Pak. J. Statist.
2015 Vol. 31(4), 353-362

A CLASS OF TRANSFORMED EFFICIENT RATIO ESTIMATORS OF FINITE POPULATION MEAN

Sajjad Ahmad Khan[1§], Hameed Ali[1], Sadaf Manzoor[1] and **Alamgir[2]**
[1] Department of Statistics, Islamia College, Peshawar, Pakistan
[2] Department of Statistics, University of Peshawar, Pakistan
[§] Corresponding author Email: sajjadkhan@icp.edu.pk

ABSTRACT

In this paper we have proposed a class of transformed efficient estimators of finite population mean by modifying Mohanty and Sahoo (1995) transformation. The approximate bias and MSE of the proposed class of estimators have been found. Conditions under which the proposed class of transformed estimators performs good, have been obtained. Examples are given based on real life data to illustrate the results.

KEYWORDS

Auxiliary Information, Transformation, Efficiency, Mean Square Error.

1. INTRODUCTION

In this paper we have proposed an efficient estimator for estimating the mean of the finite population under simple random sampling without replacement sampling and stratified random sampling schemes. Motivated by Subramani and Kunarpendyani (2012), Bedi (1996), Mohanty and Sahoo (1995), Shabbir and Gupta (2011), we have made use of auxiliary information by introducing some transformation.

The Classical estimator of the mean of the finite population $\bar{Y}$ is $\bar{y}$. Variance of $\bar{y}$ is

$$V\left(\bar{y}\right) = \lambda C_y^2, \text{ where } \lambda = \frac{1}{n} - \frac{1}{N} \tag{1.0}$$

Cochran (1940) introduced the traditional ratio type estimator of population mean, as

$$\bar{y}_{cr} = \left(\frac{\bar{y}}{\bar{x}}\right)\bar{X} \tag{1.1}$$

with the following bias and MSE

$$B\left(\bar{y}_{cr}\right) \cong \lambda \bar{Y}\left(C_x^2 - C_{yx}\right) \tag{1.2}$$

and

$$MSE\left(\bar{y}_{cr}\right) \cong \lambda \bar{Y}^2\left(C_y^2 + C_x^2 - 2\rho C_y C_x\right). \tag{1.3}$$

where $C_y^2 = \frac{S_y^2}{\bar{Y}^2}$ is the coefficient of variation of variable Y, $C_x^2 = \frac{S_x^2}{\bar{X}^2}$ is the coefficient of variation of X, $C_{yx} = \frac{S_{yx}}{\bar{Y}\bar{X}}$ is the coefficient of covariance of X and Y and $\rho = \frac{C_{yx}}{C_y C_x}$ is the coefficient of correlation between Y and X.

The classical regression estimator given by Hansen and Harwitz (1943) is

$$\bar{y}_{lr} = \bar{y} + b_{yx}\left(\bar{X} - \bar{x}\right) \tag{1.4}$$

with the following mean square error

$$MSE\left(\bar{y}_{lr}\right) \cong \bar{Y}^2 \lambda C_y^2 \left(1 - \rho^2\right) \tag{1.5}$$

This is a well-established fact that regression estimator is efficient when the linear relationship passes in the neighborhood of origin.

An exponential ratio type estimator due to Bhal and Tuteja (1991) is given by

$$\bar{y}_{BT} = \bar{y}\exp\left(\frac{\left(\bar{X} - \bar{x}\right)}{\left(\bar{X} + \bar{x}\right)}\right) \tag{1.6}$$

The bias and MSE due to $\bar{y}_{BT}$ is given by

$$B\left(\bar{y}_{BT}\right) = \lambda \bar{Y}\left(\frac{3C_x^2}{8} - \frac{C_{yx}}{2}\right) \tag{1.7}$$

and

$$MSE\left(\bar{y}_{BT}\right) = \lambda \bar{Y}^2\left(C_y^2 + \frac{C_x^2}{4} - C_{yx}\right) \tag{1.8}$$

Mohanty and Sahoo (1995), introduces the following transformation in ratio estimator of population mean

$$t_1 = \frac{\bar{y}}{\bar{z}}\bar{Z}$$

and

$$t_2 = \frac{\bar{y}}{\bar{u}}\bar{U}$$

where $z_i = \frac{x_i + x_m}{x_M + x_m}$ and $u_i = \frac{x_i + x_M}{x_M + x_m}$.

x_m is the minimum and x_M is the maximum value of auxiliary variable with in the data.

The biases and MSEs of t_1 and t_2 are given by

$$B(t_1) = \lambda \bar{Y}\left(S_z^2 - S_{yz}\right). \tag{1.9}$$

$$B(t_1) = \lambda \bar{Y}\left(S_u^2 - S_{yu}\right). \tag{1.10}$$

$$MSE(t_1) = \lambda \bar{Y}^2\left(C_y^2 + C_z^2 - 2C_{yz}\right). \tag{1.11}$$

$$MSE(t_2) = \lambda \bar{Y}^2\left(C_y^2 + C_u^2 - 2C_{yu}\right). \tag{1.12}$$

where $C_z^2 = \frac{C_x^2}{c_1^2}, C_u^2 = \frac{C_x^2}{c_2^2}, C_{yz} = \frac{C_{yx}}{c_1}$ and $c_1 = 1 + \frac{x_m}{\bar{X}}, c_2 = 1 + \frac{x_M}{\bar{X}}$

Shabbir and Gupta (2011) proposed the following transformed exponential ratio type of estimator

$$\bar{y}_{SG} = \left[\left\{\tau_1 \bar{y} + \tau_2\left(\bar{X} - \bar{x}\right)\right\}\left\{\exp\left(\frac{\bar{A} - \bar{a}}{\bar{A} + \bar{a}}\right)\right\}\right] \tag{1.13}$$

where $\bar{a} = \bar{x} + N\bar{X}$ and $\bar{A} = \bar{X} + N\bar{X}$

With the following MSE

$$MSE(\bar{y}_{SG}) = \bar{Y}^2\left[\left\{1 - \lambda \frac{C_x^2}{4(N+1)^2}\right\} - \frac{\left(1 - \lambda \frac{C_x^2}{8(N+1)^2}\right)^2}{\left\{1 + \lambda C_y^2\left(1 - \rho^2\right)\right\}}\right] \tag{1.14}$$

where the optimum value of τ_1 and τ_2 is given by

$$\tau_1 = \frac{\left\{1 - \lambda \frac{C_x^2}{8(N+1)^2}\right\}}{\left\{1 + \lambda C_y^2\left(1 - \rho^2\right)\right\}} \text{ and } \tau_2 = \frac{\bar{Y}}{\bar{X}}\left\{\frac{1}{2(N+1)} - \tau_1\left(\frac{1}{N+1} - \rho \frac{C_y}{C_x}\right)\right\}.$$

2. SUGGESTED TRANSFORMATION

Mohanty and Sahoo (1995) suggested the following transformation of the auxiliary variable

$$\nu_i = \frac{x_i + \eta}{\eta + \pi} \text{ and } \upsilon_i = \frac{x_i + \pi}{\eta + \pi} \text{ or } \bar{\nu} = \frac{\bar{x} + \eta}{\eta + \pi} \text{ and } \bar{\upsilon} = \frac{\bar{x} + \pi}{\eta + \pi} \tag{2.1}$$

where $\eta = x_m$ and $\pi = x_M$.

Further we can write $\mathrm{E}(\bar{\nu}) = \bar{V} = \dfrac{\bar{X} + \eta}{\eta + \pi}$ and $\mathrm{E}(\bar{\upsilon}) = \bar{U} = \dfrac{\bar{X} + \pi}{\eta + \pi}$.

Further we write $\mathrm{E}\left(\dfrac{\bar{\nu} - \bar{V}}{\bar{V}}\right) = \left(\dfrac{1-f}{n}\right)\dfrac{C_x^2}{c_1^2}$ and $\mathrm{E}\left(\dfrac{\bar{\upsilon} - \bar{U}}{\bar{U}}\right) = \left(\dfrac{1-f}{n}\right)\dfrac{C_x^2}{c_2^2}$,

where $c_1 = 1 + \dfrac{\eta}{\bar{X}}$ and $c_2 = 1 + \dfrac{\lambda}{\bar{X}}$. (3.2)

Obviously, for the above transformation one has to find out two features of auxiliary variable, such as η and λ for estimator of finite population mean. One of the disadvantage of the above transformation is that, only one constant are taking part in the reduction of MSE while the other has no contribution at all. In order to remove the extra feature from estimator and to incorporate only those feature which are taking contribution in the reduction of MSE, we precede as following.

We write (2.2) as

$$c = 1 + \frac{\eta}{\bar{X}} \text{ or } \eta = \bar{X}(c-1)$$

where η is constant or some function of auxiliary information.

The proposed transformation is

$$x^*_i = x_i + \eta, \text{ or } \bar{x}^* = \bar{x} + \eta \quad \bar{x}^* = \bar{x} + \bar{X}(c-1), \text{ and } E(\bar{x}^*) = \bar{X}, \text{ where } \bar{X}^* = c\bar{X}.$$

3. PROPOSED ESTIMATOR IN SIMPLE RANDOM SAMPLING

Following Bhal and Tuteja (1991) and Shabbir and Gupta (2011), we suggest the following class of estimators

$$\bar{y}_{prk} = \left\{\omega_1 \bar{y} + \omega_2 \left(\bar{X} - \bar{x}\right)\right\} \exp\left[\frac{\bar{X}^* - \bar{x}^*}{\bar{X}^* + \bar{x}^*}\right], k = 1, 2, 3, 4 \tag{3.1}$$

where $\bar{X}^* = \bar{X}c_k$ and $\bar{x}^* = \bar{x} + \bar{X}(c_k - 1)$, or $E(\bar{x}^*) = \bar{X}^*$.

Further

$$c_1 = C_x, c_2 = \rho + 1, c_3 = \frac{\rho + 1}{2}, \; c_4 = \frac{\rho + 1}{3}. \; k = 1, 2, 3, 4 .$$

where $\omega_1 \,\&\, \omega_2$ are suitable constant to be determined so that MSE is minimum, where "c_k" is to be chosen, so that the transformed data results high gain in efficiency. We will use the following terms to find the Bias and MSE due to our proposed estimator

$$e_0 = \left(\frac{\bar{Y} - \bar{y}}{\bar{Y}}\right), e_1 = \left(\frac{\bar{X} - \bar{x}}{\bar{X}}\right) \text{ or } \frac{e_1}{c_k} = \frac{\bar{x}^* - \bar{X}^*}{\bar{X}^*}$$

such that

$$E\left(e_1\right) = E\left(e_0\right) = 0,\ E\left(e_0^2\right) = \lambda C_y^2, E\left(e_1^2\right) = \lambda C_x^2$$

and

$$E\left(e_0 e_1\right) = \lambda C_{yx} \text{ or } E\left(e_0 e_1\right) = \lambda \rho C_y C_x .$$

Then we can write (3.1) as follows

$$\bar{y}_{pri} = \left\{\omega_1 \bar{Y}\left(1 + e_0\right) - \bar{X}\omega_2 e_1\right\} \exp\left\{\frac{-e_1}{2c_k}\left(1 + \frac{e_1}{2c_k}\right)^{-1}\right\}$$

or

$$\bar{y}_{pri} = \left\{\omega_1 \bar{Y}\left(1 + e_0\right) - \bar{X}\omega_2 e_1\right\}\left\{1 - \frac{e_1}{2c_k} + \frac{3e_1^2}{8{c_k}^2} + \cdots\right\} \tag{3.2}$$

Terms with power higher than two can be ignored, we have

$$\bar{y}_{pri} - \bar{Y} \cong \bar{Y}\left\{\left(\omega_1 - 1\right) + \omega_1\left(e_0 - \frac{e_1}{2c_k} - \frac{e_0 e_1}{2c_k} + \frac{3e_1^2}{8{c_k}^2}\right)\right\} - \omega_2 \bar{X}\left\{e_1 - \frac{e_1^2}{2{c_k}^2}\right\} \tag{3.3}$$

We can define the Bias due to our proposed estimator by

$$Bias\left(\bar{y}_{prk}\right) \cong E\left(\bar{y}_{prk} - \bar{Y}\right)$$

or

$$Bias\left(\bar{y}_{pri}\right) \cong \bar{Y}\left\{\left(\omega_1 - 1\right) + \omega_1 \lambda\left(\frac{3C_x^2}{8{c_k}^2} - \frac{C_{yx}}{2c_k}\right)\right\} + \omega_2 \bar{X}\lambda\left(\frac{C_x^2}{2{t_i}^2}\right) \tag{3.4}$$

For MSE, Squaring and taking expectation of equation (3.3), we have

$$E\left(\bar{y}_{prk} - \bar{Y}\right)^2 \cong \left[\left\{\bar{Y}\left(\omega_1 - 1\right) + \omega_1 \bar{Y}\left(e_0 - \frac{e_1}{2c_k} - \frac{e_0 e_1}{2c_k} + \frac{3e_1^2}{8{c_k}^2}\right)\right\} - \omega_2 \bar{X}\left\{e_1 - \frac{3e_1^2}{2c_k}\right\}\right]^2 \tag{3.5}$$

Since

$$MSE\left(\bar{y}_{prk}\right) \cong E\left(\bar{y}_{prk} - \bar{Y}\right)^2$$

$$MSE\left(\bar{y}_{prk}\right) \cong \bar{Y}^2\left[\left(\omega_1-1\right)^2+\omega_1^2\lambda\left\{C_y^2+\frac{C_x^2}{{c_k}^2}-\frac{2C_{yx}}{c_k}\right\}-2\lambda\omega_1\left\{\frac{3C_x^2}{8{c_k}^2}-\frac{C_{yx}}{2c_k}\right\}\right]$$
$$+\omega_2^2\bar{X}^2\lambda C_x^2-\omega_2\lambda\overline{YX}\,\frac{C_x^2}{c_k}+2\omega_1\omega_2\lambda\overline{YX}\left\{\frac{C_x^2}{c_k}-C_{yx}\right\} \qquad (3.6)$$

The optimum values of $\omega_1\,\&\,\omega_2$ can be obtained by differentiating MSE of $\bar{y}_{prk}$ with respect to the $\omega_1\,\&\,\omega_2$ by

Differentiating (3.6) w.r.to ω_1 and equating to zero $\frac{\partial MSE\left(\bar{y}_{prk}\right)}{\partial\omega_1}=0$ we get

$$\bar{Y}^2\left[2\left(\omega_1-1\right)+2\omega_1\lambda\left\{C_y^2+\frac{C_x^2}{{c_k}^2}-\frac{2C_{yx}}{c_k}\right\}-2\lambda\left\{\frac{3C_x^2}{8{c_k}^2}-\frac{3C_{yx}}{2c_k}\right\}\right]$$
$$+2\omega_2\lambda\overline{YX}\left\{\frac{3C_x^2}{c_k}-C_{yx}\right\}=0 \qquad (3.7)$$

Differentiating (3.6) w.r.to ω_2 and equating to zero, we get $\frac{\partial MSE\left(\bar{y}_{prk}\right)}{\partial\omega_2}=0$ we get

$$2\omega_2\bar{X}^2\lambda C_x^2-\lambda\overline{YX}\,\frac{C_x^2}{c_k}+2\omega_1\lambda\overline{YX}\left(\frac{C_x^2}{c_k}-C_{yx}\right)=0 \qquad (3.8)$$

Solving equation (3.7) & (3.8) simultaneously to get the optimum values of $\omega_1\,\&\,\omega_2$. After simplification we have

$$\omega_{1opt}=\frac{1-\frac{\lambda C_x^2}{8{c_k}^2}}{1+\lambda C_y^2\left(1-\rho^2\right)} \text{ and } \omega_{2opt}=\frac{\bar{Y}}{\bar{\bar{X}}}\left[\frac{1}{2c_k}-\omega_{1opt}\left\{\frac{3}{c_k}-\rho\frac{C_y}{C_x}\right\}\right]$$

Substituting these values in (3.6), we get

$$MSE\left(\bar{y}_{prk}\right)=\bar{Y}^2\left[\left\{1-\lambda\frac{C_x^2}{4{c_k}^2}\right\}-\frac{\left(1-\lambda\frac{C_x^2}{8{c_k}^2}\right)^2}{\left\{1+\lambda C_y^2\left(1-\rho^2\right)\right\}}\right] \qquad (3.9)$$

4. THEORETICAL COMPARISON OF THE PROPOSED ESTIMATOR IN SRSWOR

In this section we will compare the MSE of our proposed estimator with the existing estimator discussed here above.

1) $MSE\left(\bar{y}_{prk}\right) < MSE\left(\bar{y}\right)$

If

$$\left\{1-\lambda\frac{C_x^2}{4{c_k}^2}\right\}-\frac{\left(1-\lambda\frac{C_x^2}{8{c_k}^2}\right)^2}{\left\{1+\lambda C_y^2\left(1-\rho^2\right)\right\}}-\lambda C_y^2<0$$

this is hold true for all choices of c_k.

2) $MSE\left(\bar{y}_{prk}\right) < MSE\left(\bar{y}_{cr}\right)$

If

$$\frac{\left(1-\lambda\frac{C_x^2}{8{c_k}^2}\right)^2}{\left\{1+\lambda C_y^2\left(1-\rho^2\right)\right\}}+\lambda\left(C_y^2-C_x^2-2\rho C_y C_x\right)-\left\{1-\lambda\frac{C_x^2}{4{c_k}^2}\right\}>0$$

3) $MSE\left(\bar{y}_{prk}\right) < MSE\left(\bar{y}_{lr}\right)$

If

$$\frac{\left(1-\lambda\frac{C_x^2}{8{c_k}^2}\right)^2}{\left\{1+\lambda C_y^2\left(1-\rho^2\right)\right\}}+\lambda C_y^2\left(1-\rho^2\right)-\left\{1-\lambda\frac{C_x^2}{4{c_k}^2}\right\}>0$$

4) $MSE\left(\bar{y}_{prk}\right) < MSE\left(\bar{y}_{BT}\right)$

If

$$\left\{1-\lambda\frac{C_x^2}{4{c_k}^2}\right\}-\left[\frac{\left(1-\lambda\frac{C_x^2}{8{c_k}^2}\right)^2}{\left\{1+\lambda C_y^2\left(1-\rho^2\right)\right\}}+\lambda\left(C_y^2+\frac{C_x^2}{4}-C_{yx}\right)\right]>0$$

5) $MSE\left(\bar{y}_{prk}\right) < MSE\left(\bar{y}_{SG}\right)$

If

$$\left\{1-\lambda\frac{C_x^2}{4{c_k}^2}\right\}-\frac{\left(1-\lambda\frac{C_x^2}{8{c_k}^2}\right)^2}{\left\{1+\lambda C_y^2\left(1-\rho^2\right)\right\}}-\left\{1-\lambda\frac{C_x^2}{4\left(N+1\right)^2}\right\}+\frac{\left(1-\lambda\frac{C_x^2}{8\left(N+1\right)^2}\right)^2}{\left\{1+\lambda C_y^2\left(1-\rho^2\right)\right\}}<0$$

$$\left(\frac{1}{{c_k}^2}-\frac{1}{\left(N+1\right)^2}\right)-\lambda\frac{C_x^2}{4}\left(\frac{\lambda\frac{C_x^2}{8}\left(\frac{1}{\left(N+1\right)^2}-\frac{1}{{c_k}^2}\right)}{\left\{1+\lambda C_y^2\left(1-\rho^2\right)\right\}}\right)>0$$

Conditions 1 to 5 will always hold true for all types of real data especially for some choices of "c_k".

5. NUMERICAL COMPARISON OF THE PROPOSED ESTIMATOR FOR SRS

In this section we will make an assessment of the efficiency of our proposed estimator under SRS using data sets from real life examples. It should be noted that if we put $c=N+1$ or $\bar{X}^*=\bar{X}+N\bar{X}$ then our proposed estimator will reduces to Shabbir and Gupta (2011) estimator.

Parameters	**Population 1** Source: US Agricultural Statistics (1995) Y_i; Species group in 1995 X_i; Species group in 1995	**Population 3** Source: Pakistan MFA (2004) Y_i; District wise tomato production (in tons), in Pakistan (2003) X_i; District wise tomato production (2002)	**Population 2** Source: Agriculture Statistics (2010) Y_i; The State wise production of major spices in thousand metric tons of India. (2010,11) X_i; The State wise production of major spices in thousand metric tons of India.(2010,11)	**Population 4** Source: Murthy (1967) page 228
N	69	97	29	80
N	17	25	13	20
$\bar{Y}$	4514.9	3135.62	184.5	51.8264
$\bar{X}$	4505.16	3050.28	138.476	11.2664
C_x	1.483	2.327893	0.7531	0.7507
C_y	1.3756	2.3022	0.8561	0.3542
ρ	0.902	0.9871	0.4453	0.9513

Table 1
Mean Square Error of the Proposed Estimator for
$$c_1 = C_x, c_2 = \rho+1, c_3 = \frac{\rho+1}{2}, c_4 = \frac{\rho+1}{3}$$
and the Existing Estimators

Estimators	**Population 1**	**Population 2**	**Population 3**	**Population 4**
$\bar{y}$	1709951.485	2954.736	1547189.37	11067.086
$\bar{y}_{cr}$	388714.563	3282.800	40518.701	10960.842
$\bar{y}_{lr}$	317719.2201	2105.623	39623.005	8872.571
t_1	367322.320	1621.770	154253.921	9339.943
t_2	360544.4803	1733.541	158229.632	9473.291
$\bar{y}_{BT}$	381361.836	1708.851	172768.63	9574.837
$\bar{y}_{SG}$	312841.536	1982.820	39463.855	8766.500
Proposed Class of Estimators				
$\bar{y}_{pr1}$	308716.861	1960.990	39036.207	8716.599
$\bar{y}_{pr2}$	274913.551	1657.401	15267.600	8627.601
$\bar{y}_{pr3}$	300624.322	1863.417	33791.110	8711.855
$\bar{y}_{pr4}$	310484.230	1956.834	38807.953	8753.915

CONCLUSION

From the above analysis it is clear that our proposed estimators for various value of transformer $c_1 = C_x, c_2 = \rho+1, c_3 = \frac{\rho+1}{2}, c_4 = \frac{\rho+1}{3}$ are superior to all other estimators considered in this paper. Hence the transformation results considerable reduction in MSE, as obvious from the above table and high gain in efficiency. Similar strategy can be used for further optimization of existing estimators or for the development of new estimators.

REFERENCES

1. Bahl, S. and Tuteja, R.K. (1991). Ratio and product type exponential estimator. *Journal of Information and Optimization Sciences*, 12, 159-163.
2. 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), 262-275
3. Gupta, G. and Shabbir, J. (2008). On improvement in estimating the population mean in simple random sampling. *Journal of Applied Statistics*, 35(5), 559-566.
4. Hansen M.H. and Hurwitz W.N. (1943). On the theory of sampling from finite populations. *Annal of the Mathematics and Statistics*, 14, 333-62.
5. Government of Pakistan (2004). Crops Area Production. Ministry of Food and Agriculture, Islamabad, Pakistan.
6. Mohanty, S. and Sahoo, J. (1995). A note on improving the ratio method of estimation through linear transformation using certain known population parameters. *Sankhya Series* B 57(1), 93-102.
7. Shabbir, J. and Gupta, S. (2011). On estimating finite population mean in simple and stratified random sampling. *Communication in Statistics Theory and Method*, 40, 199-212.
8. Singh, H.P., Tailor, R. and Kakaran, M.S. (2004). An estimator of Population mean using power transformation. *J.I.S.A.S.*, 58(2), 223-230.
9. Srivastava, S.K. and Jhajj, H.S. (1983). A class of estimators of the population mean using multi-auxiliary information. *Calc. Statist. Assoc. Bull.*, 32, 47-56.