Generalized and transformed two phase sampling Ratio and Product Type Estimators for Population Mean Using Auxiliary Character in the presence of Unit nonresponse on study and auxiliary Character

B. B. Khare, Utkarsh and Supriya Khare
Department of Statistics, Institute of Science, BHU, Varanasi-221005, India.
Email- bbkhare56@yahoo.com
Email* - utkarshparmar11@gmail.com
Email- suprikh@gmail.com

Abstract

In this paper, we have proposed generalized and transformed two phase sampling ratio and product type estimators for population mean using auxiliary character in the presence of unit nonresponse on the study and auxiliary character. The properties of the proposed estimators have been studied. A comparative study has been made with the relevant estimators and an empirical study has been given in the support of the problem.

Keywords: Mean square error, population mean, auxiliary character, unit non-response on study and auxiliary character.

1. Introduction

The use of an auxiliary character to increase the efficiency of the estimator for population parameters is widely used in researches related to the field of socioeconomic, agriculture and biomedical sciences. The research work on the estimation of population parameters using auxiliary characters by several authors have been reviewed by Tripathi et al. (1994), Khare (2003) and Khare et.al (2013). For an example, in a forest surveys, the average amount of timber of a tree can be estimated by using the diameter of the tree as an auxiliary character. In this context, several estimators for population mean have been proposed by Kiregyera (1984), Srivastava et al.(1989), Sahoo et al.(1993), Kadilar and Cingi (2005) and Khare and Kumar (2010) Further the research work in the presence of nonresponse on study character has been reviewed by Khare et al. (2014).

While conducting a sample survey, it may not be possible to collect information on all the units selected in the sample due to nonresponse. In this case, Hansen and Hurwitz (1946) have suggested a method of sub-sampling from non-respondents and the estimator for population mean based on available information on responding units and sub sample units drawn from the non-responding units in the sample has been proposed. Further several problems on estimating the population mean in the presence of nonresponse on the study character through sample surveys in the presence of non-
response have been considered by Rao (1986, 90), Khare and Srivastava (1993, 95, 97, 2000), Singh et al. (2008, 09(a, b), 10) and Khare and Kumar (2011).

In this paper, we have proposed generalized and transformed two phase sampling ratio and product type estimators for population mean using auxiliary character in the presence of unit nonresponse on the study and auxiliary character. The properties of the proposed estimators have been studied. A comparative study has been made with the relevant estimators and an empirical study has been given in the support of the problem.

2. Proposed Estimators:

Let $y$ and $x$ denote the study character and auxiliary character having $j$th values $Y_j$ and $X_j$ ($j = 1, 2, ..., N$) with their population means $\bar{Y}$ and $\bar{X}$. The population of size $N$ is supposed to be divided into $N_1$ responding and $N_2$ non-responding units such that $N_1 + N_2 = N$. According to the Hansen Hurwitz (1946), a sample of size $n$ is drawn from the population of size $N$ by using simple random sampling without replacement (SRSWOR) scheme, it has been observed that only $n_1$ units are responding and $n_2$ units are not responding in the sample of size $n$ for the study character $y$. Further, by making some extra effort a sub-sample of size $r (r = n_2/k, k > 1)$ is drawn from $n_2$ non-responding units by using SRSWOR sampling scheme and related information on available units are collected by personal interview for study character $y$.

In case when there is no nonresponse on the auxiliary character and the population mean of the auxiliary character is not known, we draw the first phase sample of size $n' (< N)$ from the population of size $N$ by using simple random sampling without replacement (SRSWOR) scheme and estimate the population mean $\bar{X}$ by the first phase sample mean $\bar{x}'$ based on $n'$ units.

Further, a second phase sample of size $n' (< n')$ is drawn from first phase sample of size $n'$ by using SRSWOR sampling scheme and observed that only $n_1$ units are responding and $n_2$ units are not responding for the study character $y$. Again we draw the sub-sample of size $r (r = n_2/k, k > 1)$ from $n_2$ non-responding units by using SRSWOR sampling scheme and collect the related information on $r$ units using personal interview method by making extra effort.
Using Hansen and Hurwitz (1946) technique of sub sampling from nonrespondents, the estimator for $\bar{y}$ based on $n_1 + r$ observations on the study character $y$ is given as follows:

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

(2.1)

where $\bar{y}_1$ and $\bar{y}_2'$ are the means of character $y$ based on $n_1$ and $r$ units. The variance of the estimator is given by

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

(2.2)

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

Similarly, the sample mean of values of $x$ based on $n_1 + r$ values on $x$ character is given as follows:

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

(2.3)

where $\bar{x}_1$ and $\bar{x}_2'$ denote the means of auxiliary character $(x)$ based on $n_1$ and $r$ units.

The variance of the estimator $\bar{x}^*$ is given by

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

(2.4)

where $S_x^2$ and $S_x'^2$ are the population mean squares of study character $x$ for the entire population and for the non-responding part of the population.

In case when there is no nonresponse on the auxiliary character several researchers have proposed two phase sampling estimator for population mean of study character. But if there is also unit nonresponse on the auxiliary character along with the study character, in this case sampling design and the estimation procedure will be different. To estimate $\bar{X}$, we first draw a sample of size $n'$ for population of size $N$ by using SRSWOR method of sampling and observed that $n_1'$ units respond and $n_2'$ units
do not respond. Then from $n_2$ units we draw a subsample of size $r'(r'=n_2/k_0, k_0 > 1)$ and observe information on x character.

Now we estimate $\bar{X}$ by $\bar{x}^{**}$ by using $n_1'$ and $r'$ units on x character using Hansen Hurwitz (1946) technique. The estimator $\bar{x}^{**}$ for $\bar{X}$ is given as follows

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

(2.5)

where $\bar{x}_1'$ and $\bar{x}_2'$ are the sample means of x based on $n_1'$ and $r'$ units. Further we draw a sample of size n from $\hat{n}$ and observe $n_1$ units respond and $n_2$ units do not respond. Again we draw a subsample of size $r(r=n_2/k_1, k_1 > 1)$ from $n_2$ nonresponding units and the estimate for $\bar{Y}$ and $\bar{X}$ based on $n_1 + r'$ units.

The conventional and alternative ratio, product and regression type estimators for population mean under this scheme are given by Khare and Srivastava (1993, 1995) which are given as follows:

$$t_1 = \frac{\bar{Y}^*}{\bar{X}} \bar{x}' \quad \text{and} \quad t_2 = \frac{\bar{Y}^*}{\bar{X}'}$$

(2.6)

Further Khare and Kumar (2009) have proposed transformed two phase sampling ratio and product type estimators for population mean in presence of nonresponse, which are given as follows:

$$t_3 = \frac{\bar{Y}^* (\bar{x}' + D_1)}{(\bar{x} + D_1)} \quad \text{and} \quad t_4 = \frac{\bar{Y}^* (\bar{x} + D_1')}{(\bar{x} + D_1')}$$

(2.7)

In the present context, we have proposed generalized and transformed two phase sampling ratio and product type estimators for population mean using auxiliary character in the presence of unit nonresponse on the study and auxiliary character, which are given as follows:

$$T = \bar{y}^* \left( \frac{\bar{x}}{\bar{x}^{**}} \right)^{\alpha}, \quad \bar{T}_R = \bar{y}^* \left( \frac{\bar{x}^{**} + A}{\bar{x}^* + A} \right), \quad \bar{T}_P = \bar{y}^* \left( \frac{\bar{x}^{**} + A_1}{\bar{x}^* + A_1} \right)$$

(2.8)

3. **Bias and mean square error of the proposed estimator:**

In order to derive the expression for Bias and Mean square error and bias of the proposed estimator $T, T_R$ and $T_P$ we have:
Let $\bar{Y}^* = \bar{Y}(1 + \varepsilon_0)$, $\bar{x}^* = \bar{x}(1 + \varepsilon_1)$, $\bar{x}^{**} = \bar{x}(1 + \varepsilon_2)$
such that $E(\varepsilon_1) = 0$ and $|\varepsilon_l| < 1 \ \forall \ l = 0,1,2$.

\[
\begin{align*}
E(\varepsilon_0^2) &= \left( \frac{1 - \frac{1}{N}}{n} \right) c_y^2 + \frac{W_2(k-1)}{n} c_{y^2} \\
E(\varepsilon_1^2) &= \left( \frac{1 - \frac{1}{N}}{n} \right) c_x^2 + \frac{W_2(k-1)}{n} c_{x^2} \\
E(\varepsilon_0 \varepsilon_1) &= \left( \frac{1 - \frac{1}{N}}{n} \right) c_{yx} \\
E(\varepsilon_1 \varepsilon_2) &= \left( \frac{1 - \frac{1}{N}}{n} \right) c_{x^2} \\
E(\varepsilon_0 \varepsilon_2) &= \left( \frac{1 - \frac{1}{N}}{n} \right) c_{y^2}
\end{align*}
\]

(3.1)

\[
T = \bar{Y}(1 + \varepsilon_0 + \alpha \varepsilon_1 - \varepsilon_2 + \alpha \varepsilon_0 \varepsilon_1 - \varepsilon_0 \varepsilon_2 - \alpha^2 \varepsilon_1 \varepsilon_2 \\
+ \frac{\alpha (\alpha - 1)}{2} \varepsilon_1^2 + \frac{\alpha (\alpha + 1)}{2} \varepsilon_2^2) \\
= \bar{Y}(1 + \varepsilon_0 + \alpha (\varepsilon_1 - \varepsilon_2) + \alpha (\varepsilon_0 \varepsilon_1 - \varepsilon_0 \varepsilon_2) - \alpha^2 \varepsilon_1 \varepsilon_2 \\
+ \frac{\alpha (\alpha - 1)}{2} \varepsilon_1^2 + \frac{\alpha (\alpha + 1)}{2} \varepsilon_2^2) \\
\]

(3.2)

\[
T_R = \bar{Y}(1 + \varepsilon_0) \left\{ \frac{\bar{X}(1 + \varepsilon_2 + A)}{\bar{X}(1 + \varepsilon_1 + A)} \right\} \\
= \bar{Y}(1 + \varepsilon_0) \left[ 1 + \left( \frac{\bar{X}}{\bar{X} + A} \right) \varepsilon_2 \right] \left[ 1 + \left( \frac{\bar{X}}{\bar{X} + A} \right) \varepsilon_1 \right]^{-1} \\
= \bar{Y} \left[ 1 + \varepsilon_0 + \left( \frac{\bar{X}}{\bar{X} + A} \right) (\varepsilon_2 - \varepsilon_1) + \varepsilon_0 \varepsilon_2 \right] \left( \frac{\bar{X}}{\bar{X} + A} \right)^2 (\varepsilon_1^2 - \varepsilon_1 \varepsilon_2) \\
\]

(3.3)

where $\Lambda > 0$ & $\left| \frac{\bar{X}}{\bar{X} + A} \right| \varepsilon_1 < 0$

\[
T_P = \bar{Y}(1 + \varepsilon_0) \left\{ \frac{\bar{X}(1 + \varepsilon_1 + A_1)}{\bar{X}(1 + \varepsilon_2 + A_1)} \right\} \\
= \bar{Y}(1 + \varepsilon_0) \left[ 1 + \left( \frac{\bar{X}}{\bar{X} + A_1} \right) \varepsilon_1 \right] \left[ 1 + \left( \frac{\bar{X}}{\bar{X} + A_1} \right) \varepsilon_2 \right]^{-1}
\]
where \( A_1 > 0 \) and \( \left| \frac{X}{X + A_1} \right| \varepsilon_2 \) < 0

Now, the Bias and MSE of the estimators \( T, T_R \) and \( T_P \) are given as follows:

\[
\text{Bias (T)} = E[\hat{Y} \{ \varepsilon_0 + \alpha (\varepsilon_1 - \varepsilon_2) + \alpha (\varepsilon_0 \varepsilon_1 - \varepsilon_0 \varepsilon_2) - \alpha^2 \varepsilon_1 \varepsilon_2 \\
+ \frac{\alpha (\alpha - 1)}{2} \varepsilon_1^2 + \frac{\alpha (\alpha + 1)}{2} \varepsilon_2^2 \}] \\
= \hat{Y} \left\{ \alpha \left( \frac{1}{n} - \frac{1}{n'} \right) C_{yx} + \left( \frac{W_2 (k - 1)}{n} \right) C_{yx}^* \right\} - \alpha^2 \left\{ \left( \frac{1}{n} - \frac{1}{N} \right) C_x^2 \right\} \\
+ \frac{\alpha (\alpha - 1)}{2} \left\{ \frac{1}{n} - \frac{1}{N} \right\} C_x^2 + \left( \frac{W_2 (k - 1)}{n} \right) C_{x}^* \\
+ \alpha (\alpha + 1) \left\{ \frac{1}{n} - \frac{1}{N} \right\} C_x^2 + \left( \frac{W_2 (k - 1)}{n} \right) C_{x}^* \} \\
(3.5)
\]

\[
\text{Bias (T_R)} = \left( \frac{\hat{Y} X}{X + A} \right) E \left\{ \left( \varepsilon_0 \varepsilon_2 - \varepsilon_0 \varepsilon_1 \right) + \left( \frac{X}{X + A} \right) \left( \varepsilon_1^2 - \varepsilon_1 \varepsilon_2 \right) \right\} \\
= \left( \frac{\hat{Y} X}{X + A} \right) \left\{ \left( \frac{X}{X + A} \right) \left( \frac{1}{n} - \frac{1}{n'} \right) C_x^2 + \left( \frac{W_2 (k - 1)}{n} \right) C_{x}^* \right\} \\
- \left\{ \frac{1}{n} - \frac{1}{n'} \right\} C_{yx} + \frac{W_2 (k - 1)}{n} C_{yx}^* \} \\
(3.6)
\]

\[
\text{Bias (T_P)} = \left( \frac{\hat{Y} X}{X + A_1} \right) E \left\{ \left( \varepsilon_0 \varepsilon_1 - \varepsilon_0 \varepsilon_2 \right) + \left( \frac{X}{X + A_1} \right) \left( \varepsilon_2^2 - \varepsilon_1 \varepsilon_2 \right) \right\} \\
= \left( \frac{\hat{Y} X}{X + A_1} \right) \left\{ \left( \frac{1}{n} - \frac{1}{n'} \right) C_{yx} + \left( \frac{W_2 (k - 1)}{n} \right) C_{xy}^* \right\} \\
+ \left( \frac{X}{X + A_1} \right) \left( \frac{W_2 (k - 1)}{n} C_x^2 \right) \} \\
(3.7)
\]
Now differentiating (3.8) with respect to $\alpha$ and equating to zero, we have
After putting the value of $\alpha_{\text{opt.}}$ from (3.11) in (3.8) the minimum value $\text{MSE}(T)$ is given as follows:

$$\text{MSE}(T)_{\text{min}} = F^2 \left( \left( \frac{1}{n} \right) C_{yx}^2 + \frac{W_2(k-1)}{n} C_{yx}^* \right)^2 \left( \frac{1}{n} \right) C_{yx}^* + \frac{W_2(k-1)}{n} C_{yx}^* - 1$$

(3.12)

Again differentiating (3.9) with respect to $D (= \bar{X} / (\bar{X} + A))$ and equating to zero, we obtain the optimum value of $A$ which is given by

$$A_{\text{opt.}} = \bar{X} \left[ \left( \frac{1}{n} \right) C_{yx}^* + \frac{W_2(k-1)}{n} C_{yx}^* \right] - 1$$

(3.13)

After putting the value of $A_{\text{opt.}}$ from (3.13) in (3.9) the minimum value of $\text{MSE}(T_R)$ is given as follows:

$$\text{MSE}(T_R)_{\text{min}} = F^2 \left( \left( \frac{1}{n} \right) C_{yx}^2 + \frac{W_2(k-1)}{n} C_{yx}^* \right)^2 \left( \frac{1}{n} \right) C_{yx}^* + \frac{W_2(k-1)}{n} C_{yx}^* - 1$$

(3.14)

Similarly differentiating (3.10) w.r.t. $D_1 (= \bar{X} / (\bar{X} + A_1))$ and equating to zero, we obtain the optimum value of $A_1$ which is given as follows:
After putting the value of $A_{1\text{opt.}}$ from (3.15) in (3.10), the minimum value of MSE $(T_P)$ is given by:

$$MSE(T_P)_{\text{min}} = \bar{Y}^2 \left( \left( \frac{1}{n} - \frac{1}{n'} \right) C_{x_x}^2 + \frac{W_2(k-1)C_{x_x}^*}{n} \right) - \left( \left( \frac{1}{n} - \frac{1}{n'} \right) C_{y_x}^2 + \frac{W_2(k-1)C_{y_x}^*}{n} \right)$$

(3.16)

To study the precision of the estimator $T, T_R$ and $T_P$, an empirical study is considered.

4. An Empirical Study:

109 Village/Town/ward wise population of urban area under police-station-Baria, Tahasil-Champua, Orissa has been taken under consideration from District Census Hanbook, 1981, Orissa, published by Govt. of India, (Sinha (2001)). The last 25% villages (i.e. 27 villages) have been considered as non-response group of the population.

The values of the parameters of the population under study are as follows:

$\bar{Y} = 145.3028$, $\bar{X} = 259.083$, $C_y = 0.7667$, $C_x = 0.764$, $C_{x_x}^* = 0.6899$, $C_{y_x}^* = 0.5429$, $\rho_{yx} = 0.905$, $\rho_{yx}^* = 0.875$.

where average number of literate persons in the village ($y$) by using average number of non-workers in the village ($x$) as auxiliary character.
Table 1: Relative efficiency (in %) of the estimators with respect to $\bar{y}^*$ for the different values of $k$, $(N = 109, n' = 50, n = 30)$.

<table>
<thead>
<tr>
<th>Estimators</th>
<th>$1/k$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1/3</td>
<td>1/2</td>
</tr>
<tr>
<td>$\bar{y}^*$</td>
<td>100 (548.6727)</td>
<td>100 (465.6998)</td>
<td>100 (382.7269)</td>
</tr>
<tr>
<td>$t_3$</td>
<td>161.9357 (338.8213)</td>
<td>165.8717 (280.7590)</td>
<td>171.9459 (222.5856)</td>
</tr>
<tr>
<td>$t_4$</td>
<td>161.9357 (338.8213)</td>
<td>165.8717 (280.7590)</td>
<td>171.9459 (222.5856)</td>
</tr>
<tr>
<td>$T$</td>
<td>183.7947 (298.5247)</td>
<td>182.9773 (254.5124)</td>
<td>182.2467 (210.0048)</td>
</tr>
<tr>
<td>$T_R$</td>
<td>183.7947 (298.5247)</td>
<td>182.9773 (254.5124)</td>
<td>182.2467 (210.0048)</td>
</tr>
<tr>
<td>$T_P$</td>
<td>183.7947 (298.5247)</td>
<td>182.9773 (254.5124)</td>
<td>182.2467 (210.0048)</td>
</tr>
</tbody>
</table>

Figures in parentheses give the MSE (.)

From the Table, we observed that the MSE of the proposed estimators $T$, $T_R$ and $T_P$ are equal. It is also observed that the value of MSE (.) of the proposed and relevant estimators decreases as the value of $k$ decreases.

5. Conclusion:

In the case of unit nonresponse on the study and auxiliary character, generalized two phase and transformed two phase ratio and product type estimators $T$, $T_R$ and $T_P$ are found to be more efficient than the relevant estimators $t_3$, $t_4$ and $\bar{y}^*$. So it is suggested to use $T$, $T_R$ and $T_P$ in case of two phase sampling with unit nonresponse on the study and auxiliary character.

6. References:


