AN EFFICIENT ESTIMATION PROCEDURE FOR THE POPULATION MEAN UNDER NON-RESPONSE

Shashi Bhushan
Department of Mathematics and Statistics, Dr Shakuntala Misra National Rehabilitation University, Lucknow, India

Abhay Pratap Pandey 1
Department of Statistics, Ramanujan College University of Delhi, New Delhi, India

1. INTRODUCTION

Usually almost all surveys covering human populations suffer from the problem of non-response. Lack of information, absence at the time of survey, and refusal of the respondents are main reasons of the non-response. Non-respondents differ significantly from the respondents. Hansen and Hurwitz (1946) suggested a procedure of taking a sample from the non-respondent and collecting information by a more expensive method like first attempt by mail questionnaire and second attempt by personal interview. It is well known that utilizes the auxiliary information to neutralize the effect of non-response for estimating the population mean. The auxiliary information closely related to the main characteristics plays a very important role in estimation of population characteristics. The parameters can be estimated more accurately by making use such information on auxiliary variable. The ratio, product and regression methods of estimation and their generalizations are good examples in this context. Using the procedure envisaged by Hansen and Hurwitz (1946) various authors including Singh (1965, 1967), Shukla (1966), Ray and Singh (1981), Srivastava and Jhajj (1981), Diana and Tomasi (2003), Singh et al. (1994), Bahl and Tuteja (1991), Bhushan and Gupta (2015), Cochran (1977), Rao (1983), Khare and Srivastava (1995, 1997), Okafor and Lee (2000), Singh et al. (2010), Singh and Kumar (2008) have suggested improvement in the estimation procedure for population mean in presence of non-response.

An important finding of all these papers was that the difference or the corresponding regression type estimators were found to be best in terms of MSE and any ratio type estimator can at best attains the MSE of regression (difference) estimator. In this paper,

1 Corresponding Author. E-mail: abhay.pandey@ramanujan.du.ac.in
we have proposed some improvement over these estimators proposed by various authors in their earlier works.

This paper is arranged as follows: Section 2 provides a brief review of some existing estimators along with some other important results; in Section 3, we propose seven new improved estimators using auxiliary variable and the results are reported; Section 4 deals with a comparative study of the proposed estimators using auxiliary variable in comparison to conventional estimators includes an empirical study and in Section 5, some concluding remarks are made.

2. Notations and Existing Estimators

Consider a finite population mean \( U = (U_1, U_2, ..., U_N) \) of \( N \) identifiable units in the sense that they are uniquely labeled from 1 to \( N \) and the label on each unit is known. Let \((y_i, x_i)\) be the study and the auxiliary variables taking values \((y_i, x_i)\) on the \( i \)-th population units \( U_i, i = 1, 2, ..., N \). Let \( n \) be the size of a sample drawn from the population of size \( N \) by using simple random sampling without replacement (SRSWOR), where \( n_1 \) of \( n \) units respond and \( n_2 \) sample units do not respond. From the \( n_2 \) non-response units, \( r(n_2/k, k > 1) \) units are selected by making extra effort and thus giving \( n_1 + r \) observations of the study variable \( y \) in place of \( n \). In this procedure, the population \( \Omega \) of size of \( N \) is supposed to be composed of two strata, namely respondents stratum \( S_1 \) and non-respondents stratum \( S_2 \) such that \( \Omega = S_1 \cup S_2 \); having sizes \( N_1 \) and \( N_2 \) respectively. Without loss of generality, we label the data on study variable as \( \{y_i, i \in S_1\} \) for the response stratum, and as \( \{y_i, i \in S_2\} \) for the non-response stratum.

Let \( \bar{Y} = \sum_{i \in \Omega} y_i / N \) and \( S_y^2 = \sum_{i \in \Omega} (y_i - \bar{Y})^2 / (N - 1) \) denote the population mean and population variance, respectively. Let \( \bar{Y}_1 = \sum_{i \in S_1} y_i / N_1 \) and \( S_{y_1}^2 = \sum_{i \in S_1} (y_i - \bar{Y}_1)^2 / (N_1 - 1) \) denote the mean and variance of the response stratum, respectively. Similarly, let \( \bar{Y}_2 = \sum_{i \in S_2} y_i / N_2 \) and \( S_{y_2}^2 = \sum_{i \in S_2} (y_i - \bar{Y}_2)^2 / (N_2 - 1) \) denote the mean and variance of the non-response stratum respectively. The population mean can be written as \( \bar{Y} = W_1 \bar{Y}_1 + W_2 \bar{Y}_2 \), \( W_1 = N_1 / N \) and \( W_2 = N_2 / N \). The sample mean \( \hat{Y}_1 = \sum_{i=1}^{n_1} y_i / n_1 \) is unbiased for \( \bar{Y}_1 \) and hence biased for \( \bar{Y} \) having bias \( W_1(\bar{Y}_1 - \bar{Y}_2) \).

The Hansen and Hurwitz (1946) procedure is actually double sampling when strata sizes are not known (see Lohr, 1999) by effectively using the sample mean \( \hat{Y}_{2r} = \sum_{i=1}^{r} y_i / r \) which is unbiased for the mean \( \bar{Y}_2 \) of the \( n_2 \) units resulting in an unbiased estimator for the population mean \( \bar{Y} \) given by

\[
\hat{y}^* = w_1 \hat{Y}_1 + w_2 \hat{Y}_{2r},
\]

where \( w_1 = n_1 / n \) and \( w_2 = n_2 / n \). The variance of Hansen and Hurwitz (HH) mean \( \hat{y}^* \) is given by

\[
\text{Var}(\hat{y}^*) = \left( \frac{1-f}{n} \right) S_y^2 + \frac{W_2(k-1)}{n} S_{y_2}^2,
\]
where \( f = n/N \) & \( k = n_2/r \) are sampling fraction and inverse sub-sampling fraction respectively.

Let \( x_i (i = 1, 2, ..., N) \) denote an auxiliary variate correlated with study variate \( y_i (i = 1, 2, ..., N) \). The population mean of the auxiliary variable \( x \) is \( \bar{X} = \sum_{i=1}^{N} x_i / N \). Let \( \bar{X}_1 \) and \( \bar{X}_2 \) denote the means of the response and non-response groups. Let \( \bar{x} \) denote the mean of all the \( n \) units. Let \( \bar{x}_1 \) and \( \bar{x}_2 \) denote the means of the \( n_1 \) responding units and the \( n_2 \) non-responding units. Further let \( \bar{x}_{2r} = \sum_{i=1}^{r} x_i / r \) denote the mean of the sub-sampled units. The population variances of \( x \) and \( y \) are denoted by \( S^2_x \) and \( S^2_y \), and the population covariance by \( S_{xy} \). The population correlation coefficient is \( \rho = S_{xy} / S^2_x S^2_y \).

The unbiased estimator of the population mean \( \bar{X} \) of the auxiliary variable \( x \) is

\[
\bar{x}^* = w_1 \bar{x}_1 + w_2 \bar{x}_2. \tag{3}
\]

The variance of \( \bar{x}^* \) is given by

\[
\text{Var}(\bar{x}^*) = P S^2_x + Q S^2_{x2}, \tag{4}
\]

where \( S^2_{x2} = \sum_{i=N_1+1}^{N_1+N_2} (x_i - \bar{X}_2)^2 / (N_2 - 1). \)

Taking inspiration from some important works of Khare and Srivastava (1993, 1995, 1997), Okafor and Lee (2000), Singh and Kumar (2008, 2010), Cochran (1977), and Rao (1986) we have considered some regression (difference) type estimators under non-response. Due to paucity of space, we have not considered ratio, product or ratio product type estimators and have considered only the regression (difference) estimators as they are the BLUE within each category. Also, for better understanding, we have divided them under four different strategies given below.

- **Strategy I**: \( \bar{y}^*, \bar{x}^* \) and \( \bar{X} \) are used. The non-response occurs on both the study variable \( y \) and auxiliary variable \( x \), and the population mean \( \bar{X} \) of the auxiliary variable is known. The difference and regression type estimators are

\[
t_1 = \bar{y}^* + K_1 (\bar{X} - \bar{x}^*) \tag{5}
\]

\[
t_{lr_1} = \bar{y}^* + b^* (\bar{X} - \bar{x}^*). \tag{6}
\]

- **Strategy II**: \( \bar{y}^*, \bar{x} \) and \( \bar{X} \) are used. The non-response occurs on the study variable \( y \), and information on the auxiliary variable \( x \) is available from all the sample units along the population mean \( \bar{X} \) of the auxiliary variable is known. The difference and regression type estimators are

\[
t_2 = \bar{y}^* + K_2 (\bar{X} - \bar{x}) \tag{7}
\]

\[
t_{lr_2} = \bar{y}^* + b (\bar{X} - \bar{x}). \tag{8}
\]
• **Strategy III**: $\bar{y}^*$, $\bar{x}$ and $\tilde{x}$ are used. The non-response occurs on the study variable $y$, and the information on the auxiliary variable $x$ is obtained from all the sample units, but the population mean $\bar{X}$ of the auxiliary variable is not known. The difference and regression type estimators are

\[ t_3 = \bar{y}^* + K_3 (\bar{x} - \tilde{x}^*). \tag{9} \]

\[ t_{l_r} = \bar{y}^* + b_2r (\bar{x} - \bar{x}^*). \tag{10} \]

• **Strategy IV**: $\bar{y}^*$, $\bar{x}$, $\bar{x}$ and $\bar{X}$ are used. The difference and regression type estimators are

\[ t_4 = \bar{y}^* + d_1 (\bar{x} - \tilde{x}^*) + d_2 (\bar{X} - \bar{x}). \tag{11} \]

The above mentioned estimators under strategy I, II and III were improved by Singh and Kumar (2008, 2010) by incorporating all possible auxiliary information that might be available at the disposal of a survey statistician and proposed $t_4$ under strategy IV. Singh and Kumar (2008, 2010) proved that their estimator minimized both overall variance component as well as the non-response variances component in (2), as evident by (13). It is important to note that $t_4$ is a generalization over $t_3$ and $t_2$, given in Equations (7) and (9) for $d_2 = 0$ and $d_1 = 0$, respectively.

The variances of the above estimators, up to the first order of approximation, are given by

\[
\Var(t_{l_r}) = \left[ \frac{(1-f)}{n} S_y^2 (1 - \rho^2) + \frac{W_2 (k-1)}{n} S_{y_2}^2 + \beta^2 S_{x_2}^2 - 2\beta S_{xy} \right] \tag{12}
\]

\[
\min.MSE(t_1) = \left\{ \frac{(1-f)}{n} S_y^2 + \frac{W_2 (k-1)}{n} S_{y_2}^2 \right\} - \left\{ \frac{(1-f)}{n} S_{xy} + \frac{W_2 (k-1)}{n} S_{xy_2} \right\}^2 \tag{13}
\]

\[
\Var(t_{l_r_2}) = \min.MSE(t_2) = \left[ \frac{(1-f)}{n} S_y^2 (1 - \rho^2) + \frac{W_2 (k-1)}{n} S_{y_2}^2 \right] \tag{14}
\]

\[
\Var(t_{l_r_3}) = \min.MSE(t_3) = \left[ \frac{(1-f)}{n} S_y^2 + \frac{W_2 (k-1)}{n} S_{y_2}^2 (1 - \rho^2) \right] \tag{15}
\]

\[
\text{min.MSE}(t_4) = \left[ \frac{(1-f)}{n} S_y^2 (1 - \rho^2) + \frac{W_2 (k-1)}{n} S_{y_2}^2 (1 - \rho^2) \right]. \tag{16}
\]

It is important to notice that MSE of any strategy under non-response can be split as $M = M_1 + M_2$, where $M_1$ is the over all sampling variance without non-response and $M_2$ is the contribution of sub-sampling due to non-response.

Also, if we observe the construction of $t_2$ which utilizes $\bar{y}^*$, $\bar{X}$ and zero function $\omega_1 = \bar{x} - \bar{X}$ (zero function is a function whose average value is zero) under optimum
conditions minimises its $M_1$ component but not $M_2$. Now, if we observe the construction of $t_3$ which uses $\tilde{y^*}$ and zero function $\omega_2 = \tilde{x} - \tilde{x^*}$ under optimum conditions minimises its $M_2$ component but not $M_1$. While, if we observe the construction of $t_4$ which uses both $\omega_1$ and $\omega_2$ under optimum conditions minimises both $M_1$ and $M_2$ components. This was the result due to chaining of $\omega_1$ and $\omega_2$ or chaining all the available information about $x$ with $\bar{x}$ as the chaining statistic.

Also, Khare and Srivastava (1993, 1995, 1997) and Okafor and Lee (2000) proposed a double sampling scheme for ratio and regression estimation with sub sampling the non-respondent while dealing with non-response problem. Due to economy of space, we have considered only the difference and regression type estimators similar to the estimators defined in (6) to (11). Some generalisation of these estimators was proposed by Singh and Bhushan (2012). In this paper we have also considered as follow the double sampling estimators under four strategies. The double sampling strategies are given below.

- **Strategy V**: $\tilde{y^*}$, $\tilde{x^*}$ and $\tilde{x'}$ are used. The difference and regression type estimators are

  \begin{align*}
  t_5 &= \tilde{y^*} + K_5 (\tilde{x'} - \tilde{x^*}) \\
  t_{l_5} &= \tilde{y^*} + b^* (\tilde{x'} - \tilde{x^*}).
  \end{align*}

- **Strategy VI**: $\tilde{y^*}$, $\tilde{x}$ and $\tilde{x'}$ are used. The difference and regression type estimators are

  \begin{align*}
  t_6 &= \tilde{y^*} + K_6 (\tilde{x'} - \tilde{x}) \\
  t_{l_6} &= \tilde{y^*} + b^{**} (\tilde{x'} - \tilde{x}).
  \end{align*}

Singh and Kumar (2010) proposed estimators for the population mean $\bar{Y}$ by using double sampling scheme under non-response.

- **Strategy VII**: $\tilde{y^*}$, $\tilde{x^*}$, $\tilde{x}$ and $\tilde{x'}$ are used. The difference (regression) type estimators is given by

  \begin{align*}
  t_7 &= \tilde{y^*} + d_3 (\tilde{x} - \tilde{x^*}) + d_4 (\tilde{x'} - \tilde{x}).
  \end{align*}

Again, it is important to note that $t_7$ is an extension of $t_4$ under double sampling. Also, $t_7$ is a generalisation over $t_6$ and $t_5$ for $d_3 = 0$ and $d_4 = 0$ respectively. Further $t_5$ minimises both $M_1$ and $M_2$ components due to double sampling and sub-sampling of non-respondents.

The MSEs of the above estimators, up to the first order of approximation, are given by

\begin{equation}
\text{Var}(t_{l_5}) = \left[ \left( \frac{1}{n'} - \frac{1}{N} \right) s_y^2 + \left( \frac{1}{n} - \frac{1}{n'} \right) s_y^2 (1 - \rho^2) + \frac{W_2 (k - 1)}{n} S^2 \right],
\end{equation}
where \( S_{\beta}^2 = S_2^2 + \beta S_x^2 (\beta - 2 \beta_2) \).

\[
\min \text{MSE}(t_5) = \left( \frac{1-f}{n} \right) S_y^2 + \frac{W_2(k-1)}{n} S_2^2 - \left\{ \left( \frac{1}{n} - \frac{1}{n'} \right) S_{xy} + \frac{W_2(k-1)}{n} S_{xy} \right\}^2
\]

\[
\min \text{MSE}(t_6) = \left( \frac{1}{n'} - \frac{1}{N} \right) S_y^2 + \left( \frac{1}{n} - \frac{1}{n'} \right) S_y^2 (1 - \rho^2) + \frac{W_2(k-1)}{n} S_2^2
\]

\[
\min \text{MSE}(t_7) = \left[ \left( \frac{1}{n'} - \frac{1}{N} \right) S_y^2 + \left( \frac{1}{n} - \frac{1}{n'} \right) S_y^2 (1 - \rho^2) + \frac{W_2(k-1)}{n} S_2^2 \right]
\]

where \( b^* = (s^y_x / s^2_x) \), \( b = (s_x / s_x^2) \), \( b_{2r} = (s_{xy(2r)} / s_{x(2r)}^2) \), \( b^{**} = (s^y_x / s^2_x) \),

\[ s^y_{xy} = \frac{1}{n-1} \left( \sum x_j y_j + r \sum x_j y_j - n \bar{x} \bar{y} \right), \]

\[ s_{xy}^2 = \frac{1}{n-1} \left( \sum x^2 + r \sum x^2 - n \bar{x} \bar{x} \right), \]

\[ s^2_x = \left( \sum (x_i - \bar{x})^2 / (n-1) \right), \]

\[ s_{xy(2r)} = \left( \sum (x_i - \bar{x}_{2r}) (y_i - \bar{y}_{2r}) / (r-1) \right) \]

\[ s_{x(2r)}^2 = \left( \sum (x_i - \bar{x}_{2r})^2 / (r-1) \right). \]

In this paper, we have proposed the number of improved regression (difference) estimators motivated by estimators suggested by Searls (1964), Cochran (1977), Rao (1986), Okafor and Lee (2000) and Singh and Kumar (2008) under single phase and two phase sampling classified under seven different strategies with deterministic non-response setup.

3. PROPOSED ESTIMATORS

In this section, we propose to use Searls (1964) type transformation (STT) under different strategies. The STT can be defined as

\[ T = \alpha \tilde{y}^*, \]

where \( \alpha \) is a suitable chosen scalar. The following estimators under the strategies described in the previous section are proposed using the STT.

The proposed estimator under Strategy I, when \( \tilde{y}^*, \tilde{x}^* \) and \( \tilde{X} \) are used, is given by

\[ T_1 = \alpha_1 \tilde{y}^* + \beta_1 (\tilde{X} - \tilde{x}^*). \]
The proposed estimator under Strategy II, when $\bar{y}^*$, $\bar{x}$ and $\bar{X}$ are used, is given by

$$T_2 = \alpha_2 \bar{y}^* + \beta_2 (\bar{X} - \bar{x}).$$

(28)

The proposed estimator under Strategy III, when $\bar{y}^*$, $\bar{x}$ and $\bar{x}^*$ are used, is given by

$$T_3 = \alpha_3 \bar{y}^* + \beta_3 (\bar{x}^* - \bar{x}).$$

(29)

Similarly, another proposed estimator obtained by chaining all the available auxiliary information. The proposed estimator under Strategy IV, when $\bar{y}^*$, $\bar{x}$, $\bar{x}^*$ and $\bar{X}$ are used, is given by

$$T_4 = \alpha_4 \bar{y}^* + \beta_4 (\bar{x}^* - \bar{x}) + \gamma_4 (\bar{X} - \bar{x}).$$

(30)

Also, under the two phase sampling scheme, we propose the following improved difference estimators classified under different strategies. The proposed estimator under Strategy V, when $\bar{y}^*$, $\bar{x}^*$ and $\bar{x}'$ are used, is given by

$$T_5 = \alpha_5 \bar{y}^* + \beta_5 (\bar{x}' - \bar{x}^*).$$

(31)

The proposed estimator under Strategy VI, when $\bar{y}^*$, $\bar{x}$ and $\bar{x}'$ are uses, is given by

$$T_6 = \alpha_6 \bar{y}^* + \beta_6 (\bar{x}' - \bar{x}).$$

(32)

The proposed estimator under Strategy VII, when $\bar{y}^*$, $\bar{x}$, $\bar{x}^*$ and $\bar{x}'$ are used, is given by

$$T_7 = \alpha_7 \bar{y}^* + \beta_7 (\bar{x} - \bar{x}^*) + \gamma_7 (\bar{x}' - \bar{x}),$$

(33)

where $\alpha_i$, $\beta_i$ and $\gamma_j$ are suitably chosen constant.

It is important to note that all these estimators $T_i$ are generalizations over $t_i$ and we get $t_i$ if we put $\alpha_i = 1$ in $T_i$ ($i = 1, 2, ..., 7$). In this study, we have focussed our attention over improvement in conventional difference and regression type estimators, which are the BLUE.

**Theorem 1.** The bias and minimum MSE of the STD estimator $T_i$ ($i = 1, 2, ..., 7$) is given by

$$\text{Bias}(T_i) = \bar{Y}(\alpha_i - 1).$$

(34)

and

$$\min_{\alpha_i} \text{MSE}_{\alpha_i}(T_i) = \frac{\hat{Y}^2 \text{MSE}(t_i)}{\hat{Y}^2 + \text{MSE}(t_i)}.$$  

(35)

where $\text{MSE}_{\alpha_i}(T_i)$ is the first order MSE with parameter $\alpha_i$, and $\text{MSE}(t_i)$ is the first order MSE of the $t_i$ or $T_i$ when $\alpha_i = 1$.

**Proof.** The outline of derivation is given in Appendix A along with the optimum values of different scalars.
THEOREM 2. Under the optimum values of scalars, the STD estimators \( T_i \) have always lesser MSE than the conventional estimators \( t_i \), \( i = 1, 2, ..., 7 \). Alternatively,

\[
\min \text{MSE}(T_i) < \min \text{MSE}(t_i) \quad (i = 1, 2, ..., 7).
\]

PROOF. Using (35), proof is obvious.

THEOREM 3. \( T_4 \) provides minimum MSE in comparison to single phase estimators \( T_1 \), \( T_2 \) and \( T_3 \) under optimal conditions.

PROOF. Since \( \min \text{MSE}(T_i) \) can be written as

\[
\min \text{MSE}(T_i) = \bar{Y}^2 (1 - \alpha_i), i = 1, 2, 3, 4
\]

On the basis optimal value of \( \alpha_i \), we observed that \( \alpha_4 \) has greater value in comparison to \( \alpha_1 \), \( \alpha_2 \) and \( \alpha_3 \). Therefore,

\[
\min \text{MSE}(T_4) < \min \text{MSE}(T_i), i = 1, 2, 3
\]

THEOREM 4. \( T_7 \) provides minimum MSE in comparison to two phase estimators \( T_5 \) and \( T_6 \) under optimal conditions.

PROOF. Since \( \min \text{MSE}(T_i) \) can be written as

\[
\min \text{MSE}(T_i) = \bar{Y}^2 (1 - \alpha_i), i = 5, 6, 7
\]

On the basis optimal value of \( \alpha_i \), we observe that \( \alpha_7 \) has greater value in comparison to \( \alpha_5 \) and \( \alpha_6 \). Therefore,

\[
\min \text{MSE}(T_7) < \min \text{MSE}(T_i), i = 5, 6
\]

Therefore, the estimators \( T_4 \) and \( T_7 \) provides the way to minimize the MSE and to improve the efficiency.

4. Empirical study

In order to have a better understanding about the efficiency of the proposed estimators, we have conducted an empirical study on the data by Srivastava (1993, p. 50) and compared the proposed estimators with \( \bar{y}_r \) and the results are reported.

A list of 70 villages in a Tehsil of India along with their population in 1981 and cultivated area (in acres) in the same year is taken in to consideration. Here the cultivated
**TABLE 1**  
*MSE and PRE of the existing estimators and proposed estimators for the first dataset.*

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$k = 2$</th>
<th>$k = 3$</th>
<th>$k = 4$</th>
<th>$k = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{y}$</td>
<td>10160.17(100)</td>
<td>10636.88(100)</td>
<td>11113.60(100)</td>
<td>11590.32(100)</td>
</tr>
<tr>
<td>$T$</td>
<td>10054.08(101.05)</td>
<td>10520.67(101.10)</td>
<td>10986.79(101.15)</td>
<td>11452.47(101.20)</td>
</tr>
</tbody>
</table>

**Strategy I**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$k = 2$</th>
<th>$k = 3$</th>
<th>$k = 4$</th>
<th>$k = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>4294.06(236.61)</td>
<td>4765.90(223.18)</td>
<td>5237.7533(212.18)</td>
<td>5709.60(202.10)</td>
</tr>
<tr>
<td>$T_1$</td>
<td>4288.77(236.90)</td>
<td>4745.94(224.12)</td>
<td>5195.20(213.92)</td>
<td>5637.74(205.58)</td>
</tr>
<tr>
<td>$t_4$</td>
<td>4269.75(237.95)</td>
<td>4722.66(225.23)</td>
<td>5167.32(215.07)</td>
<td>5604.92(206.78)</td>
</tr>
</tbody>
</table>

**Strategy II**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$k = 2$</th>
<th>$k = 3$</th>
<th>$k = 4$</th>
<th>$k = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_2$</td>
<td>4298.93(236.34)</td>
<td>4775.64(222.73)</td>
<td>5252.36(211.60)</td>
<td>5729.08(202.30)</td>
</tr>
<tr>
<td>$T_2$</td>
<td>4279.82(237.39)</td>
<td>4752.08(223.83)</td>
<td>5223.87(212.74)</td>
<td>5695.19(203.51)</td>
</tr>
</tbody>
</table>

**Strategy III**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$k = 2$</th>
<th>$k = 3$</th>
<th>$k = 4$</th>
<th>$k = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_3$</td>
<td>10065.76(100.93)</td>
<td>10448.08(101.80)</td>
<td>10830.39(102.61)</td>
<td>11212.71(103.36)</td>
</tr>
<tr>
<td>$T_3$</td>
<td>9961.63(101.99)</td>
<td>10335.93(102.91)</td>
<td>10709.93(103.77)</td>
<td>11083.65(104.57)</td>
</tr>
</tbody>
</table>

**Strategy IV**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$k = 2$</th>
<th>$k = 3$</th>
<th>$k = 4$</th>
<th>$k = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_4$</td>
<td>4204.52(241.65)</td>
<td>4586.8435(231.99)</td>
<td>4969.1591(223.65)</td>
<td>5351.4746(216.58)</td>
</tr>
<tr>
<td>$T_4$</td>
<td>4186.25(242.70)</td>
<td>4565.09(233.00)</td>
<td>4961.16(242.80)</td>
<td>5321.89(217.78)</td>
</tr>
</tbody>
</table>

**Strategy V**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$k = 2$</th>
<th>$k = 3$</th>
<th>$k = 4$</th>
<th>$k = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_5$</td>
<td>6736.24(150.82)</td>
<td>7208.08(147.57)</td>
<td>7679.93(144.71)</td>
<td>8151.78(142.18)</td>
</tr>
<tr>
<td>$t_6$</td>
<td>6727.54(151.02)</td>
<td>7176.38(148.22)</td>
<td>7614.40(145.95)</td>
<td>8044.04(144.08)</td>
</tr>
<tr>
<td>$T_5$</td>
<td>6680.87(152.07)</td>
<td>7123.29(149.32)</td>
<td>7554.66(147.11)</td>
<td>7977.40(145.29)</td>
</tr>
</tbody>
</table>

**Strategy VI**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$k = 2$</th>
<th>$k = 3$</th>
<th>$k = 4$</th>
<th>$k = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_6$</td>
<td>6741.11(150.72)</td>
<td>7217.83(147.37)</td>
<td>7694.54(144.43)</td>
<td>8171.26(141.84)</td>
</tr>
<tr>
<td>$T_6$</td>
<td>6694.24(151.77)</td>
<td>7164.13(148.47)</td>
<td>7633.55(145.58)</td>
<td>8102.50(143.04)</td>
</tr>
</tbody>
</table>

**Strategy VII**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$k = 2$</th>
<th>$k = 3$</th>
<th>$k = 4$</th>
<th>$k = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_7$</td>
<td>6646.71(152.86)</td>
<td>7029.02(151.32)</td>
<td>7411.34(149.95)</td>
<td>7793.65(148.71)</td>
</tr>
<tr>
<td>$T_7$</td>
<td>6601.14(153.91)</td>
<td>6978.08(152.43)</td>
<td>7354.73(151.10)</td>
<td>7731.08(149.91)</td>
</tr>
</tbody>
</table>
### TABLE 2
*MSE and PRE of the existing estimators and proposed estimators for the second dataset.*

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$k = 2$</th>
<th>$k = 3$</th>
<th>$k = 4$</th>
<th>$k = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{y}$</td>
<td>0.2067(100)</td>
<td>0.2464(100)</td>
<td>0.2860(100)</td>
<td>0.3256(100)</td>
</tr>
<tr>
<td>$T$</td>
<td>0.2066(100.054)</td>
<td>0.2462(100.064)</td>
<td>0.2857(100.075)</td>
<td>0.3253(100.085)</td>
</tr>
<tr>
<td><strong>Strategy I</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_1$</td>
<td>0.0664(311.050))</td>
<td>0.0853(288.828)</td>
<td>0.1040(274.383)</td>
<td>0.1227(265.221)</td>
</tr>
<tr>
<td>$t_{1,t_1}$</td>
<td>0.0665(310.625)</td>
<td>0.0856(287.794)</td>
<td>0.1046(273.273)</td>
<td>0.1237(263.223)</td>
</tr>
<tr>
<td>$T_1$</td>
<td>0.0664(311.104)</td>
<td>0.0852(288.893)</td>
<td>0.1040(274.913)</td>
<td>0.1227(265.307)</td>
</tr>
<tr>
<td><strong>Strategy II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{t_{1,t_2}} = t_2$</td>
<td>0.0871(237.288)</td>
<td>0.1267(194.376)</td>
<td>0.1663(171.902)</td>
<td>0.2060(158.072)</td>
</tr>
<tr>
<td>$T_2$</td>
<td>0.0871(237.342)</td>
<td>0.1267(194.441)</td>
<td>0.1663(171.977)</td>
<td>0.2058(158.158)</td>
</tr>
<tr>
<td><strong>Strategy III</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{t_{1,t_3}} = t_3$</td>
<td>0.1857(111.337)</td>
<td>0.2042(120.615))</td>
<td>0.2228(128.346)</td>
<td>0.2414(134.888))</td>
</tr>
<tr>
<td>$T_3$</td>
<td>0.1856(111.392)</td>
<td>0.2041(120.680)</td>
<td>0.2227(128.421)</td>
<td>0.2412(134.974)</td>
</tr>
<tr>
<td><strong>Strategy IV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{t_{1,t_4}}$</td>
<td>0.0660(312.896)</td>
<td>0.0846(291.079)</td>
<td>0.1032(277.110)</td>
<td>0.1217(267.401)</td>
</tr>
<tr>
<td>$T_4$</td>
<td>0.0660(312.950)</td>
<td>0.0846(291.144)</td>
<td>0.1031(277.186)</td>
<td>0.1217(267.486)</td>
</tr>
<tr>
<td><strong>Strategy V</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{t_{1,t_5}}$</td>
<td>0.0914(225.982)</td>
<td>0.1105(222.865)</td>
<td>0.1296(220.665)</td>
<td>0.1486(219.029)</td>
</tr>
<tr>
<td>$t_5$</td>
<td>0.0913(226.276)</td>
<td>0.1101(223.627)</td>
<td>0.1289(221.877)</td>
<td>0.1475(220.636)</td>
</tr>
<tr>
<td>$T_5$</td>
<td>0.0913(226.331)</td>
<td>0.1101(223.692)</td>
<td>0.1288(221.952)</td>
<td>0.1475(220.722)</td>
</tr>
<tr>
<td><strong>Strategy VI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{t_{1,t_6}} = t_6$</td>
<td>0.1120(184.514)</td>
<td>0.1516(162.438)</td>
<td>0.1913(149.506)</td>
<td>0.2309(141.012)</td>
</tr>
<tr>
<td>$T_6$</td>
<td>0.1120(184.568)</td>
<td>0.1516(162.503)</td>
<td>0.1912(149.581)</td>
<td>0.2307(141.098)</td>
</tr>
<tr>
<td><strong>Strategy VII</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_7$</td>
<td>0.0910(227.205)</td>
<td>0.1095(224.869)</td>
<td>0.1281(223.210)</td>
<td>0.1467(221.971)</td>
</tr>
<tr>
<td>$T_7$</td>
<td>0.0909(227.259)</td>
<td>0.1095(224.934)</td>
<td>0.1280(223.286)</td>
<td>0.1466(222.057)</td>
</tr>
</tbody>
</table>
area (in acres) is taken as main study character and the population of village is taken as auxiliary character. The parameters of the population are as follows.

\[ N = 70, \quad n' = 40, \quad n = 25, \quad \bar{Y} = 981.29, \quad \bar{X} = 1755.53, \quad S_y = 613.66, \quad S_x = 1406.13, \quad \bar{Y}_2 = 597.29, \quad \bar{X}_2 = 1100.24, \quad S_{y_2} = 244.11, \quad S_{x_2} = 631.51, \quad \rho = 0.778, \quad \rho_2 = 0.445, \quad R = 0.5589, \]
\[ \beta = 0.3395, \quad \beta_2 = 0.1720, \quad W_2 = 0.20. \]

We have conducted an empirical study on the data by Khare and Sinha (2004, p. 53). The data belongs to the data on physical growth of upper-socio-economic group of 95 school children of Varanasi under an ICMR study. The first 25 (i.e. 24 children) units have been considered as non-response units. The values of the parameters related to the study variate \( y \) (the weight in kg) and the auxiliary variate \( x \) (the chest circumference in cm) have been given below:

\[ N = 95, \quad n' = 70, \quad n = 35, \quad \bar{Y} = 19.497, \quad \bar{X} = 55.8611, \quad S_y = 3.0435, \quad S_x = 3.2735, \quad S_{y_2} = 2.3552, \quad S_{x_2} = 2.5137, \quad \rho = 0.8460, \quad \rho_2 = 0.7290, \quad R = 0.3490, \quad \beta = 0.7865, \quad \beta_2 = 0.6829, \]
\[ W_2 = 0.25, \quad N_2 = 24, \quad N_1 = 71. \]

The MSE and percent relative efficiency (PRE) of the estimators with respect to \( \bar{y}_r \) at different values of \( k \) are given in Tables 1 and 2.

5. **Simulation Study**

In this section, simulation is conducted to evaluate the performance of the proposed class of estimators with respect to traditional estimators. For this study we have generated a population size \( N = 1,000 \) from standard normal distribution using the MVRNORM package in software R, where study and auxiliary variable are correlated with correlation \( \rho = 0.7 \), draw a large sample of size \( n' = 400 \) from the population then again select a sample of size \( n = 200 \) from \( n' \) with 35% non-response. The whole simulation process starting from the drawing sample from variable \( Y \) and auxiliary variable \( X \) from normal population and calculating the estimates was repeated 50,000 times.

It can be easily seen that in simulation study the proposed estimators are better in comparison than the conventional estimators. A comparison of STD (regression) estimators \( T_i \) (\( i = 1, 2, \ldots, 7 \)) within themselves and with \( t_i \) (\( i = 1, 2, 3, 5, 6, 7 \)) shows that STD estimator \( T_4 \) is superior under single phase sampling and \( T_7 \) is superior under two phase sampling.
TABLE 3
Percentage relative efficiency (PRE) of the proposed estimators with respect to \( \bar{y}' \) using simulation.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{y}' )</td>
<td>100</td>
</tr>
<tr>
<td>( T )</td>
<td>101.20</td>
</tr>
<tr>
<td>Strategy I</td>
<td></td>
</tr>
<tr>
<td>( t_1 )</td>
<td>249.66</td>
</tr>
<tr>
<td>( t_{1r_1} )</td>
<td>249.64</td>
</tr>
<tr>
<td>( T_1 )</td>
<td>250.85</td>
</tr>
<tr>
<td>Strategy II</td>
<td></td>
</tr>
<tr>
<td>( t_2 = t_{1r_2} )</td>
<td>129.31</td>
</tr>
<tr>
<td>( T_2 )</td>
<td>130.51</td>
</tr>
<tr>
<td>Strategy III</td>
<td></td>
</tr>
<tr>
<td>( t_3 = t_{1r_3} )</td>
<td>159.45</td>
</tr>
<tr>
<td>( T_3 )</td>
<td>160.64</td>
</tr>
<tr>
<td>Strategy IV</td>
<td></td>
</tr>
<tr>
<td>( t_4 )</td>
<td>249.67</td>
</tr>
<tr>
<td>( T_4 )</td>
<td>250.87</td>
</tr>
<tr>
<td>Strategy V</td>
<td></td>
</tr>
<tr>
<td>( t_5 )</td>
<td>205.96</td>
</tr>
<tr>
<td>( t_{1r_5} )</td>
<td>205.94</td>
</tr>
<tr>
<td>( T_5 )</td>
<td>207.16</td>
</tr>
<tr>
<td>Strategy VI</td>
<td></td>
</tr>
<tr>
<td>( t_6 = t_{1r_6} )</td>
<td>116.50</td>
</tr>
<tr>
<td>( T_6 )</td>
<td>117.70</td>
</tr>
<tr>
<td>Strategy VII</td>
<td></td>
</tr>
<tr>
<td>( t_7 )</td>
<td>205.97</td>
</tr>
<tr>
<td>( T_7 )</td>
<td>207.16</td>
</tr>
</tbody>
</table>

6. Conclusions

Diana and Perri (2013) in their study opined that the regression and difference type estimator provide best possible improvement and cannot be improved upon. In this paper, we have proposed various improved difference type estimators, which provide an improvement over the corresponding regression estimators in their respective strategies. This result is significant as it provides an improvement over BLUE. The results are proved both theoretically as well as empirically. The results of conventional estimators can be obtained as a special case of the respective proposed estimators by setting \( \alpha_i = 1 \), \( i = 1, 2, ..., 7 \).

Acknowledgements

The authors would like to thank the Editor-in-Chief, and the Referees for their careful reading and constructive comments and suggestions which greatly improved the presentation of the paper.
APPENDIX

A. OUTLINE OF THE DERIVATION OF THEOREM 1

The MSE of $T_1$ is given by

$$\text{MSE}(T_1) = (\alpha_1 - 1)^2 \bar{Y}^2 + \alpha_1^2 \left( AS_y^2 + BS_{y_2}^2 \right) + \beta_1^2 \left( AS_x^2 + BS_{x_2}^2 \right) - 2\alpha_1\beta_1 \left( AS_{xy} + BS_{xy_2} \right),$$

for optimum value of $\alpha_1$ and $\beta_1$. Differentiating above equation partially with respect to $\alpha_1$ and $\beta_1$, we get

$$\alpha_1 = \frac{\bar{Y}^2}{\left\{ \bar{Y}^2 + AS_y^2 + BS_{y_2}^2 - \frac{(AS_{xy} + BS_{xy_2})^2}{AS_y^2 + BS_{y_2}^2} \right\}}$$

$$\beta_1 = \frac{\bar{Y}^2 (AS_{xy} + BS_{xy_2})}{(AS_x^2 + BS_{x_2}^2) \left\{ \bar{Y}^2 + AS_y^2 + BS_{y_2}^2 - \frac{(AS_{xy} + BS_{xy_2})^2}{AS_y^2 + BS_{y_2}^2} \right\}}.$$

After putting these values in $\text{MSE}(T_1)$, we get

$$\text{min.MSE}(T_1) = \frac{\bar{Y}^2 \left[ \frac{AS_y^2 + BS_{y_2}^2}{\bar{Y}^2 + AS_y^2 + BS_{y_2}^2} - \frac{(AS_{xy} + BS_{xy_2})^2}{AS_y^2 + BS_{y_2}^2} \right]}{\bar{Y}^2 + \left[ \frac{AS_y^2 + BS_{y_2}^2}{\bar{Y}^2 + AS_y^2 + BS_{y_2}^2} - \frac{(AS_{xy} + BS_{xy_2})^2}{AS_y^2 + BS_{y_2}^2} \right]} = \frac{\bar{Y}^2 \left[ \text{MSE}(T_1) \right]}{\bar{Y}^2 + \text{MSE}(T_1)},$$

where $A = \left( \frac{1}{n} - \frac{1}{N} \right)$ and $B = \frac{w_2(k-1)}{n}$.

Similarly, we get the min.MSE such that

$$\text{min.MSE}(T_i) = \frac{\bar{Y}^2 \left[ \text{MSE}(T_i) \right]}{\bar{Y}^2 + \text{MSE}(T_i)},$$

for $i = 1, 2, ..., 7$.

The optimum values of scalars for different estimators involved are

$$\alpha = \frac{\bar{Y}^2}{\bar{Y}^2 + AS_y^2 + BS_{y_2}^2}, \quad \alpha_2 = \frac{\bar{Y}^2}{\bar{Y}^2 + AS_y^2 \{1 - \rho^2\} + BS_{y_2}^2}, \quad \beta_2 = \frac{\bar{Y}^2 S_{xy}}{S_y^2 \{1 - \rho^2\} \bar{Y}^2 + AS_y^2 \{1 - \rho^2\} + BS_{y_2}^2}$$
\[
\alpha_3 = \frac{\hat{Y}^2}{\{Y^2 + AS^2_2 + BS^2_2(1-\rho^2)\}}, \quad \beta_3 = \frac{\hat{Y}^2 S_{xy}}{S^2_2 \{Y^2 + AS^2_2 + BS^2_2(1-\rho^2)\}}
\]
\[
\alpha_4 = \frac{\hat{Y}^2}{\{Y^2 + AS^2_2(1-\rho^2) + BS^2_2(1-\rho^2)\}}, \quad \beta_4 = \frac{\hat{Y}^2 S_{xy}}{S^2_2 \{Y^2 + AS^2_2(1-\rho^2) + BS^2_2(1-\rho^2)\}}
\]
\[
\gamma_4 = \frac{\hat{Y}^2 S_{xy}}{S^2_2 \{Y^2 + AS^2_2(1-\rho^2) + BS^2_2(1-\rho^2)\}}
\]
\[
\alpha_5 = \frac{\hat{Y}^2}{(C_{S_2}^2 + BS_{xy}^2) \{Y^2 + AS^2_2 + BS^2_2 - (C_{S_2}^2 + BS^2_2)\}}, \quad \beta_5 = \frac{\hat{Y}^2 S_{xy}}{(C_{S_2}^2 + BS_{xy}^2) \{Y^2 + AS^2_2 + BS^2_2 - (C_{S_2}^2 + BS^2_2)\}}
\]
\[
\alpha_6 = \frac{\hat{Y}^2}{\{Y^2 + DS^2_2 + CS^2_2(1-\rho^2) + BS^2_2\}}, \quad \beta_6 = \frac{\hat{Y}^2 S_{xy}}{S^2_2 \{Y^2 + DS^2_2 + CS^2_2(1-\rho^2) + BS^2_2\}}
\]
\[
\alpha_7 = \frac{\hat{Y}^2}{\{Y^2 + DS^2_2 + CS^2_2(1-\rho^2) + BS^2_2\}}, \quad \beta_7 = \frac{\hat{Y}^2 S_{xy}}{S^2_2 \{Y^2 + DS^2_2 + CS^2_2(1-\rho^2) + BS^2_2\}}
\]
\[
\gamma_7 = \frac{\hat{Y}^2 S_{xy}}{S^2_2 \{Y^2 + DS^2_2 + CS^2_2(1-\rho^2) + BS^2_2\}}
\]

where \(C = \left(\frac{1}{n} - \frac{1}{n'}\right)\), \(D = \left(\frac{1}{n'} - \frac{1}{N}\right)\).

REFERENCES


**Summary**

This paper introduces an efficient estimation procedure for the population mean in the presence of non-response. The proposed estimators of population mean provides an improvement over the corresponding conventional estimators proposed by Cochran (1977), Rao (1983, 1986) and Singh and Kumar (2008, 2010) under the deterministic non-response in terms of efficiency. A comparative study has been performed and it has been shown that the proposed estimators perform better in comparison to the conventional estimators. The theoretical findings are supported by an empirical study.

**Keywords**: Auxiliary information; Non-response; Mean square error.