IMPROVED CLASS OF ESTIMATORS OF FINITE POPULATION MEAN USING SAMPLING FRACTION AND INFORMATION ON TWO AUXILIARY VARIABLES IN SAMPLE SURVEYS

H. P. Singh,
A. Rathour,
R. S. Solanki∗

School of Studies in Statistics, Vikram University, Ujjain-456010, M.P., India.

1. INTRODUCTION

Auxiliary information has been widely used in sample surveys to increase the efficiency of the estimators. For this reason, many authors used the auxiliary information at the estimation stage. When the correlation between study variable and auxiliary variable is positive (high) ratio method of estimation due to (Cochran, 1940) is used. On the other hand if the correlation is negative (high) product method of estimation due to (Robson, 1957) and revisited by (Murthy, 1964) is preferred.

In addition several authors such as (Srivastava, 1980); (Srivenukataramana and Tracy, 1980); (Prasad, 1989); (Bahl and Tuteja, 1991); (Mohanty and Sahoo, 1995); (Singh and Gangele, 1998); (Upadhyaya and Singh, 1999); (Singh and Ruiz-Espejo, 2003); (Singh et al., 2004, 2009, 2010); (Singh and Tailor, 2005); (Singh et al., 2007); (Singh and Karpe, 2009); (Tailor and Sharma, 2009); (Upadhyaya et al., 2011) (Singh and Solanki, 2012) and (Solanki et al., 2012) suggested some modified estimators of population mean using information on single auxiliary. (Singh, 1967); suggested a ratio-cum-product estimator of population mean using information on two auxiliary variables. In this paper we have made the use of information on two auxiliary variables and the sampling fraction in defining the estimators of population mean and study their properties under large sampling approximation. An empirical is carried out to judge the merits of the suggested estimators over usual unbiased estimator and (Singh, 1967); ratio-cum-product estimator.

Consider a finite population \( U = (U_1, U_2, ..., U_N) \) containing \( N \) units. Let \( Y, X_1 \) and \( X_2 \) are denote the study variable and the auxiliary variables taking values \( Y_j, X_{1j} \) and \( X_{2j} \) respectively on the unit \( U_j \) (\( j = 1, 2, ..., N \)) of population \( U \), where \( X_1 \) is positively and \( X_2 \) is negatively correlated with the study variable \( Y \). Let a sample of size \( n \) is drawn by simple random sampling without replacement (SRSWOR) from population \( U \) to estimate the population mean \( \bar{Y} \) of the study variable \( Y \). It is very well known that sample mean \( \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i \) of the study variable \( Y \) is an unbiased estimator of population mean \( \bar{Y} \) and under SRSWOR the variance of \( \bar{Y} \) is given by

* Corresponding author: ramkssolanki@gmail.com
\[ \text{Var}(\bar{Y}) = \text{MSE}(\bar{Y}) = \theta S^2_y = \theta \bar{Y}^2 C^2_y \]  

(1)

where \( f = (n / N) \), \( \theta = \{(1 - f) / n\} \), \( C^2_y = (S^2_y / \bar{Y}^2) \), \( S^2_y = (N - 1)^{-1} \sum_{j=1}^{N} (Y_j - \bar{Y})^2 \) and \( \text{MSE}(\bullet) \) stands for the mean square error of \( \bullet \).

Assuming the knowledge of the population mean \( \bar{X}_1 \) and \( \bar{X}_2 \) of the auxiliary variables \( X_1 \) and \( X_2 \) respectively, (Singh, 1967) proposed a ratio-cum-product estimator for population mean \( \bar{Y} \) as

\[ t_{RP} = \bar{Y} \left( \frac{\bar{X}_1}{\bar{X}_1} \right) \left( \frac{\bar{X}_2}{\bar{X}_2} \right) \]  

(2)

where \( \bar{X}_1 = n^{-1} \sum_{i=1}^{n} X_{1i} \) and \( \bar{X}_2 = n^{-1} \sum_{i=1}^{n} X_{2i} \) are sample mean of auxiliary variables \( \bar{X}_1 \) and \( \bar{X}_2 \) respectively.

To the first degree of approximation, the bias and \( \text{MSE} \) of \( t_{RP} \) are respectively given by

\[ B(t_{RP}) = \theta \bar{Y} \left[ C^2_{x1} (1 - K_{y1}) + C^2_{x2} (K_{y2} - K_{x1x2}) \right] \]  

(3)

\[ \text{MSE}(t_{RP}) = \theta \bar{Y}^2 \left[ C^2_{x1} + C^2_{x2} (1 - 2K_{y2}) + C^2_{x2} \{1 + 2(K_{y2} - K_{x1x2})\} \right] \]  

(4)

where \( C^2_{xib} = (S^2_{xb} / \bar{X}_b^2) \), \( S^2_{xb} = (N - 1)^{-1} \sum_{j=1}^{N} (X_{bi} - \bar{X}_b)^2 \), \( K_{yib} = \rho_{yib} (C_{jib} / C_{xb}) \), \( \rho_{yib} = (S_{yib} / S_{xb} S_{yib}) \), \( S_{yib} = (N - 1)^{-1} \sum_{j=1}^{N} (X_{bi} - \bar{X}_b)(Y_j - \bar{Y}) \) (\( b = 1, 2 \)); \( K_{x1x2} = \rho_{x1x2} (C_{x1} / C_{x2}) \), \( \rho_{x1x2} = (S_{x1x2} / S_{x1} S_{x2}) \) and \( S_{x1x2} = (N - 1)^{-1} \sum_{j=1}^{N} (X_{1j} - \bar{X}_1)(X_{2j} - \bar{X}_2) \).

The remaining part of the paper is as follows. In Section 2 we have suggested a generalized class of estimators of population mean and studied their properties. Some members of suggested class have been generated. Section 3 made some theoretical comparisons of suggested class with other existing estimators of population mean. An empirical study is carried out in Section 4. We have end the paper with final conclusion.

2. SUGGESTED ESTIMATORS

We have proposed the following generalized class of estimators for estimating the population mean \( \bar{Y} \) as a linear combination between \( \bar{Y} \) and \( t_{RP} \).
Improved class of estimators of finite population mean ...

\[
\begin{align*}
\tau &= [\bar{y} + (1 - w)\bar{y}_{K, h}] = [\bar{y} + (1 - w)\bar{y}(\frac{\bar{x}_1}{\bar{x}_1} + \frac{\bar{x}_2}{\bar{x}_2})]
\end{align*}
\]  
(5)

where  \( w = (a + bf)(c + df)^{-1} \) is a function of \( f \) and \((a, b, c, d)\) are suitably chosen scalars for designing the different estimators which are shown in Table 1.

To obtain the bias and MSE of proposed class of estimators \( \tau \) we write

\[
\bar{y} = \bar{Y}(1 + e_0), \quad \bar{x}_1 = \bar{X}_1(1 + e_1) \quad \text{and} \quad \bar{x}_2 = \bar{X}_2(1 + e_2),
\]

such that \( E(e_0) = E(e_1) = E(e_2) = 0 \)

and to the first degree of approximation

\[
\begin{align*}
E(e_0^2) &= \theta C_j^2, \quad E(e_1^2) = \theta C_{x_1}^2, \quad E(e_2^2) = \theta C_{x_2}^2, \quad E(e_0e_1) = \theta K_{jx_1}C_{x_1}^2, \\
E(e_0e_2) &= \theta K_{jx_2}C_{x_2}^2 \quad \text{and} \quad E(e_1e_2) = \theta K_{x_1x_2}C_{x_1}^2
\end{align*}
\]

where

\[
\begin{align*}
K_{jx_1} &= \rho_{jx_1}(C_j / C_{x_1}), \quad K_{jx_2} = \rho_{jx_2}(C_j / C_{x_2}), \quad \rho_{jx_1} = (S_{jx_1} / S_j S_{x_1}), \\
\rho_{jx_2} &= (S_{jx_2} / S_j S_{x_2}), \quad S_{jx_1} = (N - 1)^{-1}\sum_{j=1}^{N}(Y_j - \bar{Y})(X_{1j} - \bar{X}_1) \quad \text{and} \quad \\
S_{jx_2} &= (N - 1)^{-1}\sum_{j=1}^{N}(Y_j - \bar{Y})(X_{2j} - \bar{X}_2).
\end{align*}
\]
TABLE 1
Some members of generalized class of estimators $t$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Values of constants $w$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a$</td>
</tr>
<tr>
<td>$t_0 = \bar{Y}$</td>
<td>0</td>
</tr>
<tr>
<td>$t_1 = f_{10}$</td>
<td>0</td>
</tr>
<tr>
<td>$t_2 = f \bar{Y} + (1 - f) \frac{1}{\tilde{X_1}} \frac{\hat{Y} - \frac{1}{\tilde{X_1}}}{\bar{X}_2}$</td>
<td>0</td>
</tr>
<tr>
<td>$t_3 = 2f \bar{Y} + (1 - 2f) \frac{1}{\tilde{X_1}} \frac{\hat{Y} - \frac{1}{\tilde{X_1}}}{\bar{X}_2}$</td>
<td>1</td>
</tr>
<tr>
<td>$t_4 = (1 - f) \bar{Y} + f \frac{1}{\tilde{X_1}} \frac{\hat{Y} - \frac{1}{\tilde{X_1}}}{\bar{X}_2}$</td>
<td>1</td>
</tr>
<tr>
<td>$t_5 = (1 - 2f) \bar{Y} + 2f \frac{1}{\tilde{X_1}} \frac{\hat{Y} - \frac{1}{\tilde{X_1}}}{\bar{X}_2}$</td>
<td>1</td>
</tr>
<tr>
<td>$t_6 = \frac{1}{(1 + f)} \bar{Y} + \frac{f}{(1 + f)} \frac{1}{\tilde{X_1}} \frac{\hat{Y} - \frac{1}{\tilde{X_1}}}{\bar{X}_2}$</td>
<td>1</td>
</tr>
<tr>
<td>$t_7 = \frac{(1 - f)}{(1 + f)} \bar{Y} + \frac{2f}{(1 + f)} \frac{1}{\tilde{X_1}} \frac{\hat{Y} - \frac{1}{\tilde{X_1}}}{\bar{X}_2}$</td>
<td>1</td>
</tr>
<tr>
<td>$t_8 = \frac{2f}{(1 + f)} \bar{Y} + \frac{(1 - f)}{(1 + f)} \frac{1}{\tilde{X_1}} \frac{\hat{Y} - \frac{1}{\tilde{X_1}}}{\bar{X}_2}$</td>
<td>0</td>
</tr>
<tr>
<td>$t_9 = \frac{(1 - 2f)}{(1 + f)} \bar{Y} + \frac{3f}{(1 + f)} \frac{1}{\tilde{X_1}} \frac{\hat{Y} - \frac{1}{\tilde{X_1}}}{\bar{X}_2}$</td>
<td>1</td>
</tr>
<tr>
<td>$t_{10} = \frac{(1 - 2f)}{(1 + 2f)} \bar{Y} + \frac{4f}{(1 + 2f)} \frac{1}{\tilde{X_1}} \frac{\hat{Y} - \frac{1}{\tilde{X_1}}}{\bar{X}_2}$</td>
<td>1</td>
</tr>
</tbody>
</table>

Expressing (5) in terms of $e$’s we have

$$t = \bar{Y} \{w(1 + e_0) + (1 - w)(1 + e_0)(1 + e_1)^{-1}(1 + e_2)\}$$

(6)

We assume $|e_1| < 1$ so that $(1 + e_1)^{-1}$ is expendable, thus expanding, multiplying out and neglecting terms of $e$’s having power greater than two we have

$$t \cong \bar{Y} \{1 + e_0 - (1 - w)(e_1 - e_2 + e_0 e_1 - e_0 e_2 - e_1^2 + e_1 e_2)\}$$

or

$$(t - \bar{Y}) \cong \bar{Y} \{e_0 - (1 - w)(e_1 - e_2 + e_0 e_1 - e_0 e_2 - e_1^2 + e_1 e_2)\}$$

(7)

Taking expectation on both side of (7), we get the bias of $t$ to the first degree of approximation as
Improved class of estimators of finite population mean ...

\[ B(t) = (1 - w) \sum \theta [C_{x1}^2(1 - K_{y1}) + C_{x2}^2(K_{y2} - K_{y1})]. \]  \hspace{1cm} (8)

Squaring both side of (7) and neglecting terms of \( \varepsilon \)'s having power greater than two we have

\[ (t - \bar{Y})^2 \cong \sum \varepsilon^2 \left[ (1 - w)^2 \right] \hspace{1cm} (9) \]

Taking expectation on both side of (9), we get the \( MSE \) of the proposed class of estimators \( t \) to the first degree of approximation as

\[ MSE(t) = \theta \sum \varepsilon^2 \left[ C_y^2 + A + w^2B - 2wC \right], \]  \hspace{1cm} (10)

where

\[ A = [C_{x1}^2(1 - 2K_{y1}) + C_{x2}^2(1 + 2K_{y2} - K_{y1})] = [B - 2(C_{x1}^2 K_{y1} - C_{x2}^2 K_{y2})], \]

\[ B = [C_{x1}^2 + C_{x2}^2(1 - 2K_{y1})], \]

\[ C = [C_{x1}^2(1 - K_{y1}) + C_{x2}^2(1 + K_{y2} - 2K_{y1})] = [B - (C_{x1}^2 K_{y1} - C_{x2}^2 K_{y2})], \]

being \( K_{y2} < 0 < K_{y1} \), then \( (C_{x1}^2 K_{y1} - C_{x2}^2 K_{y2}) > 0 \) and \( A < C < B \).

3. EFFICIENCY COMPARISON

From (1) and (10) we have

\[ MSE(t) < MSE(\bar{Y}) \text{ if } A - 2wC + w^2B < 0 \]  \hspace{1cm} (11)

or

\[ \frac{2C}{B} - 1 < w < 1. \]  \hspace{1cm} (12)

From (4) and (10) we have

\[ MSE(t) < MSE(t_{RP}) \text{ if } \min\left(0, \frac{2C}{B}\right) < w < \max\left(0, \frac{2C}{B}\right). \]  \hspace{1cm} (13)

in order to judge the biasedness of proposed class of estimators \( t \) we have defined the quantity as
\[ Q = \left| \frac{B(t)}{\theta \bar{Y}} \right| = \left| (1-w)[C_{x_1}(1-K_{y_1}) + C_{x_2}(K_{y_2} - K_{x_1x_2})] \right|. \] (14)

The quantity \( q \) for different estimators which are members of suggested class of estimators \( t \) would be easily obtained through (4) by putting relevant value of \( w \) as shown in Table 1.

4. \textbf{EMPIRICAL STUDY}

To observe the relative performance of different estimators of \( \bar{Y} \), we have considered the natural population data set given in Steel and Torrie (1960, p. 282). The population description is given below:

\( Y = \log \) of leaf burn in sec, \( X_1 = \) Potassium percentage, \( X_2 = \) Chlorine percentage.

\( N = 30, \bar{Y} = 0.6860, C_{y_1} = 0.4803, \rho_{y_1x_1} = 0.1797, n = 6, \bar{X}_1 = 4.6537, C_{x_1} = 0.2295, \rho_{x_1x_2} = -0.4996, \bar{X}_2 = 0.8077, C_{x_2} = 0.7493, \rho_{x_1x_2} = 0.4074. \)

We have computed the percent relative efficiency (PRE) with respect to \( \bar{Y} \) and quantity \( Q \) of different estimators of population mean \( \bar{Y} \) and summarized in Table 2.

\begin{table}[h]
\centering
\begin{tabular}{ccc}
\hline
Estimator & PRE & Quantity \( Q \) \\
\hline
\( t_0 = \bar{Y} \) & 100.00 & 0.0000 \\
\( t_1 = t_{RP} \) & 75.52 & 0.2170 \\
\( t_2 \) & 107.46 & 0.1736 \\
\( t_3 \) & 142.58 & 0.1302 \\
\( t_4 \) & 135.85 & 0.0434 \\
\( t_5 \) & 157.10 & 0.0868 \\
\( t_6 \) & 130.10 & 0.0362 \\
\( t_7 \) & 153.50 & 0.0723 \\
\( t_8 \) & 131.66 & 0.1447 \\
\( t_9 \) & 154.22 & 0.1085 \\
\( t_{10} \) & 146.62 & 0.1240 \\
\hline
\end{tabular}
\caption{PRE and quantity \( Q \) of different estimators of \( \bar{Y} \)}
\end{table}

It is observed from Table 2 that the estimators \( t_i \), (\( i = 2, 3, ..., 10 \)) which are members of suggested class of estimators \( t \) are biased but marginal in magnitude. The estimator \( t_6 \) is least biased estimator followed by \( t_4 \) and \( t_7 \) among estimators \( t_i \), (\( i = 1, 2, ..., 10 \)). Further it is observed from Table 2 that the estimators \( t_i \), (\( i = 2, 3, ..., 10 \)) are more efficient than the usual
unbiased estimator $\overline{Y}$ and Singh (1967) ratio-cum-product estimator $t_{RP}$ in respect of $PRE$. The estimator $t_5$ is best in the sense of having largest $PRE$ [i.e. $PRE(t_5) = 157.10$] among all the estimators $t_i, (i = 0, 1, ..., 10)$ followed by $t_9$ and $t_7$.

5. CONCLUSION

The present article suggests a generalized class of estimators of population mean using information on two auxiliary variables and sampling fraction. The members of suggested class have been generated by choosing suitable values of scalars involved in the class. The usual unbiased estimator and (Singh, 1967) ratio-cum-product estimator are identified as members of suggested class. The bias and $MSE$ expressions of the suggested class have been obtained to the first degree of approximation. The properties of the members of the suggested class have been obtained through the properties of the suggested class. Thus the suggested class unifies several results at one place.

The conditions are obtained under which the proposed class of estimators is more efficient than the usual unbiased estimator and the (Singh, 1967) ratio-cum-product estimator. The theoretical findings have been examined with an empirical study. It has been observed that the members of suggested class are more efficient than the usual unbiased estimator and (Singh, 1967) ratio-cum-product estimator. Thus the proposal of the suggested class of estimators are justified.

ACKNOWLEDGEMENTS

Authors are thankful to the learned referee for his valuable suggestions regarding improvement of the paper.

REFERENCES


SUMMARY

Improved class of estimators of finite population mean using sampling fraction and information on two auxiliary variables in sample surveys

This paper suggested a generalized class of estimators using information on two auxiliary variables and sampling fraction in simple random sampling. The bias and mean squared error formulae of suggested class have been derived under large sample approximation and compared with usual unbiased estimator and Singh’s (1967) ratio-cum-product estimator. The theoretical findings have been satisfied with an empirical study.