A New Generalized Class of Ratio-Cum-Regression Estimator of Population Mean in Stratified Sampling.

Rafia Jan

Department of Statistics, University of Kashmir, J&K, India

Email-rafiajan836@gmail.com

Asra Nazir

Department of Statistics, University of Kashmir, J&K, India

Email-asranazir44@gmail.com

Showkat Maqbool

Division of Animal Genetics and Breeding, Skuast, Kashmir, India

Email-Showkatmaq@gmail.com

T. R. Jan

Department of Statistics, University of Kashmir, J&K, India

Email-drtrjan@gmail.com

Abstract-The proposed article, suggests a generalised estimator in stratified random sampling, with its properties. Numerous well-known estimators/classes of estimators are recognized as members of the recommended class. The recommended class of estimators gives minimum mean squared error (MSE). Some real data sets are used to detect the efficiency of the estimators.

Key words- Study variable, auxiliary variable, stratified random sampling, population mean, bias, MSE.

I. INTRODUCTION

Ratio estimator in stratified sampling was given by [1]. [2] suggested a family of estimators of population mean using auxiliary information in stratified sampling. [3] proposed an improved exponential estimator in stratified sampling. [4] proposed a generalised ratio-cum-product estimator of finite population mean in stratified sampling. [5] proposed various estimators for estimating population parameters using auxiliary information in stratified sampling.

II. NOTATIONS

consider a population *U* consist of *L* strata as $U = U_1, U_2, ...U_N$ of *N* units. Here the size of the stratum U_N is N_h, and the size of simple random sample in stratum U_N is n_h where $h = 1,2,......L.$

Population means of *y* and *x* are $\bar{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$, $\bar{X}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} x_{hi}$ and the corresponding sample means of the stratified sampling are $\bar{y}_{st} = \sum_{i=1}^{L} W_h \bar{y}_h$, $\bar{x}_{st} = \sum_{i=1}^{L} W_h \bar{x}_h$, where $\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$, $\bar{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$ are sample means in the *h*-th stratum and $W_h = \frac{N_h}{N}$ is stratum weight. $\bar{y}_{st}$ and $\bar{x}_{st}$ are unbiased estimators of population mean $\bar{Y} = \sum_{i=1}^{L} W_h \bar{Y}_h$ and $\bar{X} = \sum_{i=1}^{L} W_h \bar{X}_h$ respectively.

The separate ratio estimator is given as

$$\bar{y}_{RS} = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{\bar{X}_h}{\bar{x}_h} \right) \tag{1}$$

The separate product estimator is given by

$$\bar{y}_{PS} = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{\bar{x}_h}{\bar{X}_h} \right) \tag{2}$$

And the separate regression estimator is

$$\bar{y}_{lrs} = \sum_{h=1}^{L} W_h \left[\bar{y}_h + b_h \left(\bar{X}_h - \bar{x}_h \right) \right], \tag{3}$$

where $b_h = \frac{s_{yxh}}{s_{xh}^2}$, $s_{xh}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(x_{hi} - \bar{x}_h \right)^2$ and $s_{yxh} = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(y_{hi} - \bar{y}_h \right)\left(x_{hi} - \bar{x}_h \right)$

The combined ratio estimator is given as

$$\bar{y}_{RC} = \bar{y}_{st} \left(\frac{\bar{X}}{\bar{x}_{st}} \right) \tag{4}$$

The combined product estimator is

$$\bar{y}_{PC} = \bar{y}_{st} \left(\frac{\bar{x}_{st}}{\bar{X}} \right) \tag{5}$$

The combined regression estimator is

$$\bar{y}_{lrc} = \bar{y}_{st} + b\left(\bar{X} - \bar{x}_{st} \right), \tag{6}$$

where $b = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h s_{yxh}}{\sum_{h=1}^{L} W_h^2 \lambda_h s_{xh}^2}$, $\lambda_h = \frac{1}{n_h} - \frac{1}{N_h}$.

The variance of $\bar{y}_{st}$ and MSE of $\bar{y}_{RS}$, $\bar{y}_{PS}$, $\bar{y}_{lrs}$, $\bar{y}_{Rc}$, $\bar{y}_{pc}$ and $\bar{y}_{lrc}$ are given as

$$V\left(\bar{y}_{st} \right) = \sum_{h=1}^{L} W_h^2 \lambda_h S_{yh}^2 = \sum_{h=1}^{L} W_h^2 \lambda_h \bar{Y}_h^2 C_{yh}^2 \tag{7}$$

$$MSE\left(\bar{y}_{RS} \right) = \sum_{h=1}^{L} W_h^2 \lambda_h \left(S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h S_{yxh} \right) \tag{8}$$

$$MSE\left(\bar{y}_{PS} \right) = \sum_{h=1}^{L} W_h^2 \lambda_h \left(S_{yh}^2 + R_h^2 S_{xh}^2 + 2R_h S_{yxh} \right) \tag{9}$$

$$MSE\left(\bar{y}_{irs} \right) = \sum_{h=1}^{L} W_h^2 \lambda_h S_{yh}^2 \left(1 - \rho_{yxh}^2 \right) \tag{10}$$

$$MSE\left(\bar{y}_{RC} \right) = \sum_{h=1}^{L} W_h^2 \lambda_h \left(S_{yh}^2 + RS_{xh}^2 - 2RS_{yxh} \right) \tag{11}$$

$$MSE(\bar{y}_{PC}) = \sum_{h=1}^{L} W_h^2 \lambda_h \left(S_{yh}^2 + RS_{xh}^2 + 2RS_{yxh}\right) \quad (12)$$

$$MSE(\bar{y}_{irc}) = \sum_{h=1}^{L} W_h^2 \lambda_h S_{yh}^2 \left(1-\rho_{yx}^2\right) \quad (13)$$

III. PROPOSED SEPARATE RATIO-CUM-REGRESSION ESTIMATOR

The proposed class of separate ratio-cum-regression estimator is given by

$$t_S = \sum_{h=1}^{L} W_h \left[\bar{y}_h + \beta_h \left(\bar{X}_h - \bar{x}_h\right) \left(\frac{a\bar{X}_h + b}{a\bar{x}_{st} + b} \right)^{2g-1} \right], \quad (14)$$

where g is a suitable chosen scaler to minimize the MSE.

To find the bias and MSE, we write

$$\bar{y}_h = \bar{Y}_h (1 + e_{0h}),$$

$$\bar{x}_h = \bar{X}_h (1 + e_{1h}),$$

such that

$$E(e_{0h}) = E(e_{1h}) = 0,$$

and

$$E\left(e_{0h}^{\ 2}\right) = \lambda_h C_{yh}^{\ 2},$$

$$E\left(e_{1h}^{\ 2}\right) = \lambda_h C_{xh}^{\ 2},$$

$$E(e_{0h} e_{1h}) = \lambda_h \rho_{xyh} C_{yh} C_{xh}.$$

writing equation (14) in terms of e's, we get

$$t_S = \sum_{h=1}^{L} W_h \left[\left\{ \bar{Y}_h (1 + e_{0h}) - \beta_h \bar{X}_h e_{1h} \right\} \left\{ (1 + \theta_h e_{1h})^{-(2g-1)} \right\} \right],$$

where $\theta_h = \left(\dfrac{a\bar{X}_h}{a\bar{X}_h + b} \right).$

Solving and keeping the terms up to second order we have

$$\left(t_S - \bar{Y}_h\right) = \sum_{h=1}^{L} W_h \left(\bar{Y}_h e_{0h} - (2g-1)\theta_h \bar{Y}_h e_{1h} - \beta_h \bar{X}_h e_{1h} \right)$$

Squaring and taking expectation on both sides we get the MSE,

$$MSE(t_S) = \sum_{h=1}^{L} W_h^2 \gamma_h \begin{bmatrix} S_{yh}^2 + (2g-1)^2 R_h^2 \theta_h^2 S_{xh}^2 + \beta_h^2 S_{xh}^2 - 2(2g-1) R_h \theta_h \rho_{xyh} S_{yh} S_{xh} \\ -2\beta_h \rho_{xyh} S_{yh} S_{xh} + 2(2g-1)\beta_h R_h \theta_h S_{xh}^2 \end{bmatrix}$$

After differentiating and equating the $MSE(t_S)$ to zero, we get g_{opt} given as,

$$g_{opt} = \frac{1}{2}.$$

The $MSE_{\min}(t_S)$ is given by

$$MSE(t_S) = \sum_{h=1}^{L} W_h^2 \gamma_h \left[S_{yh}^2 \left(1 - \rho_{yxh}^2\right) \right] = MSE(\bar{y}_{lrs}) \quad (15)$$

IV. PROPOSED COMBINED RATIO-CUM-REGRESSION ESTIMATOR

The proposed class of combined ratio-cum-regression estimator is given by

$$t_C = \left[\bar{y}_{st} + \beta\left(\bar{X} - \bar{x}_{st}\right)\right]\left(\frac{a\bar{X}+b}{a\bar{x}_{st}+b}\right)^{2g-1} \qquad (16)$$

We define,

$$E(e_{0h}) = E(e_{1h}) = 0,$$

and

$$E\left(e_{0h}^{\ 2}\right) = \lambda_h C_{yh}^{\ 2},$$
$$E\left(e_{1h}^{\ 2}\right) = \lambda_h C_{xh}^{\ 2},$$
$$E(e_{0h}e_{1h}) = \lambda_h \rho_{xyh} C_{yh} C_{xh}.$$

writing equation (16) in terms of $e's$, we get

$$t_C = \left[\left\{\bar{Y}(1+e_0) - \beta\bar{X}e_1\right\}\left\{(1+\theta e_1)^{-(2g-1)}\right\}\right],$$

where $\theta = \left(\frac{a\bar{X}}{a\bar{X}+b}\right)$, $e_0 = \frac{\sum_{h=1}^{L} W_h \bar{Y}_h e_{0h}}{\bar{Y}}$ and $e_1 = \frac{\sum_{h=1}^{L} W_h \bar{X}_h e_{1h}}{\bar{X}}$,

such that

$$E(e_0) = E(e_1) = 0,$$

and

$$E\left(e_0^{\ 2}\right) = \sum_{h=1}^{L} W_h^2 \lambda_h C_{yh}^{\ 2},$$
$$E\left(e_1^{\ 2}\right) = \sum_{h=1}^{L} W_h^2 \lambda_h C_{xh}^{\ 2},$$
$$E(e_0 e_1) = \sum_{h=1}^{L} W_h \lambda_h \rho_{xyh} C_{yh} C_{xh}.$$

Solving and keeping the terms up to second order we have

$$\left(t_C - \bar{Y}\right) = \left(\bar{Y}e_0 - (2g-1)\theta\bar{Y}e_1 - \beta\bar{X}e_1\right)$$

Squaring and taking expectation on both sides we get the MSE,

$$MSE(t_C) = \sum_{h=1}^{L} W_h^2 \gamma_h \begin{bmatrix} S_{yh}^2 + (2g-1)^2 R^2\theta^2 S_{xh}^2 + \beta^2 S_{xh}^2 - 2(2g-1)R\theta\rho_{xyh} S_{yh} S_{xh} \\ -2\beta\rho_{xyh} S_{yh} S_{xh} + 2(2g-1)\beta R\theta S_{xh}^2 \end{bmatrix}$$

Putting the value of g_{opt} we get,

$$MSE(t_C) = \sum_{h=1}^{L} W_h^2 \gamma_h \left[S_{yh}^2\left(1-\rho_{yxh}^2\right)\right] = MSE\left(\bar{y}_{lrC}\right) \qquad (17)$$

V. EFFICIENCY COMPARISON

We compare the efficiency of the proposed estimator with classical estimators, the conditions are given as

$$V(\bar{y}_{st}) - MSE(t_S) > 0$$
$$= \sum_{h=1}^{L} W_h^2 \lambda_h S_{yh}^2 - \sum_{h=1}^{L} W_h^2 \gamma_h \left[S_{yh}^2 \left(1 - \rho_{yxh}^2\right)\right] > 0$$
$$= \sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^2 \rho_{yxh}^2 > 0$$
$$MSE(\bar{y}_{RS}) - MSE(t_S) > 0$$
$$= \sum_{h=1}^{L} W_h^2 \lambda_h \left(R_h^2 S_{xh}^2 - 2R_h S_{yxh} + S_{yh}^2 \rho_{yxh}^2\right) > 0$$
$$MSE(\bar{y}_{PS}) - MSE(t_S) > 0$$
$$= \sum_{h=1}^{L} W_h^2 \lambda_h \left(R_h^2 S_{xh}^2 + 2R_h S_{yxh} + S_{yh}^2 \rho_{yxh}^2\right) > 0$$
$$MSE(\bar{y}_{RC}) - MSE(t_C) > 0$$
$$= \sum_{h=1}^{L} W_h^2 \lambda_h \left(RS_{xh}^2 - 2RS_{yxh} + S_{yh}^2 \rho_{yxh}^2\right) > 0$$
$$MSE(\bar{y}_{PC}) - MSE(t_C) > 0$$
$$= \sum_{h=1}^{L} W_h^2 \lambda_h \left(RS_{xh}^2 + 2RS_{yxh} + S_{yh}^2 \rho_{yxh}^2\right) > 0$$

VI. NUMERICAL ILLUSTRATION

For numerical illustrations we have taken the data from:
(1)[6]; where *y* is total number of trees and *x* is area under orchards in hectares.
(2) [7]; here *y* is juice quantity and *x* is weight of cane.
(3) [8]; where *y* is study variable and *x* is auxiliary variable.

Table 1: Data Statistics.

Population	Data Sets	Stratum no.			values of parameters for h^th stratum				
		(h)	N_h	n_h	$\bar{Y}_h$	$\bar{X}_h$	S_{yh}^2	S_{xh}^2	S_{xyh}
(1)	$N = 25, n = 10$	1	6	3	417.33	6.81	74775.87	15.97	1007.05
	$\bar{Y} = 410.841$	2	8	3	503.38	10.12	259113.70	132.66	5709.16
	$\bar{X} = 8.3796$	3	11	4	340.00	7.97	65885.60	38.44	1404.71
(2)	$N = 25, n = 10$	1	6	3	135.00	366-67	80.00	2706.67	440.00
	$\bar{Y} = 102.6$	2	12	4	99.17	310.83	226.52	1881.06	618.94
	$\bar{X} = 325.998$	3	7	3	80.71	317.14	120.24	2890.48	444.05

(3)	$N=20, n=8$	1	10	4	149.70	1622.99	181.17	10438.71	-1072.80
	$\bar{Y}=126.15$	2	10	4	102.60	2035.95	158.76	10662.63	-655.25
	$\bar{X}=1829.47$								

The following table shows the MSE and PRE of the proposed estimator and other estimators present in the literature.

Table 2

Population	Population 1		Population 2		Population 3	
Estimator	MSE	PRE	MSE	PRE	MSE	PRE
$\bar{y}_{st}$	8274.88	100.00	11.26	100.00	12. 75	100.00
$\bar{y}_{RS}$	1014.64	815.55	3.28	343.23	*	*
$\bar{y}_{PS}$	*	*	*	*	7.19	177.17
$\bar{y}_{irs}$	842.62	**982.04**	1.74	**648.64**	7.10	**179.17**
$\bar{y}_{RC}$	1159.01	713.96	3.47	324.28	*	*
$\bar{y}_{PC}$	*	*	*	*	7.57	168.33
$\bar{y}_{irc}$	948.89	872.06	2.78	404.62	7.44	171.32
t_S	842.62	**982.04**	1.74	**648.64**	7.10	**179.17**
t_C	842.62	**982.04**	1.74	**648.64**	7.10	**179.17**

VI. CONCLUSION

We have proposed, classes of separate and combined ratio-cum-product estimators in stratified sampling. The theoretical results were numerically justified by the data sets considered in Table 1. Also, it follows from Table 2 that there is significant gain in efficiency by the proposed classes of estimators.

REFERENCES

[1] Kadilar C. and Cingi, H. (2005): *A new ratio estimator in stratified random sampling*, Comm. Statist. Theory Methods 34, pp. 597–602.

[2] Singh, H. P. and Vishwakarma, G. K. (2008): *A family of estimators of population mean using auxiliary information in stratified sampling*, Comm. Statist. Theory Methods 37, pp. 1038–1050.

[3] Koyuncu, N. and Kadilar, C. (2009): *Ratio and product estimators in stratified random sampling*, J. Statist. Plann. Inference 13, pp. 2552–2558.

[4] Tailor, R., Sharma, B. and Kim, J. M. (2011): *A generalized ratio-cum-product estimator of finite population mean in stratified random sampling*, Commun. Korean Statist. Soc. 18, pp. 111–118.

[5] Vishwakarma. G. K. and Kumar, M. (2016): *Efficient classes of ratio-cum-product estimators of population mean in stratified random sampling*, Probability and Mathematical Statistics, 36, pp. 267–278

[6] Singh, D. and Chaudhary, F. S. (1986): *Theory and Analysis of Sample Survey Designs*, New Age International, New Delhi 1986.

[7] Singh, R. and Mangat, S. N. (1996): *Elements of Survey Sampling*, Kluwer Academic Publishers, Dordrecht–Boston 1996.
[8] Japan Meteorological Society, [retrieved: May 26, 2016]. *http://www.data.jma.go.jp/obd/stats/data/en/smp/index.html*