HCTL Open International Journal of Technology Innovations and Research (IJTIR)
http://ijtir.hctl.org
Volume 14, April 2015
e-ISSN: 2321-1814, ISBN (Print): 978-1-62951-946-3

Improved Ratio in Regression type Estimator for Population Mean using known Coefficient of Variation of the Study Character in the Presence of Non-Response

B. B. Khare, Habib Ur Rehman

hrmbd007@gmail.com

Abstract

Using the known coefficient of variation of the study character, improved ratio in regression type estimator for population mean in the presence of non-response has been proposed and their properties have been studied. The conditions under which the proposed estimator is more efficient than the relevant estimators have been obtained. The empirical studies have been given in the support of the problems in the case of positive and negative correlation between the study and the auxiliary characters which show the increase in the efficiency of the proposed estimator using known coefficient of variation of the study character with respect to the relevant estimators.

Keywords: Population mean, Study variable, Coefficient of variation, Non- response.
***JEL Classification**:* C83

Introduction

The use of auxiliary information in sample surveys during the stage of planning, designing, selection of units and devising the estimation procedure has been considered mainly in the field of Agricultural, Biological, Medical and Social Sciences. The reviews of the research works related to these areas have been carried out by Tripathi et al. (1994) and Khare (2003).

Das and Tripathi (1980) have proposed improved estimators for population mean using known coefficient of variation of auxiliary character. The use of coefficient of variation of the study character in proposing the estimator for population mean of the study character has been made by Searls (1964,67) and Sen (1978). In the case of non response in the selected sample from the population, Hansen and Hurwitz (1946) have proposed the method of sub sampling from non respondents in the sample. Using Hansen and Hurwitz (1946) techniques, Khare and Srivastava (1993,95) have proposed two phase sampling estimators for population mean in the presence of non response. Khare and Kumar (2009) have proposed the estimators utilizing the coefficient of variation of the study character in the estimation of population mean using auxiliary character in the presence of non response.

In the present paper, we have proposed improved ratio in regression type estimator for the population mean using known coefficient of variation of the study character in the presence of non response. The properties of their proposed estimator have been studied and comparative studies of the estimators have been made with the relevant estimators. The empirical studies have been given in the support of the proposed estimators used in the case of positive correlation as well as negative correlation between study and auxiliary characters.

The Estimators

Let (y, x) be the study character and auxiliary character observed in the population $U{:}\,(U_1, U_2, \dots, U_N)$ of the size N having population means $(\bar{Y}, \bar{X})$ and coefficient of variations(C_y, C_x). $(\bar{y}, \bar{x})$ denote the sample means of (y, x) based on a sample of size n drawn from the population using SRSWOR method of sampling. Searls (1964) proposed an estimator $\bar{y}^{**} = a\bar{y}^*$, where $a = (1 + \frac{f}{n} C_y^2)^{-1}$ and $f = \frac{N-n}{N}$. In the case of non-response in the selected sample of size n for the study character y, we observe that n_1 units respond and n_2 units do not respond. Further, a sub sample of size $r = \frac{n_2}{K} (K > 1)$ is drawn from n_2 non-responding units by making extra efforts. Hansen and Hurwitz (1946) proposed an estimator for $\bar{Y}$ which is given as follows:

$$\bar{y}^* = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\bar{y}_2' \tag{1}$$

$$MSE(\bar{y}^*) = \frac{f}{n} S_y^2 + \frac{W_2(K-1)}{n} S_{y(2)}^2, \tag{2}$$

where $\bar{y}_1$ and $\bar{y}_2$ denote the sample means based on n_1 responding and n_2 non-responding units in the sample of size n such that $\bar{y} = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\bar{y}_2'$ and $\bar{y}_2'$ denotes the sample mean based on r units selected from n_2 non-responding units on y . S_y^2 and $S_{y(2)}^2$ denote the population mean squares for the whole population and for the non-responding part of the population for the study character y.

In case, when $\bar{X}$ is unknown, we select a large sample of size $n'(> n)$ from the population of size N by using SRSWOR method of sampling and we estimate $\bar{X}$ by $\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i$ and again draw a sub-sample of size n and observed y character.

In this case conventional and alternative two phase sampling ratio type estimators suggested by Khare and Srivastava (2010). In two phase sampling regression estimator for population mean $\bar{Y}$ using one auxiliary character in the presence of non-response have been proposed by Khare and Srivastava (1995) and generalized chain ratio in regression type estimator for the population mean using two auxiliary characters in the presence of non response have been proposed by Khare et al.(2015).which are given as follows:

$$T_1 = \bar{y}^* \left(\frac{\bar{x}^*}{\bar{x}'} \right)^{\alpha} \tag{3}$$

$$T_2 = \bar{y}^* + b_{yx}^* \left(\bar{x}' - \bar{x}^* \right) \tag{4}$$

$$T_3 = \bar{y}^* + b_{yx}^* \left[\bar{x}' \left(\frac{\bar{Z}}{\bar{z}'} \right)^{\alpha_1} - \bar{x}^* \right] \tag{5}$$

where

α and α_1 are constants and b_{yx}^* is regression coefficient.

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

Searls (1964) proposed an estimator $\bar{y}^{**} = a\bar{y}^*$ for $\bar{Y}$ in the presence of non response, where a is a constant. The value of a, for which the $MSE(\bar{y}^{**})$ will be minimum is given by:

$$a_{(opt.)} = [1 + \frac{f}{n}\frac{S_y^2}{\bar{Y}^2} + \frac{W_2(K-1)}{n}\frac{S_{y(2)}^2}{\bar{Y}^2}]^{-1}. \tag{6}$$

Since $\frac{S_y^2}{\bar{Y}^2}$ and $\frac{S_{y(2)}^2}{\bar{Y}^2}$ do not differ significantly, so, we may approximate $\frac{S_{y(2)}^2}{\bar{Y}^2} = C_y^2$ and neglecting the terms of order $\frac{1}{N}$, we have

$$\hat{a}_{(opt)} = \left[1 + \frac{C_y^2}{n}(1 + \frac{n_2}{n}(K-1))\right]^{-1}. \tag{7}$$

Now, we define an estimator $\bar{y}^{**}$ for $\bar{Y}$ by improving the estimator $\bar{y}^*$ using Searls(1964) method, which is given as follows:

$$\bar{y}^{**} = [1 + \frac{C_y^2}{n}(1 + \frac{n_2}{n}(K-1))]^{-1}\bar{y}^* \tag{8}$$

The mean square error (MSE) of $\bar{y}^{**}$ is given as follows:

$$MSE(\bar{y}^{**}) = (1 - A)\frac{S_y^2}{n} + (1 - 2B)\frac{W_2(K-1)}{n}S_{y(2)}^2 \tag{9}$$

where $A = \frac{C_y^2}{n}[1 - W_2^2(K-1)^2]$ and $B = \frac{C_y^2}{n}[1 + W_2(K-1)]$.

Comparing $MSE(\bar{y}^{**})$ with MSE $(\bar{y}^*)$, we see that

$$MSE(\bar{y}^{**}) < MSE(\bar{y}^*) \text{ when } 1 < a < 1 + \frac{1}{W_2}. \tag{10}$$

Under this condition, the estimator may be improved by replacing a better estimator $\bar{y}^{**}$ of $\bar{Y}$ than $\bar{y}^{*}$ in case of non-response in sample survey. However the estimator $\bar{y}^{**}$ may also be more efficient than $\bar{y}^{*}$ beyond the range of $a > 1 + \frac{1}{W_2}$.

Now, using known coefficient of variation of the study character, we propose improved ratio in regression type estimator for population mean in the presence of non-response, which is given as follows:

$$T_R = \bar{y}^{**} + b_{yx}^{*}\left[\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right)^{p} - \bar{x}^{*}\right] \qquad (11)$$

where $\bar{y}^{**} = a\bar{y}^{*}$, p is constant and b_{yx}^{*} is regression coefficient.

In order to derive the expressions for the mean square error of the estimators:

Let $\bar{y}^{*} = \bar{Y}(1+\epsilon_0^{*}), \bar{x}^{*} = \bar{X}(1+\epsilon_1^{*}), \bar{x}' = \bar{X}(1+\epsilon_1')$ and $\bar{z}' = \bar{Z}(1+\epsilon_2')$ such that $E(\epsilon_0^{*}) = E(\epsilon_1^{*}) = E(\epsilon_1') = E(\epsilon_2') = 0, |\epsilon_0^{*}|, |\epsilon_1^{*}|, |\epsilon_1'|, |\epsilon_2'|, < 1.$

By using simple random sampling without replacement method of sampling, we get,

$$E(\epsilon_0^{*2}) = \frac{1}{\bar{Y}^2}V(\bar{y}^{*}) = \frac{1}{\bar{Y}^2}\left\{\left(\frac{1}{n}-\frac{1}{N}\right)S_y^2 + \frac{W_2(K-1)}{n}S_{y(2)}^2\right\}$$
$$E(\epsilon_1^{*2}) = \frac{1}{\bar{X}^2}V(\bar{x}^{*}) = \frac{1}{\bar{X}^2}\left\{\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2 + \frac{W_2(K-1)}{n}S_{x(2)}^2\right\}$$
$$E\left(\epsilon_1'^{2}\right) = \frac{1}{\bar{X}^2}V(\bar{x}') = \frac{1}{\bar{X}^2}\left(\frac{1}{n'}-\frac{1}{N}\right)S_x^2$$
$$E\left(\epsilon_2'^{2}\right) = \frac{1}{\bar{Z}^2}V(\bar{z}') = \frac{1}{\bar{Z}^2}\left(\frac{1}{n'}-\frac{1}{N}\right)S_z^2$$
$$E(\epsilon_0^{*}, \epsilon_1^{*}) = \frac{1}{\bar{Y}\bar{X}}COV(\bar{y}^{*}, \bar{x}^{*}) = \frac{1}{\bar{Y}\bar{X}}\left\{\left(\frac{1}{n}-\frac{1}{N}\right)S_{yx} + \frac{W_2(K-1)}{n}S_{yx(2)}\right\}$$
$$E(\epsilon_0^{*}, \epsilon_1') = \frac{1}{\bar{Y}\bar{X}}COV(\bar{y}^{*}, \bar{x}') = \frac{1}{\bar{Y}\bar{X}}\left(\frac{1}{n'}-\frac{1}{N}\right)S_{yx}$$
$$E(\epsilon_0^{*}, \epsilon_2') = \frac{1}{\bar{Y}\bar{Z}}COV(\bar{y}^{*}, \bar{z}') = \frac{1}{\bar{Y}\bar{Z}}\left(\frac{1}{n'}-\frac{1}{N}\right)S_{yz}$$
$$E(\epsilon_1^{*}, \epsilon_2') = \frac{1}{\bar{X}\bar{Z}}COV(\bar{x}^{*}, \bar{z}') = \frac{1}{\bar{X}\bar{Z}}\left(\frac{1}{n'}-\frac{1}{N}\right)S_{xz}$$
$$E(\epsilon_1^{*}, \epsilon_1') = \frac{1}{\bar{X}^2}COV(\bar{x}^{*}, \bar{x}') = \frac{1}{\bar{X}^2}V(\bar{x}') = \frac{1}{\bar{X}^2}\left(\frac{1}{n'}-\frac{1}{N}\right)S_x^2$$
$$E(\epsilon_1', \epsilon_2') = \frac{1}{\bar{X}\bar{Z}}COV(\bar{x}', \bar{z}') = \frac{1}{\bar{X}\bar{Z}}\left(\frac{1}{n'}-\frac{1}{N}\right)S_{xz}$$

The contribution of the terms involving the powers in $\epsilon_0^{*}, \epsilon_1^{*}, \epsilon_1'$ and ϵ_2' of order higher than two in mean square errors are assumed to be negligible. So, the expressions for the MSE of the proposed estimator and relevant estimators up to the terms of order n^{-1} are given.

Mean Square Errors of the Proposed Estimator T_R

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

$$MSE(T_R)=a^2V\left(\bar{y}^*\right)+(a-1)^2\bar{Y}^2+A_1\left(B_1^2\bar{X}^2C_x^2-2aB_1\bar{X}\bar{Y}C_{yx}\right)+A_2[p^2B_1\bar{X}C_z^2-2(a-1)B_1p\bar{X}\bar{Y}C_{xz}$$

$$-2aB_1p\bar{X}\bar{Y}C_{xz}+(a-1)B_1\bar{X}\bar{Y}p(p+1)C_z^2]+A_3\left(B_1^2\bar{X}^2C_{x(2)}^2-2aB_1\bar{X}\bar{Y}C_{yx(2)}\right) \quad (12)$$

where $a_{(opt.)}=[1+\frac{f}{n}\frac{S_y^2}{\bar{Y}^2}+\frac{W_2(K-1)}{n}\frac{S_{y(2)}^2}{\bar{Y}^2}]^{-1}$,

$$p_{(opt.)}=\frac{(a-1)\bar{Y}C_{xz}-a\bar{Y}C_{yz}-\bar{Y}(a-1)C_z^2/2}{\{B_1\bar{X}+\bar{Y}(a-1)\}C_z^2},$$

Mean square errors of the estimators T_1, T_2 and T_3 are given as follows:

$$MSE(T_1)=V(\bar{y}^*)+\bar{Y}^2\left[A_1\left(\alpha^2C_x^2+2\alpha C_{yx}\right)+A_3\left(\alpha^2C_{x(2)}^2+2\alpha C_{yx(2)}\right)\right] \quad (13)$$

$$MSE(T_2)=V\left(\bar{y}^*\right)-\bar{Y}^2\left[A_1\rho_{yx}^2C_y^2-A_3\left\{B_1^2C_{x(2)}^2-2B_1C_{yx(2)}\right\}\right] \quad (14)$$

$$MSE(T_3)=V\left(\bar{y}^*\right)+A_1\left(B_1^2\bar{X}^2C_x^2-2B_1\bar{X}\bar{Y}C_{yx}\right)+A_2\left\{\alpha_1^2B_1\bar{X}C_z^2-2B_1\alpha_1\bar{X}\bar{Y}C_{xz}\right\}+A_3\left(B_1^2\bar{X}^2C_{x(2)}^2-2B_1\bar{X}\bar{Y}C_{yx(2)}\right) \quad (15)$$

where

$$V(\bar{y}^*)=\bar{Y}^2\left\{\frac{f}{n}C_y^2+\frac{W_2(k-1)}{n}C_{y(2)}^2\right\}, \quad \alpha_{1(opt.)}=\frac{\bar{Y}C_{xz}-\bar{Y}C_{yz}}{B_1\bar{X}C_z^2}, \quad \alpha_{(opt.)}=-\frac{A_1C_{yx}+A_3C_{yx(2)}}{A_1C_x^2+A_3C_{x(2)}^2},$$

$$B_1=\frac{\bar{Y}\rho_{yx}C_y}{\bar{X}C_x} \quad A_1=\left(\frac{1}{n}-\frac{1}{n'}\right), \quad A_2=\left(\frac{1}{n'}-\frac{1}{N}\right) \text{ and } A_3=\frac{W_2(k-1)}{n}$$

$\rho_{yx(2)}=\frac{S_{yx(2)}}{S_{y(2)}S_{x(2)}}$ and $(S_{yx(2)},\rho_{yx(2)})$ denote the covariance and correlation between y and x characters for the non response group of the population.

An Empirical Study

To illustrate the results, we have considered the data by Khare and Sinha (2009).The description of the population is given below:

96 village wise population of rural area under Police-station – Singur, District -Hooghly, West Bengal has been taken under the study from the District Census Handbook 1981. The 25% villages (i.e. 24 villages) whose area is greater than 160 hectares have been considered as non-response group of the population. The number of agricultural laboures in the village is taken as study character (y) while the area (in hectares) of the village, the number of cultivators in the village and the total population of the village are taken as auxiliary characters x and z respectively.
The values of the parameters of the population under study are as follows:

$\bar{Y}=137.9271$, $\bar{X}=144.8720$, $\bar{Z}=185.2188$, $S_y=182.5012$, $C_x=.8115$,

$S_{y(2)}=287.4202$, $C_{x(2)}=.9408$, $C_z=1.0529$, $C_{z(2)}=1.4876$, $\rho_{yx}=0.773$,

$\rho_{yx(2)} = 0.724$, $\rho_{yz} = 0.786$, $\rho_{yz(2)} = 0.787$, $\rho_{xz} = 0.819$, $\rho_{xz(2)} = 0.724$, $W_2 = 0.25$, $N = 96$, $n = 24$, $n' = 60$.

***Table 1:** Relative efficiency (in %) of the estimators with respect to $\bar{y}^*$ for the fixed values of n', n and different values of k (N =96, n' =60 and n =24).*

Estimators	$1/k$		
	1/4	1/3	1/2
$\bar{y}^*$	100 (32.93)*	100 (28.52)	100 (24.10)
T_1	142(23.14)	152(18.73)	168 (14.31)
T_2	132 (24.88)	143 (19.89)	162 (14.90)
T_3	146 (22.49)	140(20.38)	132(18.29)
T_R	150 (21.88)	144 (19.79)	136 (17.73)

*Figures in parenthesis give the MSE (.).

From table 1, we observe that for the fixed sample sizes (n', n), the proposed estimator T_R is more efficient in comparison to the estimators $\bar{y}^*$, T_1, T_2 and T_3. Which suggest that the use of coefficient of variation of the study character is beneficial in increasing the efficiency of the proposed estimator in comparison to relevant estimators without using coefficient of variation of study character.

References

[1] Das, A. K., Tripathi, T. P. (1980). Sampling strategies for population mean when the coefficient of variation of an auxiliary character is known. Sankhya C-42: 76-86.

[2] Hansen, M. H. and Hurwitz, W. N.(1946). The problem of non-response in sample surveys. J. Amer. Statist. Assoc., 41, 517-529.

[3] Khare, B. B. and Srivastava, S.(1993). Estimation of population mean using auxiliary character in presence of non-response. Nat. Acad. Sci. Letters, India, 16(3), 111-114.

[4] Khare, B. B. and Srivastava, S.(1995). Study of conventional and alternative two phase sampling ratio, product and regression estimators in presence of non-response. Proc. Nat. Acad. Sci., India, Sect. A 65(a) II, 195-203.

[5] Khare, B. B. (2003). Use of auxiliary information in sample surveys up to 2000- A review. In: Sen, D. K., Mishra, P.K., eds. Mathematics and Statistics in Engineering, Biotechnology and Science, M/S centre of Bio-mathematical studies, India 76-87.

[6] Khare, B. B. and Sinha, R.R.(2009). On class of estimators for population means using multi- auxiliary characters in presence of non-response. Stats. in Transition-new series.(Poland),10(1),3-14.

[7] Khare, B.B. and Kumar, S.(2009). Transformed two phase sampling ratio and product type estimators for population mean in the presence of non response. Aligarh J. Stats., 29, 91-106.

[8] Khare, B. B. and Srivastava, S.(2010). Generalized two phase sampling estimators for the population mean in the presence of nonresponse. *Aligarh. J. Stats.,* 30, 39-54.

[9] Khare, B.B., Kumar, K. and Srivastava, U. (2015). Generalized chain ratio in regression estimators for population mean using two auxiliary characters in the presence of non-response. Elect. Jour. of Appl. Stats. Anal., 8(1),67-83.

[10] Searls, D. T (1964). The utilization of a known coefficient of variation in the estimation procedure. J. Amer. Statist. Assoc. 59:1225-1226.

[11] Searls, D. T. (1967). A note on the use of an approximately known coefficient of variation. J. Amer. Statist. Assoc. 21:20–21.

[12] Sen, A. R. (1978). Estimation of the population mean when the coefficient of variation is known. Commun. Statist.- Theor. Meth., A7(7):657-672.

[13] Tripathi, T. P., Das, A. K., Khare, B. B. (1994). Use of auxiliary information in sample surveys- A review. Aligarh J. Statist. 14:79-134.

Authors

Dr. B. B. Khare, presently working as Professor of Statistics, Department of Statistics, Faculty of Science, Banaras Hindu University, Varanasi, India since 2006. Joined the Department as lecturer in 1984. Dr. Khare did his M.Sc. (Statistics) and Ph.D. (Statistics) from Banaras Hindu University in 1977 and 1984 respectively. He has more than 30 years teaching and research experience.
The fields of specialization are Sampling Theory and Population studies (physical growth and development, breast feeding and weaning practices during infancy), published 95 research papers in the reputed international/national journals. Guided 5 Ph. D. scholars for their Ph. D. Degree. About 60 papers presented in international/national conferences, Life member of National Academy of Sciences and M. N. A. Sc. (India) and reviewer of a number of international/national journals of repute.

Habib Ur Rehman, pursuing research work for Ph.D. degree in Statistics in the Department of Statistics, Banaras Hindu University Varanasi-221005, India, on the topic "Estimation Procedure for the Population Parameters Utilizing Auxiliary Variables in the Presence of Non-Response". He has published 8 research papers in the reputed international/national journals. He has done M.Sc. and M.Phil. in Statistics from C. C. S. University, Meerut, UP, India.