A Family of Estimators for Population Mean Under Model Approach in Presence of Non-Response

Ajeet Kumar Singh[1,*] and V. K. Singh[2]

[1]*Department of Statistics, University of Rajasthan, Jaipur, India*
[2]*Department of Statistics, Banaras Hindu University, Varanasi, India*
E-mail: ajeetvns.singh@gmail.com; vijay_usha_2000@yahoo.com
[*]Corresponding Author

Abstract

We have defined a class of estimators for population mean under non-response error based upon the concept of sub-sampling of non-respondents utilizing an auxiliary variable. The class is a one-parameter class of estimators which is based on the idea of exponential type estimators (ETE). The model biasness and model-mean square error of the class and some of its important members have been derived under polynomial regression model (PRM). The effect of variations in PRM specifications on the efficiency of the estimators has been discussed based upon the empirical results.

Keywords: Non-response, families of estimators, polynomial regression model, mean square error.

1 Introduction

The errors arising due to drawing inferences about the population on the basis of only a part of it; the sample. This type of errors are called 'sampling errors'. A second source of error may arise because of failure to measure some of the units in the selected sample or in the ascertainment of information some of the units or wrong reporting or recording or tabulation or processing of data. When all of these types are grouped together they are termed as 'non-sampling errors'. Further non-sampling error has been studied with illustrations by Deming (1944), Mahalanobis (1946), Moser (1958) and Zarkovich (1966). There are several literatures by Basu (1958), Cassel et al. (1976), Brewer (1963), Singh et al. (2009b) and Shukla, R. K. (2010), Royall (1971) and Royall and Herson (1973a, 1973b) stating that the model-based inference is not only desirable but almost necessary.

Sometimes, these non-sampling errors are more serious than the sampling error. Therefore an effort has been made to reduce the non-sampling error along with the sampling error incurred while conducting sample surveys. The main factors of non-sampling error can be encountered as measurement error and error due to non-response. Hansen and Hurwitz (1946) were the first to deal with the problem of non-response in mail surveys with the concept of sub-sampling the non-respondents. They developed an unbiased estimator for population mean on the basis of respondents and non-respondents.

1.1 Super Population Approach

Contrary to the classical survey sampling theory which assumes that the observations on the units of the population are fixed, there is another school of thought that advocates the use of the concepts that the finite population values are realized outcomes of a set of random variables which has been selected from a probability distribution ξ. This approach is known as *Super population Model Approach* (*SPMA*), in contrast to classical theory known as *Fixed Population Approach* (*FPA*). The SPMA has its own advantages and disadvantages. In the literature, a good number of studies are available based upon SPMA. In most of these studies, the main aim has been to investigate the "robustness of the estimators under the misspecifications of the models, that is, how much the estimator is consistent in terms of stability of its variance or mean square error over the changes of the model.

A special case of super-population model was proposed by Royall and Herson (1973a, 1973b) named as *Polynomial Regression Model (PRM)*,

which is described as

$$Y_k = \delta_0\beta_0 + \delta_1\beta_1 x_k + \delta_2\beta_2 x_k^2 + \cdots + \delta_J\beta_J x_k^J + \in_k [v(x_k)]^{1/2}$$

$$= \sum_{j=0}^{J} \delta_j\beta_j x_k^j + \in_k [v(x_k)]^{1/2} \quad \text{for } k = 1, 2, \ldots, N \tag{1}$$

with

$$E_\xi(Y_k) = h(x_k) = \sum_{j=0}^{J} \delta_j\beta_j x_k^j;$$

$$Var[Y_k] = \sigma^2 v(x_k); \quad Cov(Y_r, Y_k) = 0, \ r \neq k,$$

where Y_k is the random variable associated with the kth unit of the finite population of size N, x_k is the value of the kth unit of the population on the known auxiliary variable X typically referred to as their measures of size ($x_k > 0$ for k $= 1, 2, 3 \ldots N$), $\in_1, \in_2, \ldots \in_N$ are independent random variables each having mean zero and variance $\sigma^2, \delta_j (j = 0, 1, \ldots, J)$ is zero or one according as the term x_k^j is absent or present respectively in the model (1), $v(x_k)$ is a known function of x-values and $\beta_0, \beta_1, \ldots, \beta_J$ are unknown model parameters. Royall and Herson (1973a) denoted this model as $\xi[\delta_0, \delta_1, \delta_2, \ldots, \delta_J: v(x)]$. Chambers (1986) has mentioned that in both sample survey theory and practice, expectation of Y_k is proportional to x_k and the variance of Y_k is proportional to a known function $v(x_k)$ of the x_k are of considerable interest.

2 Notations

Let a finite population Ω of size N consists N_1 respondents and N_2 non-respondent units and a sample of size n, drawn from the population, consists n_1 respondents and n_2 non-respondents. Further let h_2 non-respondent units are randomly selected from the n_2 non-respondents and all efforts were made to take information from these non-respondents.

Let the samples of sizes n and h_2 be denoted by s and R_{h_2} respectively. Let $\Omega = R \cup \bar{R}$ where R and $\bar{R}$ be two disjoint sets of units, implying that $\bar{R}$ is the non-observed part of the population Ω. Similarly let $R = R_1 \cup R_2$ (R_1 and R_2 being the disjoint sets of units) and $R_2 = R_{h_2} \cup \bar{R}_{h_2}$, where $\bar{R}_{h_2}$ be the remaining part of R_2. Let $\sum_p$ stands for the summation over the set (subset) p of units. We define the following notations:

Let Z = Variable Y or X

$$\bar{Z} = N^{-1} \sum_{\Omega} z_k: \text{The population mean of Z}$$

$$S_Z^2 = (N-1)^{-1} \sum_{\Omega} (z_k - \bar{Z})^2: \text{Population mean square of Z.}$$

Sample means:

$$\bar{z} = n^{-1} \sum_R z_k, \quad \bar{z}_{\bar{R}} = (N-n)^{-1} \sum_{\bar{R}} z_k;$$

$$\bar{z}_{R_i} = n_i^{-1} \sum_{R_i} z_k, \quad (i = 1, 2); \quad \bar{z}_{R_{h_2}} = h_2^{-1} \sum_{R_{h2}} z_k;$$

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

Further let for the variable X

$$\bar{X}^{(j)} = N^{-1} \sum_{\Omega} x_k^j; \quad \bar{x}^{(j)} = n^{-1} \sum_R x_k^j; \quad \bar{x}_{\bar{R}}^{(j)} = (N-n)^{-1} \sum_{\bar{R}} x_k^j;$$

$$\bar{x}_{R_i}^{(j)} = n_i^{-1} \sum_{R_i} x_k^j; \quad (i = 1, 2)$$

$$\bar{x}_{R_{h_2}}^{(j)} = (h_2)^{-1} \sum_{R_{h2}} x_k^j; \quad \bar{x}_{\bar{R}_{h_2}}^{(j)} = (n_2 - h_2)^{-1} \sum_{\bar{R}_{h_2}} x_k^j \quad \text{for } j = 1, 2, 3 \ldots$$

Obviously we observed that

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

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

H-H (1946) proposed an unbiased estimator of population mean $\bar{Y}$ as follows:

$$\bar{y}^* = \frac{n_1 \bar{y}_{R_1} + n_2 \bar{y}_{R_{h_2}}}{n}. \tag{2}$$

Obviously, we have

$$E(\bar{y}_{R_{h_2}}) = \bar{y}_{R_2} \tag{3}$$

and

$$E(\bar{y}^*) = \bar{Y} \tag{4}$$

If the information on an auxiliary character is available for each and every unit of the population, it could be assumed to be known for every unit in the samples of sizes n and h_2 also. We can, therefore, utilize this information in order to define the estimator

$$\hat{\bar{y}}_{R_2} = \bar{y}_{R_{h_2}} \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) \tag{5}$$

for estimating the unknown sample mean $\bar{y}_{R_2}$ of the non-respondents present in the sample of size n_2. Based upon the estimator $\hat{\bar{y}}_{R_2}$ and $\bar{y}_{R_1}$, let us now define the estimator

$$T = \frac{n_1}{n} \bar{y}_{R_1} + \frac{n_2}{n} \hat{\bar{y}}_{R_2} \tag{6}$$

which is an alternative to the estimator defined in (7) and might be considered as an extension of the Hansen and Hurwitz estimator $\bar{y}^*$.

3 Proposed Family of Estimator

3.1 The Family $T_P^*(\delta)$

We define here a one-parameter family of estimators for population mean $\bar{Y}$ in the presence of non-response as

$$T_P^*(\delta) = T\psi^*(\delta, \bar{X}, \bar{x}) \tag{7}$$

where $\psi^*(\delta, \bar{X}, \bar{x})$ is a function of the parameter δ, population mean $\bar{X}$ and sample mean $\bar{x}$ based on the sample of size n of the auxiliary variable X and is given by

$$\psi^*(\delta, \bar{X}, \bar{x}) = \exp \left[\delta \left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}} \right) \right] \tag{8}$$

Remark 1: The function $\psi^*(\delta, \bar{X}, \bar{x})$ is developed on the lines of exponential-type estimator suggested by Bahl and Tuteja (1991).

Remark 2: Some special cases of $T_P^*(\delta)$ is of importance. Letting $\delta = 0$, we have

$$T_P^*(0) = T \tag{9}$$

which are similar to the estimators defined by H-H (1946). Further, for $\delta = 1$ and -1, $T_P^*(1)$ and $T_P^*(-1)$ respectively yield exponential-type ratio

and product estimators, proposed by Bahl and Tuteja (1991), assuming that H-H (1946) technique of sub-sampling of non-respondents was followed for tackling the problem of non-response.

4 ξ-Bias and ξ-MSE of $T_P^*(\delta)$ Under PRM $\xi[\delta_0, \delta_1, \ldots, \delta_J : v(x)]$

We now obtain the ξ-Bias and ξ-MSE of the general family $T_P^*(\delta)$ under super population approach considering the PRM $\xi[\delta_0, \delta_1, \ldots, \delta_J : v(x)]$.

***Theorem 1*:** The ξ-bias of the estimator $T_P^*(\delta)$ is

$$B_\xi[T_P^*(\delta)] = \sum_{j=0}^{J} \delta_j \beta_j \left[\psi(\delta, \bar{X}, \bar{x}) \left\{ \frac{n_1}{n} \bar{x}_{R_1}^{(j)} + \frac{n_2}{n} \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) \bar{x}_{R_{h_2}}^{(j)} \right\} - \bar{X}^{(j)} \right]$$

(10)

The proof of the expression (10) is presented in the *Appendix, Section-I*.

***Theorem 2*:** The ξ-MSE of the estimator $T_P^*(\delta)$ is given by

$M_\xi(T_P^*(\delta))$

$$= \left[\sum_{j=0}^{J} \delta_j \beta_j \left\{ \psi(\delta, \bar{X}, \bar{x}) \left(\frac{n_1}{n} \bar{x}_{R_1}^{(j)} + \frac{n_2}{n} \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) \bar{x}_{R_{h_2}}^{(j)} \right) - \bar{X}^{(j)} \right\} \right]^2$$

$$+ \left[\frac{\psi(\delta, \bar{X}, \bar{x})}{n} - \frac{1}{N} \right]^2 \sigma^2 \sum_{R_1} v(x_k)$$

$$+ \left[\frac{f_2}{n} \psi(\delta, \bar{X}, \bar{x}) \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) - \frac{1}{N} \right]^2 \sigma^2 \sum_{R_{h_2}} v(x_k)$$

$$+ \frac{\sigma^2}{N^2} \left[\sum_{\bar{R}_{h_2}} v(x_k) + \sum_{\bar{R}} v(x_k) \right]$$

(11)

The proof of the above expression is presented in the *Appendix, Section-II*.

5 Some Particular Cases of the Model $\xi[\delta_0, \delta_1, \delta_2, \ldots, \delta_J: v(x)]$

5.1

Some Particular Cases of the General PRM $\xi[\delta_0, \delta_1, \delta_2, \ldots, \delta_J: v(x)]$ may be considered with changes in two different components, viz., in the function $h(x_k) = \sum_{j=0}^{J} \delta_j \beta_j x_k^j$ and in the function $v(x_k)$. For example, some particular forms of the general model may be taken as

$$Y_k = \beta_1 x_k + \in_k [x_k]^{1/2} \tag{12}$$

$$Y_k = \beta_0 + \beta_1 x_k + \in_k [x_k^2]^{1/2} \tag{13}$$

$$Y_k = \beta_0 + \beta_1 x_k + \beta_2 x_k^2 + \in_k [x_k^g]^{1/2}. \tag{14}$$

Obviously models (12), (13) and (14) may be denoted respectively as $\xi[0, 1:x]$, $\xi[1, 1:x^2]$ and $\xi[1, 1, 1:x^g]$. Cochran (1953) and Brewer (1963) have shown that if $v(x_k)$ function is taken to be x_k^g with $0 \leq g \leq 2$ then majority of the situations occurring in practice might be covered as far as the variance function of the model is concerned. Under this consideration, therefore, we may consider the following six PRMs for further analysis:

Model I $\qquad \xi[0, 1:1] \Rightarrow h(x_k) = \beta_1 x_k, v(x_k) = x_k^0,$ $\qquad (15)$

Model II $\qquad \xi[0, 1:x] \Rightarrow h(x_k) = \beta_1 x_k, v(x_k) = x_k,$ $\qquad (16)$

Model III $\qquad \xi[0, 1:x^2] \Rightarrow h(x_k) = \beta_1 x_k, v(x_k) = x_k^2,$ $\qquad (17)$

Model IV $\qquad \xi[1, 1:1] \Rightarrow h(x_k) = \beta_0 + \beta_1 x_k, v(x_k) = x_k^0,$ $\qquad (18)$

Model V $\qquad \xi[1, 1:x] \Rightarrow h(x_k) = \beta_0 + \beta_1 x_k, v(x_k) = x_k,$ $\qquad (19)$

Model VI $\qquad \xi[1, 1:x^2] \Rightarrow h(x_k) = \beta_0 + \beta_1 x_k, v(x_k) = x_k^2.$ $\qquad (20)$

Royall and Herson (1973a) have shown that under the Model II: $\xi[0, 1:x]$, the conventional ratio estimator becomes unbiased while it is biased under design approach. On the other hand, contrary to fixed population approach, the sample mean estimator is not unbiased under the model $\xi[0, 1:x]$. All other models also have their importance and significance under different situations.

The expression for MSE of $T_P^*(\delta)$ under the six models are given as

$$M_\xi[T_P^*(\delta)]_I = \left[\beta_1 \left\{ \psi(\delta, \bar{X}, \bar{x}) \left(\frac{n_1}{n} \bar{x}_{R_1} + \frac{n_2}{n} \bar{x}_{R_2} \right) - \bar{X} \right\} \right]^2$$

$$+ \left[\frac{1}{n} \psi(\delta, \bar{X}, \bar{x}) - \frac{1}{N} \right]^2 n_1 \sigma^2$$

$$+ \left[\frac{f_2}{n} \psi(\delta, \bar{X}, \bar{x}) \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) - \frac{1}{N} \right]^2 h_2 \sigma^2$$

$$+ \frac{\sigma^2}{N^2} \{ (N - n) + (n_2 - h_2) \}. \tag{21}$$

$$M_\xi[T_P^*(\delta)]_{II} = \left[\beta_1 \left\{ \psi(\delta, \bar{X}, \bar{x}) \left(\frac{n_1}{n} \bar{x}_{R_1} + \frac{n_2}{n} \bar{x}_{R_2} \right) - \bar{X} \right\} \right]^2$$

$$+ \left[\frac{1}{n} \psi(\delta, \bar{X}, \bar{x}) - \frac{1}{N} \right]^2 \sigma^2 \sum_{R_1} x_k$$

$$+ \left[\frac{f_2}{n} \psi(\delta, \bar{X}, \bar{x}) \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) - \frac{1}{N} \right]^2 \sigma^2 \sum_{R_{h_2}} x_k$$

$$+ \frac{\sigma^2}{N^2} \left\{ \sum_{\bar{R}_{h_2}} x_k + \sum_{\bar{R}} x_k \right\}. \tag{22}$$

$$M_\xi[T_P^*(\delta)]_{III} = \left[\beta_1 \left\{ \psi\left(\delta, \bar{X}, \bar{x}\right) \left(\frac{n_1}{n} \bar{x}_{R_1} + \frac{n_2}{n} \bar{x}_{R_2} \right) - \bar{X} \right\} \right]^2$$

$$+ \left[\frac{1}{n} \psi(\delta, \bar{X}, \bar{x}) - \frac{1}{N} \right]^2 \sigma^2 \sum_{R_1} x_k^2$$

$$+ \left[\frac{f_2}{n} \psi(\delta, \bar{X}, \bar{x}) \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) - \frac{1}{N} \right]^2 \sigma^2 \sum_{R_{h_2}} x_k^2$$

$$+ \frac{\sigma^2}{N^2} \left\{ \sum_{\bar{R}_{h_2}} x_k^2 + \sum_{\bar{R}} x_k^2 \right\}. \tag{23}$$

$$M_\xi[T_P^*(\delta)]_{IV} = \left[\beta_0 \left\{ \psi\left(\delta, \bar{X}, \bar{x}\right) \left(\frac{n_1}{n} + \frac{n_2}{n} \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) \right) - 1 \right\} \right.$$

$$\left. + \beta_1 \left\{ \psi\left(\delta, \bar{X}, \bar{x}\right) \left(\frac{n_1}{n} \bar{x}_{R_1} + \frac{n_2}{n} \bar{x}_{R_2} \right) - \bar{X} \right\} \right]^2$$

$$+ \left[\frac{1}{n} \psi(\delta, \bar{X}, \bar{x}) - \frac{1}{N} \right]^2 n_1 \sigma^2$$

$$+ \left[\frac{f_2}{n} \psi(\delta, \bar{X}, \bar{x}) \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) - \frac{1}{N} \right]^2 h_2 \sigma^2$$

$$+ \frac{\sigma^2}{N^2} \{ (N - n) + (n_2 - h_2) \}. \tag{24}$$

$$M_\xi[T_P^*(\delta)]_V = \left[\beta_0 \left\{ \psi\left(\delta, \bar{X}, \bar{x}\right) \left(\frac{n_1}{n} + \frac{n_2}{n} \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) \right) - 1 \right\} \right.$$

$$\left. + \beta_1 \left\{ \psi(\delta, \bar{X}, \bar{x}) \left(\frac{n_1}{n} \bar{x}_{R_1} + \frac{n_2}{n} \bar{x}_{R_2} \right) - \bar{X} \right\} \right]^2$$

$$+ \left[\frac{1}{n} \psi(\delta, \bar{X}, \bar{x}) - \frac{1}{N} \right]^2 \sigma^2 \sum_{R_1} x_k$$

$$+ \left[\frac{f_2}{n} \psi(\delta, \bar{X}, \bar{x}) \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) - \frac{1}{N} \right]^2 \sigma^2 \sum_{R_{h_2}} x_k$$

$$+ \frac{\sigma^2}{N^2} \left\{ \sum_{\bar{R}_{h_2}} x_k + \sum_{\bar{R}} x_k \right\} \tag{25}$$

$$M_\xi[T_P^*(\delta)]_{VI} = \left[\beta_0 \left\{ \psi\left(\delta, \bar{X}, \bar{x}\right) \left(\frac{n_1}{n} + \frac{n_2}{n} \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) \right) - 1 \right\} \right.$$

$$\left. + \beta_1 \left\{ \psi(\delta, \bar{X}, \bar{x}) \left(\frac{n_1}{n} \bar{x}_{R_1} + \frac{n_2}{n} \bar{x}_{R_2} \right) - \bar{X} \right\} \right]^2$$

$$+ \left[\frac{1}{n} \psi(\delta, \bar{X}, \bar{x}) - \frac{1}{N} \right]^2 \sigma^2 \sum_{R_1} x_k^2$$

$$+ \left[\frac{f_2}{n} \psi(\delta, \bar{X}, \bar{x}) \left(\frac{\bar{x}_{R_2}}{\bar{x}_{R_{h_2}}} \right) - \frac{1}{N} \right]^2 \sigma^2 \sum_{R_{h_2}} x_k^2$$

$$+ \frac{\sigma^2}{N^2} \left\{ \sum_{\bar{R}_{h_2}} x_k^2 + \sum_{\bar{R}} x_k^2 \right\}. \tag{26}$$

6 Some Existing Estimators and Their MSEs

6.1

Singh et al. (2017) defined two families of estimators $T_s^*(\delta)$ and $t_{NR}^*(\delta)$ under non-response and compared then under different polynomial regression models. The estimators and their MSEs under $\xi[\delta_0, \delta_1 \ldots, \delta_J: v(x)]$ are as follows:

(i) $\qquad T_s^*(\delta) = \bar{y}_w^* \psi^*(\delta, \bar{X}, \bar{x}) \tag{27}$

$$M_\xi(T_s^*(\delta)) = [B_\xi(T_s^*(\delta))]^2 + \left[\frac{\psi^*(\delta, \bar{X}, \bar{x})}{n} - \frac{1}{N} \right]^2 \sigma^2 \sum_{R_1} v(x_k)$$

$$+ \left[\frac{f_2}{n} \psi^*(\delta, \bar{X}, \bar{x}) - \frac{1}{N} \right]^2 \sigma^2 \sum_{R_{h_2}} v(x_k)$$

$$+ \frac{\sigma^2}{N^2} \left[\sum_{\bar{R}_{h_2}} v(x_k) + \sum_{\bar{R}} v(x_k) \right] \tag{28}$$

(ii) $\qquad t_{NR}^*(\delta) = \bar{y}_w^{**} \frac{\bar{X}}{\bar{x}} \tag{29}$

Where

$$\bar{y}_w^{**} = \frac{n_1 \bar{y}_{R_1} + n_2 t_{R_2}(\delta)}{n}. \tag{30}$$

$$t_{R_2}(\delta) = \bar{y}_{R_{h_2}} \exp \left[\delta \left(\frac{\bar{x}_{R_2} - \bar{x}_{R_{h_2}}}{\bar{x}_{R_2} + \bar{x}_{R_{h_2}}} \right) \right] \tag{31}$$

with

$$M_\zeta[t_{NR}^*(\delta)] = \left[\sum_{j=0}^{J} \delta_j \beta_j \left(\frac{\bar{X}}{n\bar{x}} (n_1 \bar{x}_{R_1}^{(j)} + n_2 \psi(\delta, \bar{\mathrm{x}}_{R_2}, \bar{x}_{R_{h_2}}) \bar{x}_{R_{h_2}}^{(j)}) - \bar{X}^{(j)} \right) \right]^2$$

$$+ \left(\frac{\bar{X}}{n\bar{x}} - \frac{1}{N} \right)^2 \sigma^2 \sum_{R_1} v(x_k)$$

$$+ \left[\frac{n_2}{nh_2} \psi(\delta, \bar{\mathrm{x}}_{R_2}, \bar{x}_{R_{h_2}}) \frac{\bar{X}}{\bar{x}} - \frac{1}{N} \right]^2$$

$$\sigma^2 \sum_{R_{h_2}} v(x_k) + \frac{1}{N^2} \sigma^2 \sum_{\bar{R}_{h_2}} v(x_k) + \frac{1}{N^2} \sigma^2 \sum_{\bar{R}} v(x_k). \qquad (32)$$

7 Robustness and Efficiency Comparisons of $T_P^*(\delta)$

7.1

We made a study of the proposed family of estimators, $T_P^*(\delta)$ in terms of robustness criterion and compared it with the estimators $T_s^*(\delta)$ and t_{NR}^* for its efficiency. All these estimators have been developed utilizing the concept of sub-sampling of non-response in order to cope up the problem of non-response, inherent in the population and their ξ-bias and ξ-MSE are obtained under the general PRM $\xi[\delta_0, \delta_1, \delta_2, \ldots, \delta_J: v(x)]$ with different set-up of polynomial regression function $h(x_k)$ and variance function $v(x_k)$.

7.2

Since theoretical comparisons of ξ-MSEs of the estimators are not simple and any concrete conclusion can not be drawn, we have used an empirical data for this purpose. In order to make numerical comparison of robustness of the proposed estimator and MSE comparison of $T_P^*(\delta)$, $T_s^*(\delta)$ and t_{NR}^*, we have considered an empirical data presented in Singh et al. (2017). The data has been taken from Kish (1967). The details of the data have been given in Appendix-E of Kish (1967). For the data, we have the following particulars for non-response rates 15, and 30 percent respectively: The x_k represents the number of dwellings whereas y_k denotes dwelling occupied by renters.

For the data set, we obtained the following values:

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

(i) For 15 percent non-response rate:

$$n = 20, \quad n_1 = 17, \quad n_2 = 3, \quad h_2 = 2, \quad f_2 = 1.5, \quad \bar{x} = 39.55,$$

$$\bar{x}_{R_1} = 39.824, \quad \bar{x}_{R_2} = 38.0, \quad \bar{x}_{R_{h_2}} = 30.5, \quad \sum_{R_1} x_k = 677,$$

$$\sum_{R_{h2}} x_k = 61, \quad \sum_{\bar{R}_{h_2}} x_k = 53, \quad \sum_{\bar{R}} x_k = 2940, \quad \sum_{R_1} x_k^2 = 36729,$$

$$\sum_{R_{h_2}} x_k^2 = 2081, \quad \sum_{\bar{R}} x_k^2 = 179614, \quad \sum_{\bar{R}_{h_2}} x_k^2 = 2809.$$

(ii) For 30 percent non-response rate:

$$n = 20, \quad n_1 = 14, \quad n_2 = 6, \quad h_2 = 4, \quad f_2 = 1.5, \quad \bar{x} = 39.55,$$

$$\bar{x}_{R_1} = 36.5, \quad \bar{x}_{R_2} = 46.6667, \quad \bar{x}_{R_{h_2}} = 50, \quad \sum_{R_1} x_k = 511,$$

$$\sum_{R_{h2}} x_k = 200, \quad \sum_{\bar{R}_{h_2}} x_k = 80, \quad \sum_{\bar{R}} x_k = 2940, \quad \sum_{R_1} xk^2 = 25337,$$

$$\sum_{\bar{R}} xk^2 = 179614, \quad \sum_{\bar{R}_{h_2}} xk^2 = 3328.$$

7.3

Tables 1–2, 3–4 and 5–6 depict the variations in MSEs of estimators $T_s^*(\alpha)$, $T_P^*(\delta)$ and t_{NR}^* respectively for $\delta = 0, 1$ and -1 with 15, and 30 percent non-response rates over Models I–VI.

7.4 Simulation Study

Here we have used simulation study for the data given above. We have drawn 30000 times samples from the population of size 90 and take a sample of size 20 and use 15% non-response rate to find the MSEs of estimators for $\delta = 0, 1, -1$.

Table 1 Comparisons of MSEs of estimators with 15% non-response rate for $\delta = 0$ over Models I–VI

g	Model	$T_s^*(\delta)$	t_{NR}^*	$T_P^*(\delta)$
0	Model-I	221.9701	33.63682	221.9745
1	Model-II	223.2496	35.04206	223.3812
2	Model-III	291.1687	109.2985	295.6571
0	Model-IV	221.9701	34.13085	221.9745
1	Model-V	223.2496	35.53468	223.3812
2	Model-VI	291.1687	109.7911	295.6571

Table 2 Comparisons of MSEs of estimators with 30% non-response rate for $\delta = 0$ over Models I–VI

g	Model	$T_s^*(\delta)$	t_{NR}^*	$T_P^*(\delta)$
0	Model-I	19.85436	26.58954	19.8524
1	Model-II	21.38647	28.27287	21.28827
2	Model-III	110.8182	126.3833	104.4577
0	Model-IV	19.85436	26.15504	19.8524
1	Model-V	21.38647	27.83837	21.28827
2	Model-VI	110.8179	125.9488	104.4577

Table 3 Comparisons of MSEs of estimators with 15% non-response rate for d$\delta = 1$ over Models I–VI

g	Model	$T_s^*(\delta)$	t_{NR}^*	$T_P^*(\delta)$
0	Model-I	108.2256	9.458615	108.2302
1	Model-II	109.5656	10.9243	109.704
2	Model-III	180.5684	87.10091	185.2896
0	Model-IV	108.6612	9.820009	108.6657
1	Model-V	110.0011	11.28569	110.1395
2	Model-VI	181.004	87.65979	185.7251

Table 4 Comparisons of MSEs of estimators with 30% non-response rate for $\delta = 1$ over Models I–VI

g	Model	$T_s^*(\delta)$	t_{NR}^*	$T_P^*(\delta)$
0	Model-I	0.124783	6.455189	0.1227161
1	Model-II	1.730512	8.08324	1.627152
2	Model-III	95.3875	102.5966	88.69285
0	Model-IV	0.112961	6.28924	0.1108941
1	Model-V	1.71869	7.917298	1.615329
2	Model-VI	95.3756	102.4307	88.68103

Table 5   Comparisons of MSEs of estimators with 15% non-response rate for $\delta = -1$ over Models I–VI

g	Model	$T_s^*(\delta)$	t_{NR}^*	$T_P^*(\delta)$
0	Model-I	372.1006	67.9551	372.1047
1	Model-II	373.3227	69.31012	373.4478
2	Model-III	438.3187	141.6982	442.5852
0	Model-IV	371.3129	68.71755	371.317
1	Model-V	372.535	69.77258	372.6601
2	Model-VI	437.531	142.3581	441.7975

Table 6   Comparisons of MSEs of estimators with 30% non-response rate for $\delta = -1$ over Models I–VI

g	Model	$T_s^*(\delta)$	t_{NR}^*	$T_P^*(\delta)$
0	Model-I	82.59208	61.88629	82.59022
1	Model-II	84.05434	63.62918	83.96106
2	Model-III	169.4755	165.6155	163.4338
0	Model-IV	82.22125	61.0706	82.21938
1	Model-V	83.68351	62.81353	83.59023
2	Model-VI	169.1047	164.7998	163.0629

Table 7   Comparisons of MSEs of estimators with 15% non-response rate for $\delta = 0$ over Models I–VI

g	Model	$T_s^*(\delta)$	t_{NR}^*	$T_P^*(\delta)$
0	Model-I	1927.238	1727.755	558.0646
1	Model-II	1928.518	1729.029	559.3441
2	Model-III	1996.437	1796.658	627.2632
0	Model-IV	1927.238	1727.589	558.0646
1	Model-V	1928.518	1728.863	559.3441
2	Model-VI	1996.437	1796.492	627.2632

Table 8   Comparisons of MSEs of estimators with 15% non-response rate for $\delta = 1$ over Models I–VI

g	Model	$T_s^*(\delta)$	t_{NR}^*	$T_P^*(\delta)$
0	Model-I	1724.227	1124.534	53.7949
1	Model-II	1725.227	1125.977	55.31391
2	Model-III	1795.378	1199.191	135.4364
0	Model-IV	1725.312	1127.193	54.9664
1	Model-V	1726.629	1128.636	56.48532
2	Model-VI	1780.501	1202.047	136.6079

Table 9   Comparisons of MSEs of estimators with 15% non-response rate for $\delta = -1$ over Models I–VI

g	Model	$T_s^*(\delta)$	t_{NR}^*	$T_P^*(\delta)$
0	Model-I	2138.325	2272.202	1487.107
1	Model-II	2139.568	2273.383	1488.19
2	Model-III	2205.636	2337.634	1546.085
0	Model-IV	2137.135	2268.999	1481.493
1	Model-V	2138.378	2270.18	1482.576
2	Model-VI	2204.446	2334.629	1540.47

8 Conclusions

From the tables some conclusions can be drawn about the family $T_P^*(\delta)$ for $\delta = 0, -1$ and 1 regarding its robustness and precision as compared with other estimators. There are as follows:

 (i) As far as the robustness property of the estimator $T_P^*(\delta)$ is considered, it can be concluded that the data give sufficient evidence that the proposed family seems to be robust enough under the models I, II, IV and V, where the polynomial regression function changes and variance function is of the form x_k^g, $g = 0, 1$. Similarly, the estimator seems to be robust under models III and VI where variance function is proportional to x_k^2. It is also to be emphasized that this property of the estimator holds good irrespective of "the non-response rate" and the choice of the parameter δ.

 (ii) However, as the value of g increases, the performance of the estimators, irrespective of the model-choice and the choice of the parametric value, decreases.

(iii) From the Tables 7, 8 and 9 for 15% non-response rate, we conclude that this is similar to Tables 1, 3 and 5.

Appendix

Section I: We have

$$B_\xi[T_P^*(\delta)] = E_\xi[T_P^*(\delta) - \overline{Y}]$$

$$= E_\xi[\bar{y}_w^* \psi^*(\delta, \bar{X}, \bar{x}) - \overline{Y}]$$

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

$$= E_\xi \left[\psi^*(\delta, \bar{X}, \bar{x}) \left(\frac{n_1}{n} \frac{1}{n_1} \sum_{R_1} y_k + \frac{n_2}{n} \frac{1}{h_2} \sum_{R_{h_2}} y_k \right) - \frac{1}{N} \sum_{k=1}^{N} y_k \right]$$

$$\text{(B1)}$$

Now, using the PRM given in (1), we can write

$$B_\xi[T_P^*(\delta)] = E_\xi \left[\psi^*(\delta, \bar{X}, \bar{x}) \frac{1}{n} \sum_{R_1} \left(\sum_{j=0}^{J} \delta_j \beta_j x_k^j + \in_k (v(x_k))^{1/2} \right) \right.$$

$$+ \psi^*(\delta, \bar{X}, \bar{x}) \frac{f_2}{n} \sum_{R_{h_2}} \left(\sum_{j=0}^{J} \delta_j \beta_j x_k^j + \in_k (v(x_k))^{1/2} \right)$$

$$\left. - \frac{1}{N} \sum_{\Omega} \left(\sum_{j=0}^{J} \delta_j \beta_j x_k^j + \in_k (v(x_k))^{1/2} \right) \right]. \qquad \text{(B2)}$$

Since $E_\xi(\in_k) = 0$ for all k, we have

$$B_\xi[T_P^*(\delta)] = \psi^*(\delta, \bar{X}, \bar{x}) \left(\sum_{j=0}^{J} \delta_j \beta_j \frac{1}{n} \sum_{R_1} x_k^j + \sum_{j=0}^{J} \delta_j \beta_j \frac{f_2}{n} \sum_{R_{h_2}} x_k^j \right)$$

$$- \sum_{j=0}^{J} \delta_j \beta_j \frac{1}{N} \sum_{\Omega} x_k^j$$

$$= \sum_{j=0}^{J} \delta_j \beta_j \left[\psi^*(\delta, \bar{X}, \bar{x}) \left(\frac{n_1 \bar{x}_{R_1}^{(j)} + n_2 \bar{x}_{R_{h_2}}^{(j)}}{n} \right) - \bar{X}^{(j)} \right].$$

$$\text{(B3)}$$

Thus expression (10) follows.

Section II

"The ξ-MSE of the estimator" $T_P^*(\delta)$ "under the model 1 is derived as follows":

We have

$$M_\xi(T_P^*(\delta)) = E_\xi[T_P^*(\delta) - \bar{Y}]^2$$

$$= E_\xi[\bar{y}_w^* \psi^*(\delta, \bar{X}, \bar{x}) - \bar{Y}]^2$$

$$= E_\xi\left[\psi^*(\delta, \bar{X}, \bar{x})\left(\frac{1}{n}\sum_{R_1} y_k + \frac{f_2}{n}\sum_{R_{h_2}} y_k\right) - \frac{1}{N}\sum_{k=1}^N Y_k\right]^2$$

$$= E_\xi\left[\psi^*(\delta, \bar{X}, \bar{x})\frac{1}{n}\sum_{R_1}\left(\sum_{j=0}^J \delta_j \beta_j x_k^j + \in_k (v(x_k))^{1/2}\right)\right.$$

$$+ \psi^*(\delta, \bar{X}, \bar{x})\frac{f_2}{n}\sum_{R_{h_2}}\left(\sum_{j=0}^J \delta_j \beta_j x_k^j + \in_k (v(x_k))^{1/2}\right)$$

$$\left.- \frac{1}{N}\sum_\Omega\left(\sum_{j=0}^J \delta_j \beta_j x_k^j + \in_k (v(x_k))^{1/2}\right)\right]^2. \qquad (B4)$$

"Now realizing that" "$E_\xi(\in_k, \in_r) = 0$ for $r \neq k$ and" $E_\xi(\in_k^2) = \sigma^2$, "we have"

$$M_\xi(T_P^*(\delta)) = \left[\sum_{j=0}^J \delta_j \beta_j\left(\psi^*(\delta, \bar{X}, \bar{x})\left(\frac{n_1\bar{x}_{R_1}^{(j)} + n_2\bar{x}_{R_{h_2}}^{(j)}}{n}\right) - \bar{X}^{(j)}\right)\right]^2$$

$$+ \left[\frac{\psi^*(\delta, \bar{X}, \bar{x})}{n} - \frac{1}{N}\right]^2 \sigma^2 \sum_{R_1} v(x_k)$$

$$+ \left[\frac{f_2}{n}\psi^*(\delta, \bar{X}, \bar{x}) - \frac{1}{N}\right]^2 \sigma^2 \sum_{R_{h_2}} v(x_k)$$

$$+ \frac{\sigma^2}{N^2}\left[\sum_{\bar{R}_{h_2}} v(x_k) + \sum_{\bar{R}} v(x_k)\right]. \qquad (B5)$$

Expression (B5) can further be written as

$$M_\xi(T_P^*(\delta)) = [B_\xi(T_P^*(\delta))]^2 + \left[\frac{\psi^*(\delta,\bar{X},\bar{x})}{n} - \frac{1}{N}\right]^2 \sigma^2 \sum_{R_1} v(x_k)$$

$$+ \left[\frac{f_2}{n}\psi^*(\delta,\bar{X},\bar{x}) - \frac{1}{N}\right]^2 \sigma^2 \sum_{R_{h_2}} v(x_k)$$

$$+ \frac{\sigma^2}{N^2}\left[\sum_{\bar{R}_{h_2}} v(x_k) + \sum_{\bar{R}} v(x_k)\right]. \qquad \text{(B6)}$$

Hence the expression (11) follows.

References

Bahl, S. Tuteja, R. K. (1991): Ratio and product type-exponential estimator, *Information and Optimization Sciences*, XII, I, 159–163.

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

Brewer, K. R. W. (1963): Ratio estimation and finite populations: Some results deducible from the assumption of an underlying stochastic process, *Australian Journal of Statistics*, 5, 93–105.

Cassel, C. M., Sarndal, C. E. and Wretman, J. H. (1976): Some results on generalized difference estimation and generalized regression estimation for finite populations, *Biometrika*, 63, 615–620.

Chambers, R.L. (1986): Outlier robust finite population estimation, *Journal of the American Statistical Association*, 81(396), 1063–1069.

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

Deming, W. E. (1944): On errors in surveys, *American Sociological Review*, 9, 359–369.

Hansen, M. H. and Hurwitz, W. N. (1946): The problem of non-response in sample surveys, *Journal of the American Statistical Association*, 41, 517–529.

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

Mahalanobis, P. C. (1946): Recent experiments in statistical sampling in The Indian Statistical Institute, *Journal of The Royal Statistical Society*, 109A, 325–378.

Moser, C. A. (1958): *Survey Methods in Social Investigation*, Heinemann, London.

Royall, R. M. (1971): Linear regression models in finite population sampling theory, in *Foundations of Statistical Inference*, V. P. Godambe and D. A. Sprott (eds.), Toronto: Holt, Rinehart and Winston, 259–274.

Royall, R. M. and Herson, J. (1973a): Robust estimation in finite populations I, *Journal of the American Statistical Association*, 68(344), 880–889.

Royall, R. M. and Herson, J. (1973b): Robust estimation in finite populations II: Stratification on a size variable, *Journal of the American Statistical Association*, 68(344), 890–893.

Singh H.P., Solanki, R. S. (2011): Estimation of finite population mean using random non-response in survey sampling, *Pakistan Journal of Statistics and Operation Research*, 7(1), 21–41.

Singh H.P., Solanki, R. S. (2011): A General procedure for estimating the population parameter in the presence of random non-response. *Pakistan Journal of Statistics*, 27(4), 427–465 (2011).

Singh, V. K., Singh, R. V. K. and Shukla, R. K. (2009b): Model-based study of some estimators in the presence of non-response, in Population, Poverty and Health: Analytical Approaches (Eds. K. K. Singh, R. C. Yadava and Arvind Pandey), Hindustan Publishing Corporation, New Delhi, India, 360–365.

Singh, A. K. Singh, Priyanka and Singh, V. K. (2017): Model based study of families of exponential type estimators in presence of nonresponse, *Communications in Statistics – Theory and Methods*, 46,13, 6478–6490.

Shukla, R. K. (2010): Model-Based Efficiencies of Some Families of Estimators in Presence of Non-Response and Measurement Errors, Unpublished, Ph.D. Thesis submitted to Banaras Hindu University, Varanasi, India.

Zarkovich, S. S. (1966): *Quality of Statistical Data*, Food and Agricultural Organization of the United Nations, Rome.

Biographies

Ajeet Kumar Singh is Assistant Professor in Department of Statistics, University of Rajasthan, Jaipur. He received the Ph.D degree in Statistics from Banaras Hindu University. He has published more than 20 research articles in reputed international journals. Field of specializations is Sampling Theory.

V. K. Singh is retired Professor, from Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India since 2000. Joined the Department as Assistant Professor in 1972. Did M.Sc (Statistics) and Ph.D. (Statistics) from Banaras Hindu University in 1972 and 1979 respectively. Having 45 years teaching experience and 43 years research experience. Field of specializations is Sampling Theory, Stochastic Modelling, Mathematical Demography and Operations Research. Published 92 research papers in reputed international/national journals. Guided 15 Ph.D. scholars for their Ph.D. Degree. Visited United Kingdom, Australia and Sri Lanka for attending International Conferences and organizing Symposiums. Convened 2 national conferences. Life member of Indian Statistical Association, Member of International Association of Survey Statisticians (IASS), Associate Editor of Assam Statistical Review, India.