (2019) Model based estimation of population total in presence of non-ignorable non-response. PLoS ONE 14(10):e0222701

Singh AK, Singh VK (2022) (2022): A family of estimators for population mean under model approach in presence of non-response. J Reliab Stat Stud 15(1):1–20

Singh VK, Singh RVK, Shukla RK (2009) Model-based study of some estimators in the presence of non-response. In: Singh KK, Yadava RC, Pandey A (eds) Population, Poverty and Health : Analytical Approaches. Hindustan Publishing Corporation, New Delhi, India, pp 360–365

Singh AK, Singh P, Singh VK (2017) Model based study of families of exponential type estimators in presence of nonresponse. Commun Stat-Theory Methods 46(13):6478–6490

Sudman S (1976) Applied sampling. Academic Press, New York

Sukhatme PV, Sukhatme BV (1970) Sampling theory of surveys with applications. Asia Publishing House, London

Zarkovich SS (1966) Quality of statistical data. Food and Agricultural Organization of the United Nations, Rome