Journal of Social and Economic Statistics

Vol. 3, No. 1, Summer 2014

GENERALIZED RATIO CUM REGRESSION TYPE ESTIMATOR FOR POPULATION MEAN USING AUXILIARY VARIABLE WITH DOUBLE SAMPLING IN THE PRESENCE OF NON RESPONSE

B. B. KHARE[a], Habib ur REHMAN[b]

Abstract

Generalized ratio cum regression type estimator for population mean using auxiliary variable with double sampling in the presence of non response has been proposed and its properties have been studied. A comparative study of the proposed estimator has been made with the relevant estimators. For optimum value of α_1, the proposed estimator is found to be more efficient than the relevant estimator which is supported by the empirical study. A range for α_1 has been obtained for which the proposed estimator is more efficient than relevant estimators.

Keywords: Mean square error, two phase sampling, auxiliary variable, Non-response
JEL Classification: C83

Authors' Affiliation

[a]Banaras Hindu University, Faculty of Science, Department of Statistics, Varanasi, 221005, India, bbkhare56@yahoo.com

[b]Banaras Hindu University, Faculty of Science, Department of Statistics, Varanasi, 221005, India, hrmbd007@gmail.com

1. Introduction

Sample surveys are generally used in the field of agricultural, social and medical sciences. During the sample surveys, sometimes information on some units in the selected sample is not obtained due to the problem of non-response. Hansen and Hurwitz (1946) have suggested a technique of sub sampling from non-respondents. The information on the auxiliary character provides a very important contribution in the field of sample surveys. In the case when the population mean of the auxiliary character x is known, Cochran (1977), Rao (1986, 87, 90) and Khare and Srivastava (1996, 97) have proposed ratio, product and regression type estimators in the presence of non-response. In the situation when the population mean of an auxiliary character is not known, the two phase sampling ratio, product and regression type estimators have been suggested by Khare and Srivastava (1993, 95), Khare et al. (2008), Khare and Kumar (2009) and Khare and Srivastava (2010). In the case when the population mean of the main auxiliary character is not known but the population mean of an additional auxiliary character is known, which is cheaper than the main auxiliary character but less correlated to the study character than the main auxiliary character. In this situation, Chand (1975) proposed chain ratio type estimator for the population mean of the study character. Kiregyra (1980, 84) proposed chain ratio to regression, ratio in regression and regression in regression estimators for the population mean of study character. Further some class of estimators for population mean of study character has been proposed by Srivastava et al. (1990), Sahoo et al. (1993), Singh et al. (2001) and Dash and Mishra (2011). In such situation, Khare and Kumar (2010) and Khare et al. (2011) have proposed chain regression type estimators and generalized chain estimators for the population mean in the presence of non-response.

In the present paper, we have proposed generalized ratio cum regression type estimator for the population mean using auxiliary variable with double sampling in the presence of non-response. We have obtained the expressions for mean square errors of the proposed estimators for the fixed first phase sample (n'), second phase sample (n) and also for the optimum values of the constants. A comparative study of the mean square errors of the proposed estimators is made with the relevant estimators.

2. The Estimators

Let (Y_j, X_j) be the value of *j*th unit of the population of size N for the study character y and the auxiliary character x having population mean $(\bar{Y}, \bar{X})$. Suppose population is divided in N_1 responding and N_2 non responding group. In case when population mean of auxiliary character is not known, we take a first phase sample of size $n'(< N)$ from population of size N by using simple random sampling without replacement (SRSWOR) scheme of sampling and estimate the population mean $\bar{X}$ by $\bar{x}'$ which is the mean of the values of x on the first phase sample. Further we select a second phase sample of size $n(< n')$ from first phase sample of size n' by using SRSWOR scheme of sampling and observe that all n unit respond on auxiliary character but on study character y, only n_1 units respond and n_2 units do not respond from n unit. Again, we take a subsample of size $r = (n_2 k^{-1},\ k > 1)$ from n_2 non-responding units and collect information on r units by interview. Let $(\bar{y}_1, \bar{y}_2')$ be the sample means based on n_1 and r units. Hence estimator $\bar{y}^*$ suggested by Hansen and Hurwitz (1946) is given as follows:

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

Let $(\bar{x}_1, \bar{x}_2')$ be the sample means based on n_1 and r units. So the estimator $\bar{x}^*$ based on n_1 and r unit corresponding to the estimator $\bar{y}^*$ is given as:

$$\bar{x}^* = \frac{n_1}{n}\bar{x}_1 + \frac{n_2}{n}\bar{x}_2'. \tag{2}$$

The variance of the unbiased estimator $\bar{y}^*$ and $\bar{x}^*$ are given as follows:

$$V(\bar{y}^*) = f\, S_y^2 + \frac{W_2(k-1)}{n} S_{y(2)}^2 \tag{3}$$

and

$$V(\bar{x}^*) = f\, S_x^2 + \frac{W_2(k-1)}{n} S_{x(2)}^2 \tag{4}$$

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

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) and generalized ratio in regression type estimators for population mean using auxiliary variables in the presence of non-response have been proposed by Khare and Rehman (2014) which are given as follows:

$$T_1 = \bar{y}^* \frac{\bar{x}'}{\bar{x}^*},\ T_2 = \bar{y}^* \frac{\bar{x}'}{\bar{x}} \tag{5}$$

$$T_3 = \bar{y}^* + b^{**}\left(\bar{x}' - \bar{x}\right) \tag{6}$$

$$T_4 = \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}^*}\right)^{\alpha} + b^*\left(\bar{x}' - \bar{x}^*\right) \tag{7}$$

where α is constant and $\bar{x}^* = \frac{n_1}{n}\bar{x}_1 + \frac{n_2}{n}\bar{x}_2'$, $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$, $\bar{x}' = \frac{1}{n'}\sum_{j=1}^{n'} x_j, b* = \frac{\hat{S}_{yx}}{\hat{S}_x^2}, b^{**} = \frac{\hat{S}_{yx}}{s_x^2}$,

$$s_x^2 = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

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

In this case we have proposed generalized ratio cum regression type estimator for population mean using auxiliary variables with double sampling in the presence of non-response, which is given as follows:

$$T_5 = \left[\bar{y}^* + b_1\left(\bar{x} - \bar{x}^*\right) + b_2\left(\bar{x}' - \bar{x}\right)\right]\left[2 - \left(\frac{\bar{x}'}{\bar{x}^*}\right)^{\alpha_1}\right] \tag{8}$$

where α_1 is constant, b_1 and b_2 are regression coefficient of y and x.

3. Mean Squared Errors on the Proposed Estimator T_5

Using the large sample approximations, the expressions for the mean square errors of the estimator T_5 up to the terms of order (n^{-1}) are given by:

$$MSE(T_5) = V(\bar{y}^*) + \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{\left(\bar{Y}^2\alpha_1^2 + b_2^2\bar{X}^2 + 2b_1\alpha_1\bar{X}\bar{Y} - 2b_2\alpha_1\bar{X}\bar{Y}\right)C_x^2 + 2\bar{Y}^2\alpha_1 C_{yx} - 2\bar{Y}\bar{X}b_2 C_{yx}\right\}$$

$$+\frac{W_2(k-1)}{n}\left\{\left(\bar{Y}^2\alpha_1^2 + b_1^2\bar{X}^2 - 2b_1\alpha_1\bar{X}\bar{Y}\right)C_{x(2)}^2 - 2\bar{Y}\bar{X}b_1 C_{yx(2)} + 2\alpha_1\bar{Y}^2 C_{yx(2)}\right\}. \tag{9}$$

The optimum value of α_1 and regression coefficient are given as follows:

$$a_{1(opt)} = \frac{\left(\frac{1}{n} - \frac{1}{n'}\right)\left(\bar{X}b_2C_x^2 - \bar{Y}C_{yx} - \bar{X}b_1C_x^2\right) + \frac{W_2(k-1)}{n}\left(-\bar{Y}C_{yx(2)} + \bar{X}b_1C_{x(2)}^2\right)}{\left(\frac{1}{n} - \frac{1}{n'}\right)\bar{Y}C_x^2 + \frac{W_2(k-1)}{n}\bar{Y}C_{x(2)}^2},$$

$$b_1 = \frac{\bar{Y}\rho_{yx(2)}}{\bar{X}}\frac{C_{y(2)}}{C_{x(2)}} \text{ and } b_2 = \frac{\bar{Y}\rho_{yx}}{\bar{X}}\frac{C_y}{C_x}. \tag{10}$$

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

$$MSE(T_1) = V(\bar{y}^*) + \left[\left(\frac{1}{n} - \frac{1}{n'}\right)\left(C_x^2 - 2\rho_{yx}C_yC_x\right) + \frac{W_2(k-1)}{n}\left(C_{x(2)}^2 - 2\rho_{yx(2)}C_{y(2)}C_{x(2)}\right)\right]\bar{Y}^2, \tag{11}$$

$$MSE(T_2) = V(\bar{y}^*) + \left[\left(\frac{1}{n} - \frac{1}{n'}\right)\left(C_x^2 - 2\rho_{yx} C_y C_x\right)\right]\bar{Y}^2, \tag{12}$$

$$MSE(T_3) = V(\bar{y}^*) - \bar{Y}^2\left(\frac{1}{n} - \frac{1}{n'}\right)\rho_{yx}^2 C_y^2, \tag{13}$$

$$MSE(T_4) = V(\bar{y}^*) + \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{\bar{Y}^2\alpha^2 C_x^2 + b^2\bar{X}^2 C_x^2 - 2\bar{Y}^2\alpha C_{yx} - 2\bar{Y}\bar{X}bC_{yx} + 2\bar{Y}\bar{X}b\alpha C_x^2\right\}$$

$$+\frac{W_2(K-1)}{n}\left\{\bar{Y}^2\alpha^2 C_{x(2)}^2 + b^2\bar{X}^2 C_{x(2)}^2 - 2\bar{Y}^2\alpha C_{yx(2)} - 2\bar{X}\bar{Y}bC_{yx(2)} + 2\bar{X}\bar{Y}b\alpha C_{x(2)}^2\right\}. \tag{14}$$

The optimum value of α and regression coefficient are given as follows:

$$a_{opt} = \frac{\left(\frac{1}{n} - \frac{1}{n'}\right)\left(\bar{Y}C_{yx} - \bar{X}bC_x^2\right) + \frac{W_2(k-1)}{n}\left(\bar{Y}C_{yx(2)} - \bar{X}bC_{x(2)}^2\right)}{\left(\frac{1}{n} - \frac{1}{n'}\right)\bar{X}C_x^2 + \frac{W_2(k-1)}{n}\bar{Y}C_{x(2)}^2},$$

where

$$V(\bar{y}^*) = \bar{Y}^2\left\{\frac{f}{n}C_y^2 + \frac{W_2(k-1)}{n}C_{y(2)}^2\right\} \text{ and } B = \frac{\bar{Y}\rho_{yx}C_y}{\bar{X}C_x}. \tag{15}$$

4. An Empirical Study

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

Data Set- 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 labors 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:

$\bar{Y} = 137.9271$, $\bar{X} = 144.8720$, $S_y = 182.5012$, $C_x = .8115$, $S_{y(2)} = 287.4202$, $C_{x(2)} = .9408$, $C_{y(2)} = 1.0776$, $\rho_{yx} = 0.773$, $\rho_{yx(2)} = 0.724$, $W_2 = 0.25$, $N = 96$, $n = 40$, $n' = 70$.

Table 1. Relative efficiency of the estimators (in %) with respect to $\bar{y}^*$ for fixed values of n', n and different values of k (N=96, n'=70 and n=40).

Estimators	$1/k$		
	1/4	1/3	1/2
$\bar{y}^*$	100 (11169.8)*	100 (9456.1)	100 (7742.5)
T_1	184 (6055.5)	182 (5202.1)	178 (4348.6)
T_2	129 (8636.1)	137 (6922.4)	149 (5208.8)
T_3	131 (8523.1)	139 (6809.4)	152 (5095.8)
T_4	189 (5920.7)	184 (5133.5)	177 (4371.1)
T_5	**209** (5348.5)	**220** (4292.1)	**256** (3023.7)

*Figures in parenthesis give the MSE (.).

From table, we obtained that for fixed sample sizes (n', n), the proposed estimators T_5 is more efficient in comparison to the efficiency of the estimators $\bar{y}^*$, T_1, T_2, T_3 and T_4.

References

Chand, L. (1975). Some ratio-type estimators based on two or more auxiliary variables. Ph.D. Thesis submitted to Iowa State University, Ames, IOWA.

Cohran, W. G. (1977). *Sampling Techniques*, 3rd ed. New York: Wiley.

Das, P.R. and Mishra, G. (2011). An improved class of estimators in two phase sampling using two auxiliary variables. *Commun. Stats.-Theo. Meth.*, **40**(24), pp. 4347-4352.

Hansen, M. H. and Hurwitz, W. N. (1946). The problem of non-response in sample surveys. *Jour. Amer. Assoc*, **41**, pp. 517-529.

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), pp. 111-114.

Khare, B. B. and Srivastava, S. (1995). Study of conventional and alternative two phase sampling ratio, product and regression estimators in presence of nonresponse. *Proc. Nat. Acad. Sci. India*, 65(A) II, pp. 195-203.

Khare, B. B. and Srivastava, S. (1996). Transformed product type estimators for population mean in presence of softcore observations. *Proc. Math. Soc. B.H.U.*, **12**, pp. 29-34.

Khare, B. B. and Srivastava, S. (1997). Transformed ratio type estimators for population mean in the presence of nonresponse. *Commun. Statist. Theory Meth.*, **26**(7), pp. 1779-1791.

Khare, B. B., Kumar, Anupam, Sinha, R. R. and Pandey, S. K. (2008). Two phase sampling estimators for population mean using auxiliary character in presence of non-response in sample surveys. *Jour. Sci. Res. BHU*, **52**, pp. 271-281.

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

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), pp. 3-14.

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**, pp. 39-54.

Khare, B. B. and Kumar, S. (2010). Chain regression type estimators using additional auxiliary variable in two phase sampling in the presence of non-response. *Nat. Acad. Sci. Lett. India*, **33**(11&12), pp. 369-375.

Khare, B. B., Srivastava, U. and Kumar, Kamlesh (2011). Generalized chain estimators for the population mean in the presence of non-response. *Proc. Nat. Acad. Sci. India* 81(A), pt III, pp. 231-238.

Khare, B. B. and Rehman, H. U.(2014). A generalized ratio in regression type estimator for population mean using auxiliary variables in the presence of non-response. *Int. Jour. Stat. & Eco.*, **15**(3), pp. 64-68.

Kiregyera, B.(1980). A chain ratio type estimator in finite population double sampling using two auxiliary variables. *Metrika*, **27**, pp. 217-223.

Kiregyera, B.(1984). Regression - type estimators using two auxiliary variables and model of double sampling from finite populations. *Metrika*, **31**, pp. 215-226.

Rao, P. S. R. S. (1986). Ratio estimation with subsampling the nonrespondents. *Survey Methodology*, **12**(2), pp. 217-230.

Rao, P. S. R. S. (1987). Ratio and regression Estimates with subsampling the non-respondents. *Paper presented at a special contributed session of the International Statistical Association Meeting,* September, 2-16, Tokyo, Japan.

Rao, P. S. R. S. (1990). Regression estimators with sub sampling of nonrespondents. *In*: Gumer E. Liepins and V. R. R. Uppuluri, Marcel Dekker, eds. *Data quality control theory and Pragmatics*. New York, pp. 191-208.

Sahoo, J., Sahoo, L. N. and Mohanty, S. (1993). A regression approach to estimation in two-phase sampling using two auxiliary variables. *Current Science*, 65(1), pp. 73-75.

Singh, A. K., Singh, H. P. and Upadhyaya, L. N. (2001). A generalized chain estimator for finite population mean in two phase sampling. *Jour. Ind. Soc. Ag. Statist.*, **54**(3), pp. 370-375.

Srivastava, S. R., Khare, B. B. and Srivastava, S. R.(1990). A generalized chain ratio estimator for mean of finite population. *Jour. Ind. Soc. Agri. Stat.*, **42**(1), pp. 108-117.