A Study of Improved Chain Ratio-cum-Regression type Estimator for Population Mean in the Presence of Non-Response for Fixed Cost and Specified Precision

1B. B. Khare , 1Habib Ur Rehman and 2U. Srivastava

1Department Of Statistics, B.H.U, Varanasi-221005
2Statistics Section, MMV , B.H.U, Varanasi-221005

Abstract
In this paper, a study of improved chain ratio-cum regression type estimator for population mean in the presence of non-response for fixed cost and specified precision has been made. Theoretical results are supported by carrying out one numerical illustration.

Keywords Simple random sampling, non response, fixed cost, precision.

Introduction
In the field of socio, economics, researches and agricultures the problem arises due to non-response which friendly occur due to not at home, lack of interest, call back etc. In this expression a procedure of sub sampling from non respondents was suggested by Hansen and Hurwitz (1946). The use of auxiliary information in the estimators of the population parameters have helped in increased the efficiency of the proposed estimator. Using auxiliary character with known population mean of the estimators have been proposed by Rao (1986,90) and Khare and Srivastava (1996,1997). Further, Khare and Srivastava (1993,1995),Khare et al. (2008), Singh and Kumar (2010), Khare and Kumar (2009) and Khare and Srivastava(2010) have proposed different types of estimators for the estimation of population mean in the presence of non-response in case of unknown population mean of the auxiliary character.

In the present paper, we have studied an improved chain ratio-cum-regression type estimator for population mean in the presence of non-response have proposed by Khare and Rehman (2014) in the case of fixed cost and specified precision. In the present study we have obtained the optimum size of first phase sample \((n')\) and second phase sample \((n)\) is drawn from the population of size \(N\) by using SRSWOR method of sampling in case of fixed cost and also in case of specified precision \(V = V_0\). The expression for the minimum MSE of the estimator has been obtained for the optimum values of \(n'\) and \(n\) in case of fixed cost \(C \leq C_0\). The expression for minimum cost for the estimator has also been obtained in
case of specified precision $V = V_0$. An empirical study has been considered to observe the properties of the estimator in case of fixed cost and also in case of specified precision.

**The Estimators**

Let $\bar{Y}$, $\bar{X}$ and $\bar{Z}$ denote the population mean of study character $y$, auxiliary character $x$ and additional auxiliary character $z$ having $j$th value $Y_j$, $X_j$ and $Z_j$: $j = 1, 2, 3, ..., N$. Suppose the population of size $N$ is divided in $N_1$ responding units and $N_2$ not responding unit. According to Hansen and Hurwitz a sample of size $n$ is taken from population of size $N$ by using simple random sampling without replacement (SRSWOR) scheme of sampling and it has been observed that $n_1$ units respond and $n_2$ units do not respond. Again by making extra effort, a sub sample of size $r(= n_2 k^{-1})$ is drawn from $n_2$ non-responding unit and collect information on $r$ units for study character $y$. Hence the estimator for $\bar{Y}$ based on $n_1 + r$ units on study character $y$ is given by:

$$\bar{y}^* = \frac{n_1}{n} \bar{y}_1 + \frac{n_2}{n} \bar{y}_2$$

where $n_1$ and $n_2$ are the responding and non-responding units in a sample of size $n$ selected from population of size $N$ by SRSWOR method of sampling. $\bar{y}_1$ and $\bar{y}_2$ are the means based on $n_1$ and $r$ units selected from $n_2$ non-responding units by SRSWOR methods of sampling. Similarly we can also define estimator for population mean $\bar{X}$ of auxiliary character $x$ based on $n_1$ and $r$ unit respectively, which is given as:

$$\bar{x}^* = \frac{n_1}{n} \bar{x}_1 + \frac{n_2}{n} \bar{x}_2$$

Variance of the estimators $\bar{y}^*$ and $\bar{x}^*$ are given by

$$V(\bar{y}^*) = \frac{f}{n} S_y^2 + \frac{W_2 (k-1)}{n} S_{y(2)}^2$$

and

$$V(\bar{x}^*) = \frac{f}{n} S_x^2 + \frac{W_2 (k-1)}{n} S_{x(2)}^2$$

where $f = 1 - \frac{n}{N}$, $W_2 = \frac{N_2}{N}$, $(S_y^2, S_{y(2)}^2)$ and $(S_x^2, S_{x(2)}^2)$ are population mean squares of $y$ and $x$ for entire population and non-responding part of population.

In case when the population means of the auxiliary character is unknown, we select a larger first phase sample of size $n'$ units from a population of size $N$ units by using simple random sample without replacement (SRSWOR) method of sampling and estimate $\bar{X}$ by $\bar{x}'$ based on these units $n'$. Further second phase sample of size $n$ (i.e. $n < n'$) is drawn from $n'$ units by using SRSWOR method of sampling and variable $y$ under investigation is measured $n_1$ responding and $n_2$ non-responding units. Again a sub sample of size $r(n_2/k, k > 1)$ is drawn from $n_2$ non-responding units and collect information on $r$ units by personal interview.
In this case two phase sampling ratio, product and regression estimators for population mean $\bar{Y}$ using one auxiliary character in the presence of non-response have been proposed by Khare and Srivastava (1993,1995) which are given as follows:

\[
T_1 = \bar{Y}^* \frac{\bar{X}'}{\bar{X}''} \tag{5}
\]

\[
T_2 = \bar{Y}^* + b^* (\bar{X}' - \bar{X}'') \tag{6}
\]

where $\bar{X}' = \frac{n_1}{n} \bar{x}_1 + \frac{n_2}{n} \bar{x}_2$ , $\bar{X}'' = \frac{1}{n} \sum_{j=1}^{n_1} x_j$ , $\bar{x} = \frac{1}{n'} \sum_{j=1}^{n'} x_j$ , $b^* = \frac{\hat{S}_{xy}}{\hat{S}_x^2}$ , $s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$\hat{S}_{yx}$ and $\hat{S}_x^2$ are estimates of $S_{yx}$ and $S_x^2$ based on $n_1 + r$ units.

The conventional and alternative two phase sampling ratio type estimators suggested by Khare and Srivastava (2010) which are as follows:

\[
T_3 = \bar{Y}^* \left( \frac{\bar{X}'}{\bar{X}''} \right)^\alpha \tag{7}
\]

\[
T_4 = \bar{Y}^* \left( \frac{\bar{X}'}{\bar{X}''} \right)^{\alpha'}
\]

where $\alpha$ and $\alpha'$ are constants.

Singh and Kumar (2010) have proposed difference type estimator using auxiliary character in the presence of non-response which is given as follows:

\[
T_5 = \bar{Y}^* \left( \frac{\bar{X}'}{\bar{X}''} \right)^{\alpha_1} \left( \frac{\bar{X}'}{\bar{X}''} \right)^{\alpha_2} \tag{8}
\]

where $\alpha_1$ and $\alpha_2$ are constants.

In case when $\bar{X}'$ is not known than we may use an additional auxiliary character $z$ with known population mean $\bar{Z}$ with the assumption that the variable $z$ is less correlated to $y$ than $x$ i.e, ( $\rho_{yz} < \rho_{yx}$) , $x$ and $z$ are variables such that $z$ is more cheaper than $x$.

Following Chand (1975), some estimators have been proposed by Kiregyera (1980,84), Srivasatava et al. (1990) and Khare & Kumar (2011). In the case of non-response on the study character, the chain regression type and generalized chain type estimators for the population mean in the presence of non-response have been proposed by Khare & Kumar (2010) and Khare et al. (2011). An improved chain ratio-cum-regression type estimator for population mean in the presence of non-response have been proposed by Khare & Rehman (2014), which is given as follows:

\[
T_6 = \bar{Y}^* \left( \frac{\bar{X}'}{\bar{X}''} \right)^p \left( \frac{\bar{Z}}{\bar{Z}'} \right)^q + b_{xy} (\bar{x}' - \bar{x}) + b_{xz} (\bar{z}' - \bar{z}') \tag{9}
\]

where $p$ and $q$ are constants. $b_{xy}$ and $b_{xz}$ are regression coefficients. $\bar{Z}$ and $\bar{z}'$ population mean and sample mean based on first phase sample of size $n'$ units selected from population of size $N$ by SRSWOR method.
Mean Square Errors of the Study Estimator

Using the large sample approximations, the expressions for the mean square errors of the estimator proposed by Khare & Rehman (2014) up to the terms of order \((n^{-1})\) are given by

\[
MSE(T_k) = V(\hat{y}^*) + \left(1 - \frac{1}{n'}\right)\left[\bar{y}^2 + \frac{y^2}{n'} - \frac{1}{n'^2}\right]
\]

\[
+ \left(1 - \frac{1}{n'}\right)\left[\bar{y}^2 - \bar{x}^2 + \frac{x^2}{n'} - \frac{1}{n'^2}\right]
\]

\[
+ \frac{W_2(K-1)}{n}\left[\bar{y}^2 - \bar{x}^2 + \frac{x^2}{n'} - \frac{1}{n'^2}\right]
\]

The optimum values of \(p\) and \(q\) and the values of regression coefficient are given as follows:

\[
p_{opt} = \frac{\left(1 - \frac{1}{n'}\right)\bar{y}^2 - \bar{x}_y^2}{\bar{y}^2 - \bar{x}_y^2}
\]

\[
q_{opt} = \frac{\bar{y}^2 - \bar{x}_y^2}{\bar{y}^2 - \bar{x}_y^2}
\]

\[
b_{xy} = \frac{\bar{x}^2 - \bar{x}_y^2}{\bar{x}^2 - \bar{x}_y^2}
\]

Mean square errors of the estimators \(T_1, T_2, T_3, T_4\) and \(T_5\) are given as follows:

\[
MSE(T_{1\min}) = V(\hat{y}^*) + \bar{y}^2\left(1 - \frac{1}{n'}\right)\left[\bar{y}^2 - \bar{x}_y^2 + \frac{W_2(k-1)}{n}\left(\bar{y}^2 - \bar{x}_y^2 + \bar{x}^2 - \bar{x}_x^2\right)\right]
\]

\[
MSE(T_{2\min}) = V(\hat{y}^*) - \bar{y}^2\left(1 - \frac{1}{n'}\right)\left[\bar{y}^2 + \frac{W_2(k-1)}{n}\left(\bar{y}^2 - \bar{x}_y^2 + \bar{x}^2 - \bar{x}_x^2\right)\right]
\]

\[
MSE(T_{3\min}) = V(\hat{y}^*) - \bar{y}^2\left(1 - \frac{1}{n'}\right)\left[\bar{y}^2 - \bar{x}_y^2 + \frac{W_2(k-1)}{n}\left(\bar{y}^2 - \bar{x}_y^2 + \bar{x}^2 - \bar{x}_x^2\right)\right]
\]

\[
MSE(T_{4\min}) = V(\hat{y}^*) - \bar{y}^2\left(1 - \frac{1}{n'}\right)\left[\bar{y}^2 - \bar{x}_y^2 + \frac{W_2(k-1)}{n}\left(\bar{y}^2 - \bar{x}_y^2 + \bar{x}^2 - \bar{x}_x^2\right)\right]
\]

and

\[
MSE(T_{5\min}) = \bar{y}^2\left(1 - \frac{1}{n'}\right)\left[\bar{y}^2 - \bar{x}_y^2 + \frac{W_2(k-1)}{n}C_y^2\left(1 - \frac{1}{n'}\right)\left[\bar{y}^2 - \bar{x}_y^2 + \bar{x}^2 - \bar{x}_x^2\right]\right]
\]

where \(V(\hat{y}^*) = \bar{y}^2\left(1 - \frac{1}{n'}\right)\left[\bar{y}^2 - \bar{x}_y^2 + \frac{W_2(k-1)}{n}C_y^2\right]\) and \(B = \frac{\bar{y}^2}{\bar{x}_x^2}\)
Determination of $n', n$ and $k$ for the Fixed Cost $C \leq C_0$

Let us assume that $C_0$ be the total cost (fixed) of the survey apart from overhead cost. The expected total cost of the survey apart from overhead cost is given as follows:

$$C = (e'_1 + e'_2)n' + n \left( e_1 + e_wW_1 + e_3 \frac{W_2}{k} \right),$$  \hspace{1cm} (19)

where

- $e'_1$ : the cost per unit of obtaining information on auxiliary character $x$ at the first phase.
- $e'_2$ : the cost per unit of obtaining information on additional auxiliary character $z$ at the first phase.
- $e_1$ : the cost per unit of mailing questionnaire/visiting the unit at the second phase.
- $e_2$ : the cost per unit of collecting, processing data obtained from $n_1$ responding units.
- $e_3$ : the cost per unit of obtaining and processing data (after extra efforts) for the sub sampling units.

The expression for, $MSE(T_6)$ can be expressed in terms of $D_0, D_1, D_2$ and $D_3$ which are the coefficients $\frac{1}{n}, \frac{1}{n'}, \frac{1}{n}$ and $\frac{1}{N}$ respectively. The expression of $MSE(T_6)$ is given as follows:

$$MSE(T_6)_{\text{min}} = \frac{D_0}{n} + \frac{D_1}{n'} + \frac{k D_2}{n} - \frac{D_3}{N},$$  \hspace{1cm} (20)

For obtaining the optimum values of $n', n, k$ for the fixed cost $C \leq C_0$, we define a function $\phi$ which is given as:

$$\phi = MSE(T_6)_{\text{min}} + \lambda (C - C_0),$$  \hspace{1cm} (21)

where $\lambda$ is the Lagrange’s multiplier.

We differentiating $\phi$ with respect to $n', n, k$ and equating zero, we get optimum values of $n', n$ and $k$ which are given as follows:

$$n'_{\text{opt}} = \frac{D_1}{\lambda(e'_1 + e'_2)},$$  \hspace{1cm} (22)

$$n_{\text{opt}} = \frac{(D_0 + k_{\text{opt}}D_2)}{\lambda \left( e_1 + e_wW_1 + e_3 \frac{W_2}{k_{\text{opt}}} \right)},$$  \hspace{1cm} (23)

and

$$k_{\text{opt}} = \frac{D_0 e_3 W_2}{D_2 \left( e_1 + e_wW_1 \right)},$$  \hspace{1cm} (24)

where

$$\sqrt{\lambda} = \frac{1}{C_0} \left[ \sqrt{D_1 (e'_1 + e'_2)} + \sqrt{(D_0 + k_{\text{opt}}D_2) \left( e_1 + e_wW_1 + e_3 \frac{W_2}{k_{\text{opt}}} \right)} \right],$$  \hspace{1cm} (25)
The minimum value of $MSE(T_6)$ for the optimum values of $n', n$ and $k$ in the expression $MSE(T_6)$, we get:

$$MSE(T_6)_{\text{min}} = \frac{1}{C_0} \left[ \sqrt{D_1(e'_1 + e'_2)} + \sqrt{(D_0 + k_{opt}D_2) \left( e_1 + e_2W_1 + e_3 \frac{W_2}{k_{opt}} \right)} \right]^2 - \frac{D_2}{N}, \quad (26)$$

Now neglecting the term of $O \left( N^{-1} \right)$, we have

$$MSE(T_6)_{\text{min}} = \frac{1}{C_0} \left[ \sqrt{D_1(e'_1 + e'_2)} + \sqrt{(D_0 + k_{opt}D_2) \left( e_1 + e_2W_1 + e_3 \frac{W_2}{k_{opt}} \right)} \right]^2 \quad (27)$$

**Determination of $n', n$ and $k$ for the Specified Precision $V = V_0'$**

Let $V_0'$ be the specified variance of the estimator $T_6$ which is fixed in advance, so we have

$$V_0' = \frac{D_0}{n} + \frac{D_1}{n'} + \frac{kD_2}{n} - \frac{D_3}{N}, \quad (28)$$

To find the optimum values of $n', n, k$ and minimum expected total cost, we define a function $\psi$ which is give as follows:

$$\psi = (e'_1 + e'_2)n' + n \left( e_1 + e_2W_1 + e_3 \frac{W_2}{k_{opt}} \right) + \mu(MSE(T_6)_{\text{min}} - V_0')$$

where $\mu$ is the Lagrange’s multiplier.

After differentiating $\psi$ with respect to $n', n, k$ and equating to zero, we find the optimum value of $n', n$ and $k$ which are given as;

$$n'_{opt} = \frac{\mu D_1}{(e'_1 + e'_2)} \quad (30)$$

$$n_{opt} = \frac{\mu(D_0 + k_{opt}D_2)}{\left( e_1 + e_2W_1 + e_3 \frac{W_2}{k_{opt}} \right)} \quad (31)$$

and

$$k_{opt} = \frac{D_3W_2e_3}{D_2(e_1 + e_2W_1)} \quad (32)$$

where

$$\sqrt{\mu} = \left[ \sqrt{D_1(e'_1 + e'_2)} + \sqrt{(D_0 + k_{opt}D_2) \left( e_1 + e_2W_1 + e_3 \frac{W_2}{k_{opt}} \right)} \right]^2 - \frac{D_2}{N}, \quad (33)$$

The minimum expected total cost incurred on the use of $T_6$ for the specified variance $V_0'$ will be given as follows:
Sampling Strategies for Finite Population Using Auxiliary Information

We have

\[
C_{6\text{min}} = \sqrt{D_1(e'_1 + e'_2)} + \sqrt{(D_0 + k_{opt}D_2)} \left( e_1 + e_2W_1 + e_3 \frac{W_2}{k_{opt}} \right),
\]

\[V_0 + \frac{D_3}{N},\]

Now neglecting the terms of \(O(N^{-1})\), we have

\[
C_{6\text{min}} = \sqrt{D_1(e'_1 + e'_2)} + \sqrt{(D_0 + k_{opt}D_2)} \left( e_1 + e_2W_1 + e_3 \frac{W_2}{k_{opt}} \right),
\]

\[V_0',\]

An Empirical Study

To illustrate the results we use the data considered by Khare and Sinha (2007). The description of the population is given below:

The data on physical growth of upper socio-economic group of 95 schoolchildren of Varanasi under an ICMR study, Department of Pediatrics, B.H.U., during 1983-84 has been taken under study. The first 25% (i.e., 24 children) units have been considered as non-responding units. Here we have taken the study variable \(y\), auxiliary variable \(x\) and the additional auxiliary variable \(z\) are taken as follows:

- \(y\): weight (in kg.) of the children.
- \(x\): skull circumference (in cm) of the children.
- \(z\): chest circumference (in cm) of the children.

The values of the parameters of the \(y, x\) and \(z\) characters for the given data are given as follows:

\[\bar{Y} = 19.4968, \quad \bar{Z} = 51.1726, \quad \bar{X} = 55.8611, \quad C_y = 0.15613, \quad C_z = 0.03006, \quad C_x = 0.05860, \quad C_y(2) = 0.12075, \quad C_z(2) = 0.02478, \quad C_x(2) = 0.05402, \quad \rho_{yz} = 0.328, \quad \rho_{yx} = 0.846, \quad \rho_{xz} = 0.297, \quad \rho_{yz(2)} = 0.570, \quad W_2 = 0.25, \quad W_1 = 0.74, \quad N = 95, \quad n = 35\]

Table 1. Relative efficiency (in %) of the estimators with respect to \(\bar{y}^*\) (for the fixed cost \(C \leq C_0 = Rs. 220, \quad c_1 = Rs. 0.90, \quad c_2 = Rs. 0.10, \quad c_1 = Rs. 2, \quad c_2 = Rs. 4, \quad c_3 = Rs. 25\)).

<table>
<thead>
<tr>
<th>Estimators</th>
<th>(k_{opt})</th>
<th>(n'_{opt})</th>
<th>(n_{opt})</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\bar{y}^*)</td>
<td>2.68</td>
<td>---</td>
<td>30</td>
<td>100 (0.3843)*</td>
</tr>
<tr>
<td>(\bar{T}_1)</td>
<td>2.89</td>
<td>58</td>
<td>23</td>
<td>117 (0.3272)</td>
</tr>
<tr>
<td>(\bar{T}_2)</td>
<td>2.03</td>
<td>74</td>
<td>19</td>
<td>131 (0.2941)</td>
</tr>
</tbody>
</table>
From table 1, we obtained that for the fixed cost \( C \leq C_0 \) the study estimator \( T_6 \) is more efficient in comparison to the estimators \( \bar{y}^*, T_1, T_2, T_3, T_4 \) and \( T_5 \).

**Table 2.** Expected cost of the estimators for the specified variance \( V_0'=0.2356 \) : (\( c_1'=Rs.\ 0.90 \), \( c_2'=Rs.\ 0.10 \), \( c_1=Rs.\ 2 \), \( c_2=Rs.\ 5 \), \( c_3=Rs.\ 25 \))

<table>
<thead>
<tr>
<th>Estimators</th>
<th>( k_{opt} )</th>
<th>( n'_{opt} )</th>
<th>( n_{opt} )</th>
<th>Expected Cost (in Rs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{y}^* )</td>
<td>2.68</td>
<td>---</td>
<td>61</td>
<td>502</td>
</tr>
<tr>
<td>( T_1 )</td>
<td>2.89</td>
<td>107</td>
<td>40</td>
<td>418</td>
</tr>
<tr>
<td>( T_2 )</td>
<td>2.03</td>
<td>115</td>
<td>25</td>
<td>332</td>
</tr>
<tr>
<td>( T_3 )</td>
<td>2.61</td>
<td>88</td>
<td>20</td>
<td>246</td>
</tr>
<tr>
<td>( T_4 )</td>
<td>1.06</td>
<td>92</td>
<td>16</td>
<td>275</td>
</tr>
<tr>
<td>( T_5 )</td>
<td>2.68</td>
<td>87</td>
<td>21</td>
<td>244</td>
</tr>
<tr>
<td>( T_6 )</td>
<td>2.67</td>
<td>69</td>
<td>20</td>
<td>231</td>
</tr>
</tbody>
</table>

From table 2, we obtained that for the specified variance the study estimator \( T_6 \) has less cost in comparison to the cost incurred in the estimators \( \bar{y}^*, T_1, T_2, T_3, T_4 \) and \( T_5 \).
Conclusion

The information on additional auxiliary character and optimum values of increase the efficiency of the study estimators in comparison to corresponding estimators in case of the fixed cost $C \leq C_0$ and specified precision $V = V_0$.

References


