International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Modified Product Type Estimator Under Adaptive Cluster Sampling

Arshid Ahmad Bhat[1*], Manish Sharma[2], R. K. Salgotra[3], Anil Bhat[4], Mehraj Ud Din Bhat[5], Mubeena Akhter[6], Mohammad Saleem Mir[7]

[1*]MRHRU, NIP, Indian Council of Medical Research, New- Delhi, 110001

[2] Division of Statistics and Computer Science, Sher-e-Kashmir University of Agricultural Sciences and Technology-Jammu, 180009

[3]School of Biotechnology, Sher-e-Kashmir University of Agricultural Sciences and Technology-Jammu, 180009

[4]Division of Econ & ABM, FoA, Sher-e-Kashmir University of Agricultural Sciences and Technology-Jammu, 180009

[5]Jammu and Kashmir Entrepreneurship Development Institute-Pampore, Jammu and Kashmir, 191101

[6]Independent Researcher, 192123

[7]MRHRU, NIP, Indian Council of Medical Research, New- Delhi, 110001

DOI: https://doi.org/10.55248/gengpi.4.1223.123544

ABSTRACT:

Adaptive Cluster Sampling (ACS) is one of the sampling methods used for the computation of population mean & variance when the population is uncommon and concentrated. In this article a product type estimator has been developed for the estimation of population mean with the support of auxiliary variable on the condition that the study and the auxiliary variable are negatively correlated both at unit level and at network level. The mathematical expression of the proposed estimator for mean squared error (MSE) has been obtained upto first order of approximations. This theoretical expression has been examined to check the performance of the developed estimator. Results showed that the proposed product type estimator outperforms the existing conventional and ACS product type estimators.

Keywords: ACS, product type estimator, Negative correlation, MSE.

1. Introduction

The computation of population mean for uncommon, rare, spatially challenging-to-reach population units is believed to be ideally supported by adaptive cluster sampling (ACS) proposed by Thompson in 1990. The key feature of this design is that, it provides more relevant samples and produces more realistic measurements of mean, variance etc. in comparison to other traditional sampling designs. In order to enhance the estimates of population parameters like average, variation etc. auxiliary information is often collected in parallel with the main variable. To obtain better estimation results the ratio, product, and regression estimation methods are mostly used as auxiliary variables. Many authors improved the efficiency of the mean estimators using auxiliary data and suggested various estimators such as ratio-product estimators, exponential ratio-product, weighted ratio-product, and weighted exponential ratio-product under ACS etc. that combine the ratio and product methods of estimate simultaneously. The ratio and product estimators for the estimation of population mean using simple random sampling proposed by Cochran (1940) and Robson (1957) were without using the auxiliary information. Later on Sisodia and Dwivedi (1981), Upadhyay and Singh (1999), Singh and Tailor (2003), Kadilar and Cingi (2003), Sharma and Bhatnagar (2008), Yan and Tian (2010), Jeelani et al., (2017), Kumar et al., (2018), Hussain et al., (2021), Arshid et al. proposed different ratio type estimators for the estimation of population mean by using coefficient of variation (CV), kurtosis, correlation coefficient, skewness etc. as the auxiliary variables.

2. Methodology:

In ACS, an initial sample is selected by conventional sampling design especially by simple random sampling without replacement (SRSWOR). A specific condition is defined in advance in order to adopt the neighbouring units. The neighbourhood is the spatially adjacent units in the east, west, north and south of the selected units that meets the pre-specified condition. If the specified condition is satisfied in the initially selected samples selected by SRSWOR the neighbouring units are examined and added to the sample, this process continues until the neighbouring units don't meet the condition. The units that meet the condition form the network and the units that don’t meet the condition form edge units. The collection of network and edge units forms as cluster.

0	0	3*	5	0*	0	0	0	0	0
0	0	0	24*	14	0	0*	10	103	0
0	0	0	0	2	3	2	0	13639	1
0	0	0	0	0	0	0	0	14*	122
0	0	0	0	0	0	2	0	0	177

Figure 1 population of study variable (y)

0	0	3*	**5**	**0***	0	0	**0**	**0**	0
0	0	**0**	19*	19	**0**	**0***	2344.16	2344.16	**0**
0	0	0	**0**	**2**	3	2	**0**	2344.16	**1**
0	0	0	0	0	0	0	**0**	2344.16*	2344.16
0	0	0	0	0	0	2	0	**0**	2344.16

Figure 2 transformed population with average values of networks ($\boldsymbol{w_y}$)

1	1	1*	1	1*	1	1	1	1	1
1	1	1	0*	1	1	1*	1	0	1
1	1	1	1	1	1	1	1	0	1
1	1	1	1	1	1	1	1	1*	0
1	1	1	1	1	1	1	1	1	0

Figure 3 population of auxiliary variable (x)

1	1	1*	**1**	**1***	1	1	**1**	**1**	1
1	1	**1**	0.5*	0.5	**1**	**1***	0.33	0.33	**1**
1	1	1	**1**	**1**	1	1	**1**	0.33	**1**
1	1	1	1	1	1	1	**1**	0.33*	0.33
1	1	1	1	1	1	1	1	**1**	0.33

Figure 4 transformed population with average values of networks ($\boldsymbol{w_x}$)

The figure 1 and 3 illustrates an example of a ACS design each containing 5x10 = 50 population units. The units superscripted with a star (*) are first chosen units. The condition of adaption $y \geq 10$ is pre-defined for a unit to be included in the network. The units that are in the east, west, north and south of the first chosen sample are named as first order neighborhood. The cells in the shaded region in figure 2 and 4 form a network while the units in bold numerals are respective edge units of a network. A cluster is comprised of the network and its associated edge units. In figure 2 there are 42 networks of size 1 including edge units and two clusters where first cluster contains 8 units with 6 edge units and second cluster contains 14 units with 8 edge units.

Let the usual finite population consists N distinct units labelled from 1,2,…, N. the variablesy_i and x_i (i=1,2,3…..,N) denote the ith value for the survey and auxiliary variables respectively, 'n' denote the initial sample size. Let the population is divided into K exhaustive networks where φ_i denotes the network that includes i units with m_i number of units in the ith network. The mean, standard deviation, coefficient of variation, correlation coefficient and covariance of survey and auxiliary variable at network level is denoted by $\bar{w}_y, \bar{w}_x, \sigma_{wy}, \sigma_{wx}, C_{wy}, C_{wx}, \rho_{wxwy}, \sigma_{wywx}$ respectively. The parameters of study and auxiliary variables are as:

Study variable

Mean: $\mu_y = N^{-1}\sum_{i=1}^{N} y_i$

Variance: $\sigma_y^2 = (N-1)^{-1}\sum_{i=1}^{N}(y_i - \mu_y)^2$

Auxiliary variable

Mean: $\mu_x = N^{-1}\sum_{i=1}^{N} x_i$

Variance: $\sigma_x^2 = (N-1)^{-1}\sum_{i=1}^{N}(x_i - \mu_x)^2$

Let $w_{yi} = \frac{1}{m_i}\sum_{j\epsilon\varphi_i} y_i$ and $w_{xi} = \frac{1}{m_i}\sum_{j\epsilon\varphi_i} x_i$ be the transformed survey and auxiliary variables in the ith network respectively.

The transformed population parameters of study and auxiliary variables are:

Study variable

Mean: $\mu_{wy} = \mu_y = N^{-1}\sum_{i=1}^{N} w_{yi}$

Variance: $\sigma_{wy}^2 = (N-1)^{-1}\sum_{i=1}^{N}(w_{yi} - \mu_{wy})^2$

Auxiliary variable

Mean: $\mu_{wx} = \mu_x = N^{-1}\sum_{i=1}^{N} w_{xi}$

Variance:$\sigma_{wx}^2 = (N-1)^{-1}\sum_{i=1}^{N}(w_{xi} - \mu_{wx})^2$

The transformed population statistics of study and auxiliary are:

Study variable

Mean: $\bar{w}_y = n^{-1}\sum_{i=1}^{n} w_{yi}$

Variance: $s_{wy}^2 = (n-1)^{-1}\sum_{i=1}^{n}(w_{yi} - \bar{w}_y)^2$

Auxiliary variable

Mean: $\bar{w}_x = n^{-1}\sum_{i=1}^{n} w_{xi}$

Variance: $s_{wx}^2 = (n-1)^{-1}\sum_{i=1}^{n}(w_{xi} - \bar{w}_x)^2$

2.1 Existing estimators

The classical estimator proposed by Robson (1957) is:

$$\hat{\mu}_{yR} = \mu_y \frac{\bar{x}}{\hat{\mu}_x}$$

The mean square error is:

$$\text{MSE}\left(\hat{\mu}_{yR}\right) = \theta\mu_y^2\left(C_y^2 + C_x^2 + 2\rho_{xy}C_xC_y\right) \tag{1}$$

Bahl and Tuteji (1991) proposed the exponential product type estimator as:

$$\hat{\mu}_{yBT} = \mu_y\left(\frac{\bar{x} - \mu_x}{\bar{x} + \mu_x}\right)$$

The mean square error is:

$$\text{MSE}(\hat{\mu}_{yBT}) = \theta\mu_y^2(C_y^2 + \frac{C_x^2}{4} + \rho_{xy}C_xC_y \tag{2}$$

Where; $\theta = \frac{N-n}{Nn} = \frac{1-f}{n}, f = \frac{n}{N}$ and f is the sampling fraction.

C_y(Coefficient of variation of study variable) $= \frac{S_y}{\mu_y}$, C_x(Coefficient of Variation of auxiliary variable) $= \frac{S_x}{\mu_x}$

ρ_{xy}(Population Correlation Coefficient) $= \frac{S_{xy}}{S_xS_y}$

$S_y^2 = (N-1)^{-1}\sum_{i=1}^{N}\left(y_i - \mu_y\right)^2$, $S_x^2 = (N-1)^{-1}\sum_{i=1}^{N}(x_i - \mu_x)^2$,$S_{xy} = (N-1)^{-1}\sum_{i=1}^{N}(x_i - \mu_x)\left(y_i - \mu_y\right)$

Thompson (1990) proposed the mean unbiased estimator based on the modification of HH type estimator under ACS as:

$$\hat{\mu}_{yT} = n^{-1}\sum_{i=1}^{n} w_{yi}$$

The variance is:

$$\text{var}\left(\hat{\mu}_{yT}\right) = \theta\mu_y^2C_{wy}^2 \tag{3}$$

Shahzad and Hanif (2016) proposed classical product and exponential product type estimator in ACS as:

$$\hat{\mu}_{SH} = \bar{w}_y\frac{\bar{w}_x}{\hat{\mu}_{wx}}$$

$$\hat{\mu}_{SH'} = \bar{w}_y\exp\left(\frac{\bar{w}_x - \mu_x}{\bar{w}_x - \mu_x}\right)$$

The mean square error is:

$$\text{MSE}\left(\hat{\mu}_{ySH}\right) = \theta\mu_y^2\left(C_{wy}^2 + C_{wx}^2 + 2\rho_{wxwy}C_{wx}C_{wy}\right) \tag{4}$$

$$\text{MSE}(\hat{\mu}_{ySH'}) = \theta\mu_y^2(C_{wy}^2 + \frac{Cw_x^2}{4} + \rho_{wxwy}C_{wx}C_{wy} \tag{5}$$

3. Proposed estimator in ACS:

Keeping in consideration the above estimators the proposed estimator under adaptive cluster sampling design is:

$$\hat{\mu}_{wy(k,k')}^{p} = \bar{w}_y\left(\frac{k\bar{w}_x - \mu_x}{k'\bar{w}_x + \mu_x}\right)$$

Where k and k′are the constants to be determined such that the proposed estimator is efficient. For obtaining the theoretical expression of Mean square error of the proposed estimator let's define

$$U_y = \frac{\bar{w}_y - \mu_y}{\mu_y} \text{ and } U_x = \frac{\bar{w}_x - \mu_x}{\mu_x}$$

Under SRSWOR the expected values of these quantities are:

$$E\left(U_y\right) = E(U_x) = 0, E\left(U_y\right)^2 = \theta C_{wy}^2, E(U_x)^2 = \theta C_{wx}^2, E\left(U_yU_x\right) = \theta\rho_{wxwy}C_{wy}C_{wx}$$

On rewriting the proposed estimator in terms of U_y and U_x we obtained the following expression

$$\hat{\mu}_{wy(k,k')}^{p} = \left(1 + U_y\right)\mu_y\frac{k(1 + U_x)\mu_x - \mu_x}{k'(1 + U_x)\mu_x + \mu_x}$$

On solving the above equation upto second degree the following expression is obtained

The mean square error of the proposed estimator is:

$$MSE\left(\hat{\mu}^{p}_{wy(k,k')}\right) = \mu_y^2\{(\alpha-1)^2 + \alpha^2\theta C_{wy}^2 + [(\beta-\alpha)^2 + 2(\alpha-1)(\alpha\gamma^2-\beta\gamma)]\theta C_{wx}^2 + [2(\alpha-1)(\beta-\alpha\gamma) + 2\alpha(\beta-\alpha)]\theta\rho_{wxwy}C_{wx}C_{wy}\} \quad (7)$$

Where $\alpha = \frac{k-1}{k'+1}, \beta = \frac{k}{k'+1}, \gamma = \frac{k'}{k'+1}, \theta = n^{-1} - N^{-1}, C_{wy}^2 = \frac{S_{wy}^2}{\mu_{wy}^2}$

4. Results and Discussion:

Theoretical efficiency comparison:

These are the algebraic expressions that when employed to the proposed estimators, would result in the lowest MSE as compared to literature-based estimators.

(a) Proposed estimator performs better than Robson (1999) if,

$$MSE\left(\hat{\mu}^{p}_{wy(k,k')}\right) < MSE\left(\hat{\mu}_{yR}\right)$$

$$\{(\alpha-1)^2 + \theta(\alpha^2 C_{wy}^2 - C_y^2) + \theta\{[(\beta-\alpha)^2 + 2(\alpha-1)(\alpha\gamma^2-\beta\gamma)]C_{wx}^2 - C_x^2\} + 2\theta([(\alpha-1)(\beta-\alpha\gamma) + \alpha(\beta-\alpha)]\rho_{wxwy}C_{wx}C_{wy} - \rho_{xy}C_xC_y)\} < 0 \quad (7)$$

(b) Proposed estimator performs better than Bahl and Tuteji (1991) if,

$$MSE\left(\hat{\mu}^{p}_{wy(k,k')}\right) < MSE\left(\hat{\mu}_{yBT}\right)$$

$$\left\{(\alpha-1)^2 + \theta(\alpha^2 C_{wy}^2 - C_y^2) + \theta\left\{[(\beta-\alpha)^2 + 2(\alpha-1)(\alpha\gamma^2-\beta\gamma)]C_{wx}^2 - \frac{C_x^2}{4}\right\} + \theta([2(\alpha-1)(\beta-\alpha\gamma) + 2\alpha(\beta-\alpha)]\rho_{wxwy}C_{wx}C_{wy} - \rho_{xy}C_xC_y)\right\} < 0 \quad (8)$$

(c) Proposed estimator performs better than Thompson (1990) if,

$$MSE\left(\hat{\mu}^{p}_{wy(k,k')}\right) < MSE\left(\hat{\mu}_{yT}\right)$$

$$\{(\alpha-1)^2 + (\alpha^2-1)\theta C_{wy}^2 + [(\beta-\alpha)^2 + 2(\alpha-1)(\alpha\gamma^2-\beta\gamma)]\theta C_{wy}^2 + [2(\alpha-1)(\beta-\alpha\gamma) + 2\alpha(\beta-\alpha)]\theta\rho_{wxwy}C_{wx}C_{wy}\} < 0 \quad (9)$$

(d) Proposed estimator performs better than Shahzad and Hanif (2016) ($\hat{\mu}_{SH}$) if,

$$MSE\left(\hat{\mu}^{p}_{wy(k,k')}\right) < MSE\left(\hat{\mu}_{ySH}\right)$$

$$\{(\alpha-1)^2 + (\alpha^2-1)\theta C_{wy}^2 + \{[(\beta-\alpha)^2 + 2(\alpha-1)(\alpha\gamma^2-\beta\gamma)] - 1\}\theta C_{wx}^2 + \{[2(\alpha-1)(\beta-\alpha\gamma) + 2\alpha(\beta-\alpha)] - 1\}\theta\rho_{wxwy}C_{wx}C_{wy}\} < 0 \quad (10)$$

(e) Proposed estimator performs better than Shahzad and Hanif (2016) ($\hat{\mu}_{SH}{}'$) if,

$$MSE\left(\hat{\mu}^{p}_{wy(k,k')}\right) < MSE\left(\hat{\mu}_{ySH'}\right)$$

$$\left\{(\alpha-1)^2 + (\alpha^2-1)\theta C_{wy}^2 + \left\{[(\beta-\alpha)^2 + 2(\alpha-1)(\alpha\gamma^2-\beta\gamma)] - \frac{1}{4}\right\}\theta C_{wx}^2 + \{[2(\alpha-1)(\beta-\alpha\gamma) + 2\alpha(\beta-\alpha)] - 2\}\theta\rho_{wxwy}C_{wx}C_{wy}\right\} < 0 \quad (11)$$

Which is true and is therefore proposed estimator is theoretically efficient than the existing estimators taken in literature.

Numerical study:

Table 1: Data statistics:

N=50	$\mu_y = 282.42$	$C_y = 6.825$	$C_{wy} = 2.21$	$\rho_{xy} = -0.442$	$\bar{w}_y = 473.232$
n=5	$\mu_x = 0.900$	$C_x = 0.337$	$C_{wx} = 0.425$	$\rho_{wxwy} = -0.753$	$\bar{w}_x = 0.766$

Table 2: MSE of existing estimators:

$MSE(\hat{\mu}_{yR}) = 1800317$	$MSE(\hat{\mu}_{yBT}) = 654570$	$var(\hat{\mu}_{yT}) = 196882$	$MSE(\hat{\mu}_{ySH}) = 52406$	$MSE(\hat{\mu}_{ySH'}) = 60615$

Table 3: MSE of proposed estimator $\hat{\mu}^{p}_{wy(k,k')}$ at different values of K and K':

$\hat{\mu}^{p}_{wy(k,k')}$	α	β	γ	MSE
$\hat{\mu}^{p}_{wy(-1,-2')}$	0.6667	1	2	32393.44
$\hat{\mu}^{p}_{wy(-2,-3')}$	0.75	1	1.5	39904.68
$\hat{\mu}^{p}_{wy(-3,-4')}$	0.8	1	1.333333	44559.32
$\hat{\mu}^{p}_{wy(-4,-5')}$	0.8333	1	1.25	47976.12
$\hat{\mu}^{p}_{wy(-5,-6')}$	0.8571	1	1.2	50603.15

The table 3 reveals that the MSE of the proposed estimator at $k = -1, k' = -2$ is lowest and is considered to be best for comparison with existing estimators.

Table 4: Percentage relative efficiency (PRE) of the proposed estimator with respect to existing estimators (EE), where the MSE of proposed estimator (PE1) =32393, (PE2) = 39905, (PE2) = 44559, (PE2) = 47976, (PE2) = 50603.

Existing Estimator (EE)	MSE of EE	PRE of PE1	PRE of PE2	PRE of PE3	PRE of PE4	PRE of PE5
$\hat{\mu}_{yR} = \mu_y \frac{\bar{x}}{\hat{\mu}_x}$	1800317	5557.659	4511.544	4040.270	3752.527	3557.717
$\hat{\mu}_{yBT} = \mu_y \left(\frac{\bar{x} - \mu_x}{\bar{x} + \mu_x}\right)$	654570	2020.687	1640.334	1468.986	1364.366	1293.536
$\hat{\mu}_{yT} = n^{-1} \sum_{i=1}^{n} w_{yi}$	196882	607.783	493.380	441.842	410.375	389.070
$\hat{\mu}_{SH} = \bar{w}_y \frac{\bar{w}_x}{\hat{\mu}_{wx}}$	52406	161.779	131.328	117.609	109.233	103.562
$\hat{\mu}_{SH'} = \bar{w}_y \exp\left(\frac{\bar{w}_x - \mu_x}{\bar{w}_x - \mu_x}\right)$	60615	187.121	151.899	136.032	126.344	119.785

$$PRE = \frac{MSE\ of EE}{MSE\ of\ PE} \times 100$$

The table 4 shows that the proposed estimator performs better than the existing estimators and will be used for estimating the population mean of rare and clustered population.

5. Conclusion:

ACS is the best sampling design while dealing with the rare and clustered population While using the product type estimator the network level negative correlation is important rather than a unit level negative correlation, because unit level negative correlation doesn't guarantee the network level negative correlation. This happens when the study and auxiliary variable are found together irrespective of their negative correlation at unit level. ACS design is preferable only when there is a negative correlation at network level, known population mean of auxiliary variable Keeping all these assumptions in consideration we concluded that the proposed estimator at all values of k and k' performs well as compared to the existing estimators of both conventional and ACS, but is most efficient at k=-1, k'=-2 and should be preferred.

References:

[1] Sisodia, B.V.S. and Dwivedi, V.K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. J. Indian Soc. Agric. Statist., 33, 13-18.

[2] Bahl, S. and Tuteja, R. K. (1991). Ratio and product type exponential estimator. Information and Optimization Sciences, 12, 159-163.

[3] Bennitt, E., Bonyongo, M. C., and Harris, S. (2014). Habitat selection by African buffalo (syncerus caffer) in response to landscape-level fluctuations in water availability on two temporal scales. PLOS ONE, 9(7: e101346),1–14.

[4] Kadilar, C. and Cingi, H. (2003). A study on the chain ratio-type estimator. Hacettepe Journal of Mathematics and Statistics,32, 105-108.

[5] Chaudhry, M. S. and Hanif, M. (2016). Product and Exponential Product Estimators in Adaptive Cluster Sampling under different Population Situations. Proceedings of the Pakistan Academy of Sciences: Pakistan Academy of Sciences, A. Physical and Computational Sciences, 53 (4), 447–457.

[6] Cochran, W. G. (1940). The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce. The Journal of Agricultural Science, 30 (2), 262–75.

[7] Gattone, S. A., Mohamed, E., Dryver, A. L. and Munnich, R. T. (2016). Adaptive cluster sampling for negatively correlated data. Environmetrics, 27(2), 103–113.

[8] Singh, H. P. and Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean. Statistics in Transition, 6, 555-560.

[9] Hussain, S., Sharma M., and Chandra, H. (2021), Optimum exponential ratio type estimator for estimating the population mean. J. Stat. Appl. Prob. Lett., 8(2), 73-82.

[10] Jeelani, M. I., Rizvi, S.E.H., Sharma, M. K., Mir, S. A., Raja, S.A., Maqbool, S. A., Kumar, B., Sharma, M. K., Rizvi, S.E. H. and Jeelani, M. I. (2018). Estimation of Population Mean Through Improved Class Of Ratio Type Estimators Using Auxiliary Information. Journal of Reliability and Statistical Studies, 11(1), 01-07.

[11] Nazir, N., & Jeelani, F. (2017). Modified Ratio Estimator under Rank Set Sampling. Revista Investigacion Operacional, 38(1), 103-106, DOI:10.15740/HAS/IRJAES/8.1/100-103.

[12] Prins, H. H. T. (1996). Ecology and Behaviour of the African Buffalo: Social Inequality and Decision Making. Chapman & Hall, London.

[13] Robson, D. S.)1957). Application of multivariate polykays to the theory of unbiased ratio-type estimation. Journal of the American Statistical Association, 52 (280), 511–22.

[14] Sharma, M. & Bhatnagar, S. (2008). A General Class Of Estimator For The Estimation Of Population Mean Using Auxiliary Information. Journal of the Indian Statistical Association, 46(2), 155-161.

[15] Thompson, S. K. 1990. Adaptive Cluster Sampling. Journal of the American Statistical Association, 85(412), 1050–9.

[16] Upadhyaya, L. N., Singh, H. P., Chatterjee, S. and Yadav, R. (2011). Improved ratio and product exponential type estimators. Journal of Statistics Theory and Practice, 5(2), 285-302.

[17] Upadhyaya, N. and Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. Biometri. J., 41, 627-636.

[18] Yan and Tian, B. (2010). Ratio Method to the Mean Estimation Using Co-efficient of Skewness of Auxiliary Variable', ICICA 2010, Part II, CCIS 106, 103-110.