ORIGINAL PAPER

EFFICIENT PREDICTIVE ESTIMATOR FOR FINITE POPULATION MEAN USING SUPPLEMENTARY VARIABLE UNDER POST STRATIFICATION

M. YOUNIS SHAH[1,*], SYED EJAZ HUSAIN RIZVI[1], MANISH KR. SHARMA[1]
M. IQBAL JEELANI[2]

Manuscript received: 20.01.2023; Accepted paper: 01.06.2023;
Published online: 30.06.2023.

Abstract. *In the current investigation, we have developed the efficient predictive estimator for finite population mean using auxiliary variable in case of post stratification. Up to the first order of approximation, the expressions for bias and mean square error (MSE) are derived for the proposed estimator. This also reveals the constant's ideal value, which reduces the MSE of the developed estimator. The developed estimator performs better than the existing estimators.Numerical study is also carried out by using the real data sets.*
Keywords: *post stratification; auxiliary variable; bias; mean square error; efficiency.*

1. INTRODUCTION

The application of stratified random sampling assures that the sizes and structure of sampling frames for each stratum are already defined. Whereas the total population size and the percentage of the unit that belongs to each stratum may be known in many existing system, it is possible that the sample frame for every stratum is neither available or would be costly and difficult to construct. We can't employ stratified random sampling in such types of situations.

In order to resolve these difficulties, post stratification technique is applied, in which a sample of necessary size is first selected from the population employing SRS, and it is then stratified using the stratification factor.

Initially, the post stratification idea was explained by Hansen et al. [1]. The classic Cochran [2] ratio estimator was later investigated by Ige and Tripathi [3] in the problem of post stratification.

In the area of post stratification, contributions have been made by Jagers et al. [4], Jagers [5], Agrawaland Panda [6], and Singh and Ruiz Espezo [7]. The characteristics of the post stratification product and ratio type exponential estimators by Bahl and Tuteja [8] and Tailor et al. [9] were recently specified by Rather et al. [10].

[1] Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Faculty of Basic Sciences, Division of Statistics & Computer Science, 180009 Jammu, Jammu and Kashmir, India.
[2] Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Faculty of Forestry, Division of Social and Basic Science, 180009 Jammu, Jammu and Kashmir, India.
** Corresponding author: syeedunis121@gmail.com.*

2. MATERIALS AND METHODS

Consider a population of size N that is finite and partitioned into L stratum of sizes $N_1, N_2, \ldots, N_L$ such that $\sum_{h=1}^{L} N_h = N$. Suppose y serve as study variable and x serve as auxiliary variable, which should be positively and negatively connected with y, respectively. The observation on the i^{th} unit of the h^{th} stratum for study variable will be y_{hi}, and the observation on the i^{th} unit of the h^{th} stratum for auxiliary variable will be $x_{hi.}$ Population means for study variables y and x are represented by $\bar{Y}$ and $\bar{X}$, respectively, where h^{th} stratum means are represented by $\bar{Y}_h$ and $\bar{X}_h$ respectively. Through the use of SRSWOR, a sample of size n is taken from the entire population. Following the SRS selection from the population, it is specified that how much and which units belong to the h^{th} stratum. Let n_h be the size of the sample falling in h^{th} stratum such that $\sum_{h=1}^{L} n_h = n$ here, it is assumed that n is so large that probability of n_h being zero is very small.

In case of post stratification, the usual unbiased estimator is defined as.

$$\bar{Y}_{ps} = \sum_{h=1}^{L} w_h \bar{y}_h \tag{2.1}$$

Where $\bar{y}_h$ is the sample mean of the n_h sample units that lie in the h^{th} stratum and $w_h = \frac{N_h}{N}$ is the weight of the h^{th} stratum. Using Stephen's [11] findings, the variance is represented as

$$V\left(\bar{Y}_{ps}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} w_h S_{yh}^2 + \frac{1}{n^2}\sum_{h=1}^{L}\left(1 - w_h\right) S_{yh}^2 \tag{2.2}$$

Ige and Tripathi [3] defined ratio type estimator in post stratification as

$$\bar{Y}_R = \bar{y}_{PS}\left(\frac{\bar{X}}{\bar{x}_{PS}}\right) \tag{2.3}$$

where

$$\bar{X} = \sum_{h=1}^{L} w_h X_h, \quad \bar{y}_{ps} = \sum_{h=1}^{L} w_h \bar{y}_h, \ and \ \bar{x}_{ps} = \sum_{h=1}^{L} w_h x_h,$$

Bias and MSE expressions of $\bar{Y}_R$ up to the fda is given follows

$$B\left(\bar{Y}_R\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{\bar{X}}\sum_{h=1}^{L} w_h\left(R_1 S_{xh}^2 - S_{yxh}\right) \tag{2.4}$$

$$MSE\left(\bar{Y}_R\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} w_h\left(S_{yh}^2 + R_1^2 S_{xh}^2 - 2R_1 S_{yxh}\right) \tag{2.5}$$

where $R_1 = \frac{\bar{Y}}{\bar{X}}$.

Tailor et al. [12] defined the ratio type exponential estimator from Bahl and Tuteja [8] as follows

$$\left(\overline{Y}_{RC}\right)=\overline{y}_{ps}\exp\left(\frac{\overline{x}_{ps}-\overline{X}}{\overline{x}_{ps}+\overline{X}}\right) \tag{2.6}$$

Bias and MSE expressions of $\overline{Y}_{RC}$ up to the fda is given follows

$$B\left(\overline{Y}_{RC}\right)=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{1}{\overline{X}}\sum_{h=1}^{L}w_h\left(\frac{3}{8}R_1S_{xh}^2+\frac{1}{2}S_{yxh}\right) \tag{2.7}$$

$$MSE\left(\overline{Y}_{Rc}\right)=\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{h=1}^{L}w_h\left(S_{yh}^2+\frac{1}{4}R_1^2S_{xh}^2-R_1S_{yxh}\right) \tag{2.8}$$

$$\left(\overline{Y}_{PC}\right)=\overline{y}_{ps}\exp\left(\frac{\overline{X}-\overline{x}_{ps}}{\overline{X}+\overline{x}_{ps}}\right) \tag{2.9}$$

Bias and MSE expressions of $\overline{Y}_{PC}$ up to the fda is given follows

$$B\left(\overline{Y}_{PC}\right)=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{1}{\overline{X}}\sum_{h=1}^{L}w_h\left(\frac{3}{8}R_1S_{xh}^2-\frac{1}{2}S_{yxh}\right) \tag{2.10}$$

$$MSE\left(\overline{Y}_{Pc}\right)=\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{h=1}^{L}w_h\left(S_{yh}^2+\frac{1}{4}R_1^2S_{xh}^2+R_1S_{yxh}\right) \tag{2.11}$$

3. PROPOSED ESTIMATOR

As a result of Yadav and Mishra [13] inspiration, we have developed the improved ratio cum product type exponential estimator in predictive estimation approach under post-stratification as,

$$\left(\overline{Y}_{YS}\right)=\left[\eta P_{RE}+(1-\eta)P_{PE}\right] \tag{3.1}$$

where η is an unknown constant whose value is to be estimated later, P_{RE} and P_{PE} are the estimators, which are defined as

$$\begin{aligned}P_{RE}&=\left[\frac{n}{N}\overline{Y}_h+\left(\frac{N-n}{N}\right)\overline{Y}_h\exp\left(\frac{\overline{X}_h-\overline{x}_h}{\overline{X}_h-\overline{x}_h}\right)\right]\\&=\left[\frac{n}{N}\overline{Y}_h+\left(\frac{N-n}{N}\right)\overline{Y}_h\exp\left(\frac{N\left(\overline{X}_h-\overline{x}_h\right)}{N\left(\overline{X}_h-\overline{x}_h\right)-2n\overline{x}_h}\right)\right]\end{aligned} \tag{3.2}$$

and

$$P_{PE} = \left[\frac{n}{N}\overline{Y}_h + \left(\frac{N-n}{N}\right)\overline{Y}_h \exp\left(\frac{\overline{X}_h - \overline{x}_h}{\overline{X}_h - \overline{x}_h}\right)\right]$$
$$= \left[\frac{n}{N}\overline{Y}_h + \left(\frac{N-n}{N}\right)\overline{Y}_h \exp\left(\frac{N\left(\overline{X}_h - \overline{x}_h\right)}{N\overline{X}_h + (N-2n)\overline{x}_h}\right)\right] \tag{3.3}$$

In order to obtain the Bias and MSE of($\overline{Y}_{YS}$). Let us suppose,

$$\varepsilon_0 = \frac{1}{\overline{Y}_h}\sum_{h=1}^{L} w_h \overline{y}_h, \ and\ \varepsilon_1 = \frac{1}{\overline{X}_h}\sum_{h=1}^{L} w_h \overline{x}_h,$$

Such that,

$$\varepsilon_0 = \varepsilon_1 = 0,\ E\left(\varepsilon_0^2\right) = \frac{1}{\overline{Y}_h^2}\left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} w_h S_{yh}^2 \quad ,$$

$$E\left(\varepsilon_1^2\right) = \frac{1}{\overline{X}_h^2}\left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} w_h S_{xh}^2,\ and\ E\left(\varepsilon_0\varepsilon_1\right) = \frac{1}{\overline{X}_h^2\overline{Y}_h^2}\left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} w_h S_{yxh}^2$$

Expressing (P_{RE}) in terms of errors, we have

$$P_{RE} = \overline{Y_h}\left(1+\varepsilon_0\right)\left[\frac{n}{N} + \frac{N-n}{N}\exp\left(\frac{N\varepsilon_1}{2(N-n)+(N-2n)\varepsilon_1}\right)\right]$$

$$P_{RE} = \overline{Y_h}\left(1+\varepsilon_0\right)\left[\gamma_h + \left(1-\gamma_h\right)\exp\left(-\frac{\varepsilon_1}{2\left(1-\gamma_h\right)+\left(1-2\gamma_h\right)\varepsilon_1}\right)\right] \tag{3.4}$$

$$P_{RE} = \overline{Y_h}\left(1+\varepsilon_0\right)\left[\gamma_h + \left(1-\gamma_h\right)\exp\left\{-\frac{\varepsilon_1}{2\left(1-\gamma_h\right)}\left(1+\frac{\left(1-2\gamma_h\right)}{2\left(1-\gamma_h\right)}\varepsilon_1\right)^{-1}\right\}\right]$$

After simplifying eq. (3.4) and neglecting the terms up to the first order of approximation, we have

$$P_{RE} = \overline{Y_h}\left[1+\varepsilon_0 - \frac{\varepsilon_1}{2} - \frac{\varepsilon_0\varepsilon_1}{2} + \frac{\varepsilon_1^2}{8}\left(3-4\gamma_h\right)\right] \tag{3.5}$$

Similarly expressing (P_{PE}) in terms of errors, we have

$$P_{PE} = \overline{Y_h}\left(1+\varepsilon_0\right)\left[\frac{n}{N} + \frac{N-n}{N}\exp\left(\frac{N\varepsilon_1}{2(N-n)+N\varepsilon_1}\right)\right] \tag{3.6}$$

$$P_{PE} = \overline{Y_h}(1+\varepsilon_0)\left[\gamma_h + (1-\gamma_h)\exp\left(-\frac{\varepsilon_1}{2(1-\gamma_h)+\varepsilon_1}\right)\right]$$

$$P_{PE} = \overline{Y_h}(1+\varepsilon_0)\left[\gamma_h + (1-\gamma_h)\exp\left\{\frac{\varepsilon_1}{2(1-\gamma_h)}\left(1+\frac{\varepsilon_1}{2(1-\gamma_h)}\right)^{-1}\right\}\right]$$

After simplifying eq. (3.6) and neglecting the terms up to the first order of approximation, we have

$$P_{PE} = \overline{Y_h}\left[1+\varepsilon_0+\frac{\varepsilon_1}{2}+\frac{\varepsilon_0\varepsilon_1}{2}-\frac{\varepsilon_1^2}{8(1-\gamma_h)}\right] \tag{3.7}$$

Now express $(\bar{Y}_{\mathrm{YS}})$from eq. (3.1) in terms errors by using eq. (3.5) and eq. (3.7), we have

$$(\overline{Y}_{YS}) = \overline{Y}_h\left[\begin{array}{l}\eta\left\{1++\varepsilon_0-\frac{\varepsilon_1}{2}-\frac{\varepsilon_0\varepsilon_1}{2}+\frac{\varepsilon_1^2}{8}(3-4\gamma_h)\right\}\\ +(1-\eta)\left\{1+\varepsilon_0+\frac{\varepsilon_1}{2}+\frac{\varepsilon_0\varepsilon_1}{2}-\frac{\varepsilon_1^2}{8(1-\gamma_h)}\right\}\end{array}\right]$$

$$(\overline{Y}_{YS}) = \overline{Y}_h\left[\begin{array}{l}1+\varepsilon_0-(2\eta-1)\frac{\varepsilon_1}{2}-(2\eta-1)\frac{\varepsilon_0\varepsilon_1}{2}+\\ \frac{\varepsilon_1^2}{8}\left\{\frac{4\eta-1-7\gamma_h+4\gamma_h^2}{1-\gamma_h}\right\}\end{array}\right] \tag{3.8}$$

Subtracting both sides of eq. (3.8) and the bias of proposed estimator $(\bar{Y}_{\mathrm{YS}})$ to terms of order n^{-1} can be obtained by taking expectation and using the values of errors as

$$Bias(\overline{Y}_{YS}) = \left[\begin{array}{l}\frac{1}{8}\left\{\frac{4\eta-1-7\gamma+4\gamma_h^2}{1-\gamma_h}\right\}\\ \sum_{i=1}^{h}W_h\frac{1}{\overline{X}_h}\left\{R_hS_{xh}^2-\frac{(2\eta-1)}{2}R_hS_{yxh}\right\}\end{array}\right] \tag{3.9}$$

We derive the MSE of developed estimator to terms of order n^{-1} by first squaring both sides of equation (3.8), followed by taking the expectation.

$$MSE(\overline{Y}_{YS}) = E(\overline{Y}_{YS}-\overline{Y}_h)^2$$

$$MSE\left(\overline{Y}_{YS}\right)=E\left[\overline{Y}_{h}\left\{\varepsilon_{0}(2\eta-1)\frac{\varepsilon_{1}}{2}\right\}\right]^{2}$$

$$MSE\left(\overline{Y}_{YS}\right)=E\left[\overline{Y}_{h}\left\{\varepsilon_{0}-\eta_{1}\frac{\varepsilon_{1}}{2}\right\}\right]^{2}$$

where $\eta_1=(2\eta-1)$

$$MSE\left(\overline{Y}_{YS}\right)=\overline{Y}_{h}^{2}\left[\left(E\left(\varepsilon_{0}^{2}\right)+\eta_{1}^{2}\frac{E\left(\varepsilon_{1}^{2}\right)}{4}-\eta_{1}E\left(\varepsilon_{0}\varepsilon_{1}\right)\right)\right]$$

$$MSE\left(\overline{Y}_{YS}\right)=\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{h=1}^{L}W_{h}\left[S_{yh}^{2}+\eta_{1}^{2}\frac{R_{h}^{2}S_{xh}^{2}}{4}-\eta_{1}R_{h}S_{yxh}\right] \tag{3.10}$$

The optimum value of η_1 is obtained by when $MSE(\bar{Y}_{\mathrm{YS}})$is differentiated with respect to η_1 and the derivative is equal to zero

$$\eta_{1Opt}=2\frac{\sum_{h=1}^{L}W_{h}S_{yxh}}{\sum_{h=1}^{L}W_{h}R_{1}S_{xh}^{2}}$$

The reduced form of $MSE\left(\bar{Y}_{YS(min)}\right)$ is obtained by Substituting the value of η_{1Opt} in (3.10),

$$MSE\left(\overline{Y}_{YS(\min)}\right)=\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=1}^{h}\left[S_{yh}^{2}+2\left(\frac{W_{h}S_{yxh}^{2}}{W_{h}S_{xh}^{2}}\right)\right] \tag{3.11}$$

4. EFFICIENCY COMPARISON

From (2.1), (2.5), (2.8), (2.11) and (3.11), it is concluded that the developed estimator would be much precise than other if

I. $\left[Var\left(\bar{Y}_{ps}\right)-MSE\left(\bar{Y}_{YS(min)}\right)\right]>0$

$$\frac{1}{n^{2}}\sum_{h=1}^{L}(1-w_{h})S_{yh}^{2}-\sum_{h=1}^{L}w_{h}\left(\frac{2S_{yxh}^{2}}{S_{xh}^{2}}\right)>0 \tag{4.1}$$

II. $\left[MSE\left(\bar{Y}_{R}\right)-MSE\left(\bar{Y}_{YS(min)}\right)\right]>0$

$$\sum_{h=1}^{L}w_{h}\left(R_{1}^{2}-2R_{1}S_{yxh}\right)-\sum_{h=1}^{L}w_{h}\left(\frac{2S_{yxh}^{2}}{S_{xh}^{2}}\right)>0 \tag{4.2}$$

III. $\left[MSE(\bar{Y}_{RC}) - MSE(\bar{Y}_{YS(min)})\right] > 0$

$$\sum_{h=1}^{L} w_h\left(\frac{1}{4}R_1^2 S_{xh}^2 - R_1 S_{yxh}\right) - \sum_{h=1}^{L} w_h\left(\frac{2S_{yxh}^2}{S_{xh}^2}\right) > 0 \tag{4.3}$$

IV. $\left[MSE(\bar{Y}_{PC}) - MSE(\bar{Y}_{YS(min)})\right] > 0$

$$\sum_{h=1}^{L} w_h\left(\frac{1}{4}R_1^2 S_{xh}^2 + R_1 S_{yxh}\right) - \sum_{h=1}^{L} w_h\left(\frac{2S_{yxh}^2}{S_{xh}^2}\right) > 0 \tag{4.4}$$

Because the requirements from (4.1) to (4.4) are always met, it is seen that $\bar{Y}_{YS(min)}$ is always more efficient than the classical estimators $\bar{Y}_{ps}$, $\bar{Y}_R$, $\bar{Y}_{RC}$ and $\bar{Y}_{PC}$.

5. EMPIRICAL STUDY

We will take into account three natural population datasets to evaluate the merits of the developed estimator. Below is a description of the populations:

Dataset I[14]: Y: Output; X: Fixed cost

Table 5.1. Statistics Description

Constants	N_h	n_h	$\bar{Y}_h$	$\bar{X}_h$	S_{yh}	S_{xh}	S_{yxh}
Stratum I	05	04	1925	214.40	615.92	74.87	36360.68
Stratum II	05	04	3115.60	333.80	340.38	66.35	22356.50

Dataset II[14]: Y: Output; X: Fixed capital

Table 5.2. Staistics Description

Constants	N_h	n_h	$\bar{Y}_h$	$\bar{X}_h$	S_{yh}	S_{xh}	S_{yxh}
Stratum I	05	02	1925	214.40	615.92	74.87	36360.68
Stratum II	05	02	3115.60	333.80	340.38	66.35	22356.50

Dataset III [15]: Y: Productivity (MT/ h); X: Production in kT

Table 5.3. Statistics Description

Constants	N_h	n_h	$\bar{Y}_h$	$\bar{X}_h$	S_{yh}	S_{xh}	S_{yxh}
Stratum I	10	04	1.70	10.41	0.50	3.53	1.60
Stratum II	10	04	3.67	289.14	1.41	111.61	144.87

Table 5.4. Competences of estimators and suggested estimator with respect to $V(\bar{y}_{PS})$

Estimators	**Population I**		**Population II**		**Population III**	
	MSE	**PRE**	**MSE**	**PRE**	**MSE**	**PRE**
$V(\bar{y}_{PS})$	10059.10	100	52617.10	100	0.08524	100
$\bar{Y}_R$	2580.75	389.75	15483.10	339.83	0.037271	228.71
$\bar{Y}_{RC}$	15927.19	63.16	95578.40	55.10	0.023031	370.10
$\bar{Y}_{PC}$	1741.38	577.65	10443.11	503.86	0.219963	38.75
$\bar{Y}_{YS}$	**1433.430**	**701.70**	**8595.925**	**612.16**	**0.01940**	**439.16**

Table 5.4, revealed the percent relative proficiencies (PRE) of estimators for population I-III. Here we see that the improved ratio cum product type exponential estimator in predictive estimation approach under post-stratification in the sense of having highest percent relative efficiency than usual unbiased estimators

6. CONCLUSIONS

Section 4 gives the situations under which the developed estimator $\bar{Y}_{YS}$has more efficiency as compared to existed estimators. Expressions (4.1), (4.2), (4.3) and (4.4) are conditions under which the developed estimator would have less MSE in comparison to$\bar{Y}_{ps}$, $\bar{Y}_{R}$, $\bar{\mathrm{Y}}_{\mathrm{RC}}$and $\bar{Y}_{S}$.

Finally, from the outcome of numerical analysis and theoretical exchange of views, it can be seen that the recommended estimator is highly precise as compared to the previous available estimators, for obtaining the population mean of characteristic variable under optimal conditions. It is clear that the recommended estimator is preferred than the previous available estimators as shown in the table 5.4.

REFERENCES

[1] Hansen, M.H., Hurwitz, W.N. and Madow, W.G. 1953. *Sample Survey Methods and Theory*, Vol. 1, John Wiley and Sons, New York, 1953.
[2] Cochran, W.G., *Sampling techniques*, 3rd Ed., Wiley Eastern Limited, 1977.
[3] Ige, A.F., Tripathi, T.P., *Abacus*, **18**, 265, 1989.
[4] Jagers, P., Oden, A., Trulsson, L., *International Statistical Review*, **53**, 221, 1985.
[5] Jagers, P., *International Statistical Review*, **55**, 159, 1986.
[6] Agrawal, M.C. Panda, K.B., *Metron*, **51**, 179, 1993.
[7] Singh, H.P., Ruiz Espejo, M.*Bulletin of the International Statistical Institute*, **54**(2), 341, 2003.
[8] Bahl, S. Tuteja, R.K., *Journal of Information and Optimization Sciences*, **12**, 159, 1991.
[9] Tailor, R., Chouhan, S., *Communications in Statistics - Theory and Methods*, **46**(21), 10387, 2017.
[10] Rather, K.U.I., Jeelani, M.I., Shah, M.Y., Rizvi, S.E.H., Sharma, M., *Journal of Applied Mathematics Statistics and Informatics*, **18**(1), 29, 2022.
[11] Stephan, F., *Annals of Mathematical* Statistics, **16**, 50, 1945.
[12] Tailor R., Mehta, P., *Journal of Statistics Applications & Probability*, **8**(3), 241, 2019.
[13] Yadav, K.S., Mishra, S.S., *Statiska*, **95** (1), 2015.
[14] Murthy, M.N., *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta, 1967.
[15] http://nhb.gov.in/statistics/area-production-statistics.html; official website of National Horticulture Board, India.