Revista Colombiana de Estadística - Theoretical statistics
July 2023, volume 46, issue 2, pp. 93–110
http://doi.org/10.15446/rce.v46n2.104591

Some Improved Combined Estimators of Population Mean in Stratified Ranked Set Sampling

Algunos estimadores combinados mejorados de la media de la población en el muestreo de conjuntos clasificados estratificados

Rajesh Singh[a], Anamika Kumari[b]

Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India

Abstract

This paper presents improved population mean estimators using auxiliary variables in Stratified Ranked Set Sampling. We have derived the expressions for bias and mean square errors up to the first order of approximation and shown that the proposed estimators under optimum conditions are more efficient than other estimators taken in this paper. In an attempt to verify the efficiencies of proposed estimators, theoretical results are supported by numerical illustrations and simulation study for which we have considered two populations.

***Key words*:** Auxiliary variable; Bias; Mean square error; Ranked set sampling; Study variable.

Resumen

Este artículo presenta estimadores mejorados de la media de la población utilizando variables auxiliares en el muestreo de conjuntos ordenados estratificados. Hemos derivado las expresiones para el sesgo y los errores cuadráticos medios hasta el primer orden de aproximación y hemos demostrado que los estimadores propuestos en condiciones óptimas son más eficientes que otros estimadores tomados en este artículo. En un intento por verificar las eficiencias de los estimadores propuestos, los resultados teóricos están respaldados por ilustraciones numéricas y estudios de simulación para los cuales hemos considerado dos poblaciones.

***Palabras clave*:** Error cuadrático medio; Inclinación; Muestreo de conjuntos clasificados; Variable de estudio, Variable auxiliar.

[a]Ph.D. E-mail: rsinghstat@gmail.com
[b]Ph.D. E-mail: anamikatiwari1410@gmail.com

1. Introduction

In theory of sampling it is evident that suitable use of auxiliary information improves the efficiency of the estimator. These auxiliary information may be used either at the design phase or the estimation phase or at both phases. Cochran (1977) was the first to introduce a ratio estimator of Population Mean using auxiliary information. Shabbir & Gupta (2007), Koyuncu & Kadilar (2009) and Chaudhary et al. (2009) have considered the problem of estimating population mean taking into consideration information on auxiliary variable.

When population is heterogenous stratified random sampling (SSRS) is used for better accuracy. Several authors like Kadilar & Cingi (2003), Shabbir & Gupta (2006) and Haq & Shabbir (2013) have proposed estimators in stratified random sampling using information on a single auxiliary variable. Singh & Kumar (2012) have proposed improved estimators of population mean using two auxiliary variables in stratified random sampling. Recently, Muneer et al. (2017) have proposed family of chain exponential estimators in SSRS.

Ranked set sampling (RSS) is an improved sampling method over Simple Random Set Sampling (SRS). McIntyre (1952) was the first to explain RSS for estimating the population means. Takahasi & Wakimoto (1968) gave the necessary mathematical theory of RSS. Samawi & Muttlak (1996) suggested ratio estimators of population mean in RSS and showed that the RSS estimators gave improved results over their SRS counterparts. Ganeslingam & Ganesh (2006) compared RSS with SRS for estimation of the unknown mean of study variable and the ratio of study variable to auxiliary variable. He concluded that RSS gives a better estimate for both the mean and the ratio. Singh et al. (2014) suggested a general procedure for estimating the population mean using RSS. Al-Omari & Bouza (2014) and Bouza et al. (2018) provided a review of RSS, its modification, and its application. Stratified ranked set sampling (SRSS) was first introduced by Samawi & Muttlak (1996) for estimation of population mean. Samawi & Siam (2003) have proposed the combined and the separate ratio estimators in SRSS.

Following is how the rest of the article is organized. The SRSS sampling approach is presented in section 2. Section 3 examines a review of existing works. In section 4 and 5, the proposed estimators and their characteristics are discussed. Numerical illustrations are provided in section 6. In section 7, using a simulated populations, the effectiveness of proposed estimators is compared to that of existing estimators. Section 8 draws the conclusion.

2. Sampling Methodology

In ranked set sampling (RSS), we rank randomly selected units from the population merely by observation or prior experience after which only a few of these sampled units are measured. In RSS, k independent random sets each of size k are selected from the population and each unit in the set is being selected with equal probability (SRS). The members of each random set are ranked with respect to the characteristic of the auxiliary variable. Then the smallest unit is selected from the first ordered set and the second smallest unit is selected from the second ordered set. By this way, this procedure is continued until the largest rank is chosen from the k^{th} set. This cycle may be repeeated r times, so $r(k=n)$ units have been measured during this process.

SRSS takes the following steps.

- Step 1: Select k_h^2 bivariate sample units randomly from the h^{th} stratum of the population.
- Step 2: Arrange these selected units randomly into k_h sets, each of size k_h.
- Step 3: The procedure of ranked set sampling (RSS) is then applied, on each of the sets to obtain the k_h ranked set sample unit. Here ranking is done with respect to the auxiliary variable X_h.
- Step 4: Repeat the above steps r times for each stratum to get the desired sample of size $n_h = k_h r$.

Consider a finite population $U = (U_1, U_2, \ldots, U_N)$ based on N identifiable units with a study variable Y and auxiliary variables X associated with each unit U_i, $i = 1, 2, \ldots, N$ of the population. Let the population be divided into L disjoint strata with stratum h based on N_h, $h = 1, 2, \ldots, L$ units.

Let $(Y_{h[1]j}, X_{h(1)j}), (Y_{h[2]j}, X_{h(2)j}), \ldots, (Y_{h[k_h]j}, X_{h(k_h)j})$ be the stratified ranked set sample for j^{th}, $j = 1, 2, \ldots, r$ cycle in h^{th} stratum.

Let $\overline{y}_{SRSS} = \sum_{h=1}^{L} W_h \overline{y}_{h[rss]}$ and $\overline{x}_{SRSS} = \sum_{h=1}^{L} W_h \overline{x}_{h(rss)}$

respectively be the stratified ranked set sample means corresponding to the population means $\overline{Y} = \sum_{h=1}^{L} W_h \overline{Y}_h$ and $\overline{X} = \sum_{h=1}^{L} W_h \overline{X}_h$ of variables Y and X, where $W_h = \frac{N_h}{N}$ is the weight in stratum h.

Let $\overline{y}_{h[rss]} = \sum_{i=1}^{k_h} \sum_{j=1}^{r} \frac{Y_{h[i]j}}{k_h r}$ and $\overline{x}_{h(rss)} = \sum_{i=1}^{k_h} \sum_{j=1}^{r} \frac{X_{h[i]j}}{k_h r}$ be the stratified ranked set sample means corresponding to the population means $\overline{Y}_h = \sum_{j=1}^{N_h} \frac{Y_{h[i]j}}{N_h}$ and $\overline{X}_h = \sum_{j=1}^{N_h} \frac{X_{h(i)j}}{N_h}$ of variables Y and X in stratum h.

Let $s_{yh}^2 = \frac{1}{n_h - 1} \sum_{h=1}^{L} \left(Y_{h[i]} - \overline{y}_{h[rss]}\right)^2$, $s_{xh}^2 = \frac{1}{n_h - 1} \sum_{h=1}^{L} \left(X_{h(i)} - \overline{x}_{h(rss)}\right)^2$ and $s_{xyh} = \frac{1}{n_h - 1} \sum_{h=1}^{L} (Y_{h[i]} - \overline{y}_{h[rss]})(X_{h(i)} - \overline{x}_{h(rss)})$, respectively be the sample variances and covariances corresponding to the population variances and covariances $S_{yh}^2 = \frac{1}{N_h - 1} \sum_{h=1}^{L} \left(Y_{h[i]} - \overline{Y}_h\right)^2$, $S_{xh}^2 = \frac{1}{N_h - 1} \sum_{h=1}^{L} \left(X_{h(i)} - \overline{X}_h\right)^2$ and $S_{xyh} = \frac{1}{N_h - 1} \sum_{h=1}^{L} (Y_{h[i]} - \overline{Y}_h)(X_{h(i)} - \overline{X}_h)$ in the stratum h.

Let C_{yh} and C_{xh} respectively be the population coefficient of variation of variables Y and X.

3. Existing Estimators

The conventional combined estimator of the population mean $\overline{Y}$ under SRSS is given by

$$t^c = \overline{y}_{SRSS} \tag{1}$$

The variance of the estimator t^c is given by

$$Var(t^c) = \overline{Y}^2 V_{20} \tag{2}$$

The classical combined ratio estimator of the population mean $\overline{Y}$ given by Samawi & Siam (2003) under SRSS is defined as

$$t_r^c = \overline{y}_{SRSS} \frac{\overline{X}}{\overline{x}_{SRSS}} \tag{3}$$

The Mean Squared Error (MSE) of the estimator t_r^c is given by

$$MSE(t_r^c) = \overline{Y}^2 [V_{20} + V_{02} - 2V_{11}] \tag{4}$$

The classical combined regression estimator of the population mean $\overline{Y}$ under SRSS is given as

$$t_{lr}^c = \overline{y}_{SRSS} + \beta(\overline{X} - \overline{x}_{SRSS}) \tag{5}$$

The Mean Squared Error (MSE) of the estimator t_{lr}^c is given by

$$MSE(t_r^c) = \overline{Y}^2 V_{20} + \beta^2 \overline{X}^2 V_{02} - 2\beta \overline{Y}\overline{X} V_{11} \tag{6}$$

where β is the regression coefficient of Y on X.

4. Proposed Estimators

Motivated by Bhushan et al. (2020), we suggest some estimators of the population mean $\overline{Y}$ using SRSS as

$$t_{p1}^c = \overline{y}_{SRSS} \exp\left(\alpha_1 \left(\frac{\overline{x}_{SRSS}}{\overline{X}} - 1\right)\right) \tag{7}$$

$$t_{p2}^c = \overline{y}_{SRSS} \exp\left(\alpha_2 \log \frac{\overline{x}_{SRSS}}{\overline{X}}\right) \tag{8}$$

where α_1 and α_2 are constants such that MSE of the estimators is minimum.

Proposition 1. *The Bias, MSE and min MSE of the proposed estimators are*

$$Bias(t_{p1}^c) = \overline{Y}\left(\frac{\alpha_1^2}{2} V_{02} + \alpha_1 V_{11}\right)$$

$$Bias(t_{p2}^c) = \overline{Y}\left(\frac{(\alpha_2^2 - \alpha_2)}{2} V_{02} + \alpha_2 V_{11}\right)$$

$$MSE(t_{p1}^c) = \overline{Y}^2 \left(V_{20} + \alpha_1^2 V_{02} + 2\alpha_1 V_{11}\right)$$

$$MSE(t_{p2}^c) = \overline{Y}^2 \left(V_{20} + \alpha_2^2 V_{02} + 2\alpha_2 V_{11}\right)$$

$$MinMSE(t_{p1}^c) = \overline{Y}^2 \left(V_{20} - \frac{V_{11}^2}{V_{02}}\right)$$

$$MinMSE(t_{p2}^c) = \overline{Y}^2 \left(V_{20} - \frac{V_{11}^2}{V_{02}}\right)$$

Proof. Outline of the derivations are given in Appendix. □

5. Some Other Proposed Estimators

We propose modified estimators of population mean by $\overline{Y}$ under SRSS as

$$t_{p3}^c = [(1+w_1)\overline{y}_{SRSS} + w_2(\overline{X} - \overline{x}_{SRSS})]\frac{\overline{X}}{\overline{x}_{SRSS}} \tag{9}$$

$$t_{p4}^c = [(1+w_3)\overline{y}_{SRSS} + w_4(\overline{X} - \overline{x}_{SRSS})]\exp\left(\frac{\overline{X} - \overline{x}_{SRSS}}{\overline{X} + \overline{x}_{SRSS}}\right) \tag{10}$$

$$t_{p5}^c = w_5\overline{y}_{SRSS} + w_6\exp\left(\frac{\overline{X} - \overline{x}_{SRSS}}{\overline{X} + \overline{x}_{SRSS}}\right)\left(1 + log\frac{\overline{x}_{SRSS}}{\overline{X}}\right) \tag{11}$$

$$t_{p6}^c = w_7\overline{y}_{SRSS} + w_8\left(\frac{\overline{X}}{\overline{x}_{SRSS}}\right)\exp\left(\frac{\overline{X} - \overline{x}_{SRSS}}{\overline{X} + \overline{x}_{SRSS}}\right) \tag{12}$$

Proposition 2. *The Bias, MSE and min MSE of the proposed estimators are*

$$bias(t_{p3}^c) = \overline{Y}w_1 + \overline{Y}(V_{02} + w_1V_{02} + w_2\delta V_{02} - V_{11} - w_1V_{11})$$

$$bias(t_{p4}^c) = \overline{Y}w_3 + \overline{Y}\left(\frac{3}{8}V_{02} + \frac{3}{8}w_3V_{02} + \frac{1}{2}w_4\delta V_{02} - \frac{1}{2}V_{11} - \frac{1}{2}w_3V_{11}\right)$$

$$Bias(t_{p5}^c) = (w_5 - 1)\overline{Y} + w_6\left(1 - \frac{5}{8}V_{02}\right)$$

$$Bias(t_{p6}^c) = (w_7 - 1)\overline{Y} + w_6\left(1 + \frac{15}{8}V_{02}\right)$$

$$MSE(t_{p3}^c) = \overline{Y}^2(A_1 + w_1^2B_1 + w_2^2C_1 + 2w_1D_1 - 2w_2E_1 - 2w_1w_2F_1)$$

$$MSE(t_{p4}^c) = \overline{Y}^2(A_2 + w_3^2B_2 + w_4^2C_2 + 2w_3D_2 - 2w_4E_2 - 2w_3w_4F_2)$$

$$minMSE(t_{p3}^c) = \overline{Y}^2\left(A_1 + \frac{C_1D_1^2 + B_1E_1^2 - 2D_1E_1F_1}{F_1^2 - B_1C_1}\right)$$

$$minMSE(t_{p4}^c) = \overline{Y}^2\left(A_2 + \frac{C_2D_2^2 + B_2E_2^2 - 2D_2E_2F_2}{F_2^2 - B_2C_2}\right)$$

***Proof*.** Outline of the derivations are given in Appendix. □

Case 1: Sum of Weights is Unity ($w_5 + w_6 = 1$ & $w_7 + w_8 = 1$).

Proposition 3. *The MSE and min MSE of the proposed estimators are The MSE of the estimator t_{p5}^c is given by*

$$MSE(t_{p5}^c) = \overline{Y}^2(V_{20} + w_6^2V_{02} - 2w_6V_{11})$$

$$MSE(t_{p6}^c) = \overline{Y}^2(V_{20} + w_8^2V_{02} - 2w_8V_{11})$$

$$MinMSE(t_{p5}^c) = \overline{Y}^2\left(V_{20} - \frac{V_{11}^2}{V_{02}}\right)$$

$$MinMSE(t_{p6}^c) = \overline{Y}^2\left(V_{20} - \frac{V_{11}^2}{V_{02}}\right)$$

Proof. Outline of the derivations are given in Appendix. □

Case 2: Sum of Weights is Flexible ($w_5 + w_6 \neq 1$ **&** $w_7 + w_8 \neq 1$)

Proposition 4. *The MSE and min MSE of the proposed estimators are*

$$MSE(t^c_{p5}) = C_3 + w_5^2 A_3 + w_6^2 B_3 - 2w_5 C_3 - 2w_6 D_3 + 2w_5 w_6 E_3$$

$$MSE(t^c_{p6}) = C_4 + w_7^2 A_4 + w_8^2 B_4 - 2w_7 C_4 - 2w_8 D_4 + 2w_7 w_8 E_4$$

$$minMSE(t^c_{p5}) = C_3 + \frac{B_3 C_3^2 + A_3 D_3^2 - 2C_3 D_3 E_3}{E_3^2 - A_3 B_3}$$

$$minMSE(t^c_{p6}) = C_4 + \frac{B_4 C_4^2 + A_4 D_4^2 - 2C_4 D_4 E_4}{E_4^2 - A_4 B_4}$$

Proof. Outline of the derivations are given in Appendix. □

6. Numerical Illustrations

In this section, we compare the performance of the proposed estimators with the other estimators considered in this paper. For comparison, we have taken a real data of area under tobaco production and production of tobaco of specified countries during 1998 given by Singh (2003) (Appendix) given in Table 1, where y is production (study variable) in metric tons and x is area (auxiliary variable) in hectares. For the above population, the parameters are given as below: For total, $N = 67$, $\overline{Y} = 72247.6$, $\overline{X} = 26438$.

Table 1: Data

Stratum 1	Stratum 2	Stratum 3
$N_1 = 20$	$N_2 = 30$	$N_3 = 17$
$n_1 = 12$	$n_2 = 18$	$n_3 = 9$
$W_1 = 0.29851$	$W_2 = 0.44776$	$W_3 = 0.25373$
$\overline{X}_1 = 6801.25$	$\overline{X}_2 11025.3$	$\overline{X}_3 = 82464.1$
$\overline{Y}_1 = 17511.7$	$\overline{Y}_2 = 18937.4$	$\overline{Y}_3 = 377960.5$
$S^2_{x1} = 175539558$	$S^2_{x2} = 595679198.4$	$S^2_{x3} = 20255478994$
$S^2_{y1} = 1366895911$	$S^2_{y2} = 2421559069$	$S^2_{y3} = 687956456787$
$S_{y1x1} = 489224338$	$S_{y2x2} = 1174423304$	$S_{y3x3} = 46735680920$
$C_{x1} = 1.94804$	$C_{x2} = 2.21368$	$C_{x3} = 1.72586$
$R_1 = 2.57477$	$R_2 = 1.71763$	$R_3 = 4.58333$

Case 1 (when initial sample is taken by SRS): From this population we took ranked set samples of sizes $k_1 = 4$, $k_2 = 6$ and $k_3 = 3$ from the stratum 1^{st}, 2^{nd} and 3^{rd} respectively. Further each ranked set sample from each stratum were repeated with number of cycles $r = 3$.

Case 2 (when initial sample is taken by PPS): From this population k independent random sets (initial sample) each of size k are selected from the population

Table 2: The MSE and PRE of the estimators

Estimators	MSE	PRE
t^c	566408434	100
t^c_r	201822821	280.64
t^c_{lr}	201169647	281.55
t^c_{p1}	201169647	281.55
t^c_{p2}	201169647	281.55
t^c_{p3}	193114626	293.30
t^c_{p4}	155280723	364.76
t^c_{p5}	55059116.5	1028.72
t^c_{p6}	41886045.6	1352.26

by using probability proportion to size (pps) sampling. We took ranked set samples of size $k = 3$ from each stratum. Further each ranked set sample from each stratum were repeated with number of cycles $r_1 = 3$, $r_2 = 2$ and $r_3 = 2$ from the stratum 1^{st}, 2^{nd} and 3^{rd}, respectively.

Table 3: The MSE and PRE of the estimators under PPS sampling

Estimators	MSE	PRE
t^c	98340720.29	100
t^c_r	14809479.67	664.03
t^c_{lr}	14474608.66	679.40
t^c_{p1}	14474608.66	679.40
t^c_{p2}	14474608.66	679.40
t^c_{p3}	14192040.24	692.92
t^c_{p4}	6775003.38	1451.52
t^c_{p5}	3539691.30	2778.22
t^c_{p6}	2914206.65	3374.52

The formula for Percent Relative Efficiency (PRE) is $PRE(estimators) = \frac{MSE(t^c)}{MSE(estimator)} \times 100$.

From Table 2, it is observed that

- The estimators t^c_{p1} and t^c_{p2} are almost equally efficient estimators as combined linear regression estimators under SRSS as these estimators show the MSE almost equal to the MSE of the combined linear regression estimator (t^c_{lr}). These two estimators t^c_{p1} and t^c_{p2} are more efficient estimators than that the other competitive estimators.
- t^c_{p3} ,t^c_{p4}, t^c_{p5} and t^c_{p6} are more efficient than other estimators used in this paper. It is observed that t^c_{p3}, t^c_{p4}, t^c_{p5} and t^c_{p6} are more efficient than convention, ratio estimator and linear regression estimator under SRSS.
- From Table 2, we can conclude that the proposed estimators perform better than existing estimators as our proposed estimators have greater PRE.
- From Table 3, we can conclude that the proposed estimators when initial sample is taken by PPS show same trend as usual proposed estimators. The

estimators t^c_{p1} and t^c_{p2} are almost equally efficient estimators as combined linear regression estimators under SRSS. t^c_{p3}, t^c_{p4}, t^c_{p5} and t^c_{p6} are more efficient than other estimators used in this paper. From Table 3, we can conclude that the proposed estimators perform better than existing estimators as our proposed estimators have greater PRE.

7. Simulation Study

To generalize the results of the numerical study, we have conducted simulation study over two hypothetically generated normal populations. The simulation procedure is explained in the following points:

- We generated bivariate random observations of size $N = 600$ units from a bivariate normal distribution with parameters $\mu_y = 20$, $\sigma_y = 15$, and $\mu_x = 15$, $\sigma_x = 10$ and passably chosen values of $\rho = 0.3, 0.5, 0.6, 0.7, 0.8, 0.9$.
- Similarly, generate the population-2 with the parameters $\mu_y = 120$, $\sigma_y = 25$, and $\mu_x = 100$, $\sigma_x = 20$.
- The population generated above is divided into 3 equal strata and a stratified ranked set sample of size 12 units with number of cycles 4 and set size 3 is drawn from each stratum.
- Compute the required statistics.
- Iterate the above steps 10000 times to calculate the MSE and PRE of various combined estimators using the following expression.

$$MSE(T) = \frac{1}{10000} \sum_{i=1}^{10000} (T_i - \overline{Y})^2 \tag{13}$$

$$PRE = \frac{Var(t^c)}{MSE(T)} \times 100 \tag{14}$$

The MSE and PRE of the combined estimators are calculated using (13) and (14) and the results are reported for various values of correlation coefficients in Table 4.

Table 4: The MSE and PRE of the estimators

r_{yx}	Estimators	Population 1		Population 2	
		MSE	PRE	MSE	PRE
0.9	t^c	0.009399	100	0.009977	100
	t^c_r	0.008983	104.632314	0.007649	130.441398
	t^c_{lr}	0.008140	115.470738	0.007368	135.409878
	t^c_{p1}	0.008165	115.112126	0.007377	135.246341
	t^c_{p2}	0.008143	115.426215	0.007323	136.245249

Continued on next page

Table 4. Continued from previous page

r_{yx}	Estimators	Population 1		Population 2	
	t^c_{p3}	0.005397	174.155787	0.004591	217.296845
	t^c_{p4}	0.004126	227.787861	0.003749	266.095249
	t^c_{p5}	0.001845	509.395864	0.003496	285.359492
	t^c_{p6}	0.001612	582.794752	0.002739	364.269207
0.8	t^c	0.008071	100	0.007391	100
	t^c_r	0.008009	100.781241	0.006988	105.760711
	t^c_{lr}	0.003936	205.057579	0.006895	107.191263
	t^c_{p1}	0.003954	204.104535	0.006832	108.179959
	t^c_{p2}	0.003923	205.726039	0.006898	107.139678
	t^c_{p3}	0.002592	311.383468	0.005186	142.501132
	t^c_{p4}	0.001537	524.918546	0.003622	204.034130
	t^c_{p5}	0.001390	580.591840	0.002845	259.729142
	t^c_{p6}	0.001186	680.007413	0.002617	282.358244
0.7	t^c	0.007355	100	0.009191	100
	t^c_r	0.005736	128.219695	0.008556	107.426159
	t^c_{lr}	0.004616	159.320334	0.004802	191.378071
	t^c_{p1}	0.004687	156.904577	0.004823	190.561527
	t^c_{p2}	0.004643	158.398764	0.004855	189.293929
	t^c_{p3}	0.003559	206.628985	0.002697	340.760512
	t^c_{p4}	0.003389	217.030976	0.002335	393.501322
	t^c_{p5}	0.002699	272.486505	0.001865	492.606328
	t^c_{p6}	0.001856	396.239252	0.001135	809.643520
0.6	t^c	0.009910	100	0.018945	100
	t^c_r	0.008116	122.109429	0.014026	135.069295
	t^c_{lr}	0.008083	122.603248	0.012800	148.004437
	t^c_{p1}	0.008044	123.195799	0.012770	148.354436
	t^c_{p2}	0.008078	122.674271	0.012438	152.320271
	t^c_{p3}	0.005600	176.957222	0.010820	175.097502
	t^c_{p4}	0.005533	179.089460	0.005823	325.314051
	t^c_{p5}	0.004191	236.468491	0.005641	335.814228
	t^c_{p6}	0.004054	244.411288	0.003085	614.006261
0.5	t^c	0.011061	100	0.032002	100
	t^c_r	0.014576	75.886529	0.032978	97.039310
	t^c_{lr}	0.009580	115.454465	0.025516	125.420912
	t^c_{p1}	0.009572	115.560486	0.025929	125.848529
	t^c_{p2}	0.009578	115.488096	0.025836	123.864115
	t^c_{p3}	0.008217	134.614214	0.020097	159.234941
	t^c_{p4}	0.008118	136.249085	0.010095	292.123303
	t^c_{p5}	0.005430	203.696951	0.010087	317.254054
	t^c_{p6}	0.004944	223.693975	0.008502	376.398745
0.3	t^c	0.019186	100	0.036245	100
	t^c_r	0.030482	62.941238	0.055141	65.731736
	t^c_{lr}	0.018669	102.764342	0.035146	103.126333
	t^c_{p1}	0.018532	103.524003	0.035112	103.227368
	t^c_{p2}	0.018933	101.335755	0.035409	102.361822
	t^c_{p3}	0.018019	106.471769	0.033719	107.489961
	t^c_{p4}	0.017812	107.712086	0.032991	109.863842
	t^c_{p5}	0.015881	120.810271	0.029482	122.937858
	t^c_{p6}	0.014371	133.496614	0.027864	130.076404

Table 4 also shows that our proposed estimators perform better than the existing estimators. The MSE of the estimators decreases when the correlation and sample size increases for the population 1 and 2.

8. Conclusions

In this article we have proposed estimators for the population mean in stratified Ranked set sampling using the information of auxiliary variables. The expressions for Bias and MSE of the suggested estimators have been derived up to the first order of approximation. Numerical illustrations and simulation study for comparing the efficiency of the proposed estimators with other estimators have been used. The results have been shown in the Tables 2, 3 and 4. The Tables 2 and 3 show that the proposed estimators turn out to be more efficient as compared to the other estimators for both cases when initial samples are taken by using SRS and PPS respectively.

The proposed estimators are found to be rather improved in terms of lesser MSE and greater PRE as compared to the existing estimators in both real and simulated data sets. It is also observed from the simulation that the MSE of the proposed estimators decreases as the values of the correlation coefficient increase whereas the PRE of the suggested estimators increases as the values of the correlation coefficients increase. Based on our numerical illustrations and simulation study, we can conclude that our proposed estimators can be preferred over the other estimators taken in this paper in several real situations.

[Received: September 2022 — Accepted: May 2023]

References

Al-Omari, Ibrahim, A. & Bouza, C. (2014), 'Review of ranked set sampling: modifications and applications', *Investigación Operacional* **35**(3), 215–235.

Bhushan, S., Gupta, R., Singh, S. & Kumar, A. (2020), 'A new efficient log type class of estimators using auxiliary variable', *International Journal of Statistics and Systems* **15**(1), 19–28.

Bouza, C., Singh, P. & Singh, R. (2018), 'Ranked set sampling and optional scrambling randomized response modeling', *Investigación Operacional* **39**(1), 100–107.

Chaudhary, M., Singh, R., Shukla, R., Kumar, M. & Smarandache, F. (2009), 'A family of estimators for estimating population mean in stratified sampling under non-response', *Pakistan Journal of Statistics and Operation Research* pp. 47–54.

Cochran, W. (1977), *Sampling Techniques*, Wiley Eastern Limited, New Delhi.

Ganeslingam, S. & Ganesh, S. (2006), 'Ranked set sampling versus simple random sampling in the estimation of the mean and the ratio', *Journal of Statistics and Management Systems* **9**(2), 459–472.

Haq, A. & Shabbir, J. (2013), 'Improved family of ratio estimators in simple and stratified random sampling', *Communications in Statistics-Theory and Methods* **42**(5), 782–799.

Kadilar, C. & Cingi, H. (2003), 'Ratio estimators in stratified random sampling', *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **45**(2), 218–225.

Koyuncu, N. & Kadilar, C. (2009), 'Family of estimators of population mean using two auxiliary variables in stratified random sampling', *Communications in Statistics-Theory and Methods* **38**(14), 2398–2417.

McIntyre, G. (1952), 'A method for unbiased selective sampling, using ranked sets', *Australian journal of agricultural research* **3**(4), 385–390.

Muneer, S., Shabbir, J. & Khalil, A. (2017), 'Estimation of finite population mean in simple random sampling and stratified random sampling using two auxiliary variables', *Communications in Statistics-Theory and Methods* **46**(5), 2181–2192.

Samawi, H. & Muttlak, H. (1996), 'Estimation of ratio using rank set sampling', *Biometrical Journal* **38**(6), 753–764.

Samawi, H. & Siam, M. (2003), 'Ratio estimation using stratified ranked set sample', *Metron* **61**(1), 75–90.

Shabbir, J. & Gupta, S. (2006), 'A new estimator of population mean in stratified sampling', *Communications in Statistics-Theory and Methods* **35**(7), 1201–1209.

Shabbir, J. & Gupta, S. (2007), 'On estimating the finite population mean with known population proportion of an auxiliary variable', *Pakistan Journal of Statistics* **23**(1), 1.

Singh, H., Tailor, R. & Singh, S. (2014), 'General procedure for estimating the population mean using ranked set sampling', *Journal of Statistical Computation and Simulation* **84**(5), 931–945.

Singh, R. & Kumar, M. (2012), 'Improved estimators of population mean using two auxiliary variables in stratified random sampling', *Pakistan Journal of Statistics and Operation Research* pp. 65–72.

Takahasi, K. & Wakimoto, K. (1968), 'On unbiased estimates of the population mean based on the sample stratified by means of ordering', *Annals of the institute of statistical mathematics* **20**(1), 1–31.

Appendix

This section consider the proof of the Theorems of Section 4 and 5. To derive the MSE of the proposed estimators, the following notations will be used throughout the paper.

$$\overline{y}_{srss} = \overline{Y}(1+\epsilon_0)$$
$$\overline{x}_{srss} = \overline{X}(1+\epsilon_1)$$
$$V_{rs} = \sum_{h=1}^{L} W_h^{r+s} \frac{E[(\overline{y}_{srss}-\overline{Y})^r(\overline{x}_{srss}-\overline{X})^s]}{\overline{Y}^r\overline{X}^s}$$

such that $E(\epsilon_0) = E(\epsilon_1) = 0$

$$E(\epsilon_0^2) = \sum_{h=1}^{L} W_h^2(\eta_h C_{yh}^2 - D_{yh[i]}^2) = V_{20}$$

$$E(\epsilon_1^2) = \sum_{h=1}^{L} W_h^2(\eta_h C_{xh}^2 - D_{xh[i]}^2) = V_{02}$$

$$E(\epsilon_0\epsilon_1) = \sum_{h=1}^{L} W_h^2(\eta_h C_{xyh} - D_{xyh[i]}) = V_{11}$$

where $\eta_h = \frac{1}{k_h r}$, $C_{xh} = \frac{S_{xh}}{\overline{X}}$, $C_{yh} = \frac{S_{yh}}{\overline{Y}}$, $D_{xh[i]}^2 = \frac{1}{k_h^2 r \overline{X}^2}\sum_{i=1}^{k_h}(\overline{X}_{h(i)} - \overline{X}_h)^2$, $D_{yh[i]}^2 = \frac{1}{k_h^2 r \overline{Y}^2}\sum_{i=1}^{k_h}(\overline{Y}_{h[i]} - \overline{Y}_h)^2$ and $D_{xyh[i]} = \frac{1}{k_h^2 r \overline{Y}\overline{X}}\sum_{i=1}^{k_h}(\overline{Y}_{h[i]} - \overline{Y}_h)(\overline{X}_{h(i)} - \overline{X}_h)$ where $\overline{Y}_{h[i]}$ and $\overline{X}_{h(i)}$ are the means of the $i^t h$ is ranked set and are given by

$$\overline{Y}_{h[i]} = \frac{1}{r}\sum_{j=1}^{r} Y_{h[i]j}, \overline{X}_{h(i)} = \frac{1}{r}\sum_{j=1}^{r} X_{h(i)j}$$

Now, consider the estimator

$$t_{p1}^c = \overline{y}_{SRSS} \exp\left(\alpha_1\left(\frac{\overline{x}_{SRSS}}{\overline{X}} - 1\right)\right)$$

Using the above notations we have

$$t_{p1}^c = \overline{Y}(1+\epsilon_0) \exp\left(\alpha_1\left(\frac{\overline{X}(1+\epsilon_1)}{\overline{X}} - 1\right)\right)$$

The bias of the estimator t_{p1}^c is given by

$$Bias(t_{p1}^c) = \overline{Y}\left(\frac{\alpha_1^2}{2}V_{02} + \alpha_1 V_{11}\right)$$

The MSE of the estimator t_{p1}^c is given by

$$MSE(t_{p1}^c) = \overline{Y}^2\left(V_{20} + \alpha_1^2 V_{02} + 2\alpha_1 V_{11}\right) \tag{A1}$$

To find out the minimum MSE for t^c_{p1}, we partially differentiate equation (A1) w.r.t. α_1 and equating to zero we get

$$\alpha^*_1 = -\frac{V_{11}}{V_{02}}$$

Putting the optimum value of α_1 in the equation (A1), we get a minimum MSE of t^c_{p1} as

$$MinMSE(t^c_{p1}) = \overline{Y}^2\left(V_{20} - \frac{V^2_{11}}{V_{02}}\right)$$

Similarly, we can obtain the optimum values of constants and minimum MSEs of other proposed estimators which are given as

$$t^c_{p2} = \overline{Y}(1+\epsilon_0)\exp\left(\alpha_2 \log\frac{\overline{X}(1+\epsilon_1)}{\overline{X}}\right)$$

The bias of the estimator t^c_{p2} is given by

$$Bias(t^c_{p2}) = \overline{Y}\left(\frac{(\alpha^2_2 - \alpha_2)}{2}V_{02} + \alpha_2 V_{11}\right)$$

The MSE of the estimator t^c_{p2} is given by

$$MSE(t^c_{p2}) = \overline{Y}^2\left(V_{20} + \alpha^2_2 V_{02} + 2\alpha_2 V_{11}\right) \quad \text{(A2)}$$

To find out the minimum MSE for t^c_{p2}, we partially differentiate equation (A2) w.r.t. α_2 and equating to zero we get

$$\alpha^*_2 = -\frac{V_{11}}{V_{02}}$$

Putting the optimum value of α_2 in the equation (A2), we get a minimum MSE of t^c_{p2} as

$$MinMSE(t^c_{p2}) = \overline{Y}^2\left(V_{20} - \frac{V^2_{11}}{V_{02}}\right)$$

$$t^c_{p3} = [(1+w_1)\overline{Y}(1+\epsilon_0) + w_2\epsilon_1](1-\epsilon_1+\epsilon^2_1)$$

$$t^c_{p3} - \overline{Y} = \overline{Y}[(\epsilon_0 + W_1 + \epsilon_0 w_1 - \epsilon_1 - \epsilon_1 w_1 - \epsilon_0\epsilon_1 - \epsilon_0\epsilon_1 w_1 + \epsilon^2_1 + w_1\epsilon^2_1) - w_2\delta(\epsilon_1 - \epsilon^2_1)]$$

The bias of the estimator t^c_{p3} is given by

$$Bias(t^c_{p3}) = \overline{Y}w_1 + \overline{Y}(V_{02} + w_1 V_{02} + w_2\delta V_{02} - V_{11} - w_1 V_{11})$$

The MSE of the estimator t^c_{p3} is given by

$$\begin{aligned} MSE(t^c_{p3}) = \overline{Y}^2(&V_{20} + V_{02} - 2V_{11} + w^2_1(1 + V_{20} + 3V_{02} - 4V_{11}) + w^2_2\delta^2 V_{02} \\ &+ 2w_1(V_{20} + 2V_{02} - 3V_{11}) - 2w_2\delta(V_{11} - V_{02}) - 2w_1w_2\delta(V_{11} - 2V_{02})) \end{aligned}$$

$$MSE(t_{p3}^c) = \overline{Y}^2(A_1 + w_1^2 B_1 + w_2^2 C_1 + 2w_1 D_1 - 2w_2 E_1 - 2w_1 w_2 F_1) \tag{A3}$$

where

$$A_1 = V_{20} + V_{02} - 2V_{11}$$
$$B_1 = 1 + V_{20} + 3V_{02} - 4V_{11}$$
$$C_1 = \delta^2 V_{02}, \delta = \frac{\overline{X}}{\overline{Y}}$$
$$D_1 = V_{20} + 2V_{02} - 3V_{11}$$
$$E_1 = \delta(V_{02} - V_{11})$$
$$F_1 = \delta(V_{02} - 2V_{11})$$

To find out the minimum MSE for t_{p3}^c, we partially differentiate equation (A3) w.r.t. w_1 and w_2 and equating to zero we get

$$w_1^* = \frac{C_1 D_1 - E_1 F_1}{F_1^2 - B_1 C_1}$$

$$w_2^* = \frac{D_1 F_1 - B_1 C_1}{F_1^2 - B_1 C_1}$$

Putting the optimum values of w_1 and w_2 in the equation (A3), we get a minimum MSE of t_{p3}^c as

$$MinMSE(t_{p3}^c) = \overline{Y}^2\left(A_1 + \frac{C_1 D_1^2 + B_1 E_1^2 - 2D_1 E_1 F_1}{F_1^2 - B_1 C_1}\right)$$

$$t_{p4}^c = [(1 + w_3)\overline{Y}(1 + \epsilon_0) + w_4\epsilon_1]\left(1 - \frac{3}{2}\epsilon_1 + \frac{15}{8}\epsilon_1^2\right)$$

$$t_{p4}^c - \overline{Y} = \overline{Y}[(\epsilon_0 + W_3 + \epsilon_0 w_3 - \frac{1}{2}\epsilon_1 - \frac{1}{2}\epsilon_1 w_3 - \frac{1}{2}\epsilon_0\epsilon_1 - \frac{1}{2}\epsilon_0\epsilon_1 w_3 + \frac{3}{8}\epsilon_1^2 + \frac{3}{8}w_3\epsilon_1^2) - w_4\delta(\epsilon_1 - \epsilon_1^2)]$$

The bias of the estimator t_{p4}^c is given by

$$bias(t_{p4}^c) = \overline{Y}w_3 + \overline{Y}\left(\frac{3}{8}V_{02} + \frac{3}{8}w_3 V_{02} + \frac{1}{2}w_4\delta V_{02} - \frac{1}{2}V_{11} - \frac{1}{2}w_3 V_{11}\right)$$

The MSE of the estimator t_{p4}^c is given by

$$MSE(t_{p4}^c) = \overline{Y}^2(V_{20} + \frac{1}{4}V_{02} - V_{11} + w_3^2(1 + V_{20} + V_{02} - 2V_{11}) + w_4^2\delta^2 V_{02} + 2w_3(V_{20} + \frac{5}{4}V_{02} - frac32V_{11})2w_4\delta(V_{11} - \frac{1}{2}V_{02}) - 2w_3 w_4\delta(V_{11} - V_{02}))$$

$$MSE(t^c_{p4}) = \overline{Y}^2(A_2 + w_3^2B_2 + w_4^2C_2 + 2w_3D_2 - 2w_4E_2 - 2w_3w_4F_2) \qquad \text{(A4)}$$

where

$$A_2 = V_{20} + \frac{1}{4}V_{02} - V_{11}$$

$$B_2 = 1 + V_{20} + V_{02} - 2V_{11}$$

$$C_2 = \delta^2 V_{02}, \delta = \frac{\overline{X}}{\overline{Y}}$$

$$D_2 = V_{20} + \frac{5}{4}V_{02} - \frac{3}{2}V_{11}$$

$$E_2 = \delta\left(V_{02} - \frac{1}{2}V_{11}\right)$$

$$F_2 = \delta(V_{02} - V_{11})$$

To find out the minimum MSE for t^c_{p4}, we partially differentiate equation w.r.t. (A4) w_3 and w_4 and equating to zero we get

$$w_3^* = \frac{C_2D_2 - E_2F_2}{F_2^2 - B_2C_2}$$

$$w_4^* = \frac{D_2F_2 - B_2C_2}{F_2^2 - B_2C_2}$$

Putting the optimum values of w_3 and w_4 in the equation (A4), we get a minimum MSE of t^c_{p4} as

$$minMSE(t^c_{p4}) = \overline{Y}^2\left(A_2 + \frac{C_2D_2^2 + B_2E_2^2 - 2D_2E_2F_2}{F_2^2 - B_2C_2}\right)$$

$$t^c_{p5} = w_5\overline{Y}(1 + \epsilon_0) + w_6 \exp\left(\frac{-\epsilon_1}{2 + \epsilon_1}\right)(1 + log(1 + \epsilon_1))$$

$$t^c_{p5} - \overline{Y} = (w_5 - 1)\overline{Y} + w_5\overline{Y}\epsilon_0 + w_6\left(1 + \frac{\epsilon_1}{2} - \frac{5}{8}\epsilon_1^2\right)$$

$$Bias(t^c_{p5}) = (w_5 - 1)\overline{Y} + w_6\left(1 - \frac{5}{8}V_{02}\right)$$

$$t^c_{p6} = w_7\overline{Y}(1 + \epsilon_0) + w_8 \exp\left(\frac{-\epsilon_1}{2 + \epsilon_1}\right)(1 + \epsilon_1)^{-1}$$

$$t^c_{p6} - \overline{Y} = (w_7 - 1)\overline{Y} + w_7\overline{Y}\epsilon_0 + w_8\left(1 - \frac{3}{2}\epsilon_1 - \frac{15}{8}\epsilon_1^2\right)$$

$$Bias(t^c_{p6}) = (w_7 - 1)\overline{Y} + w_6\left(1 + \frac{15}{8}V_{02}\right)$$

Case 1: Sum of Weights is Unity $(w_5 + w_6 = 1$ **&** $w_7 + w_8 = 1)$. The MSE of the estimator $t_(p5^c)$ is given by

$$MSE(t_{p5}^c) = \overline{Y}^2(V_{20} + w_6^2 V_{02} - 2w_6 V_{11}) \tag{A5}$$

To find out the minimum MSE for t_{p5}^c, we partially differentiate equation (A5) w.r.t. w_6, and equating to zero we get

$$w_6^* = \frac{V_{11}}{V_{02}}$$

Putting the optimum value of w_6 in the equation (A5), we get a minimum MSE of t_{p5}^c as

$$MinMSE(t_{p5}^c) = \overline{Y}^2\left(V_{20} - \frac{V_{11}^2}{V_{02}}\right)$$

The MSE of the estimator t_{p6}^c is given by

$$MSE(t_{p6}^c) = \overline{Y}^2(V_{20} + w_8^2 V_{02} - 2w_8 V_{11}) \tag{A6}$$

To find out the minimum MSE for t_{p6}^c, we partially differentiate equation (A6) w.r.t. w_8, and equating to zero we get

$$w_8^* = \frac{V_{11}}{V_{02}}$$

Putting the optimum value of w_8 in the equation (A6), we get a minimum MSE of t_{p6}^c as

$$MinMSE(t_{p6}^c) = \overline{Y}^2\left(V_{20} - \frac{V_{11}^2}{V_{02}}\right)$$

Case 2: Sum of Weights is Flexible $(w_5 + w_6 \neq 1$ **&** $w_7 + w_8 \neq 1)$

$$t_{p5}^c - \overline{Y} = (W_5 - 1)\overline{Y} + w_5\overline{Y}\epsilon_0 + w_6\left(1 + \frac{\epsilon_1}{2} - \frac{5}{8}\epsilon_1^2\right)$$

Squaring on both sides we get

$$\begin{aligned}(t_{p5}^c - \overline{Y})^2 = \overline{Y}^2 + \overline{Y}^2 w_5^2(1+\epsilon_0^2) + w_6^2(1-\epsilon_1^2) - 2w_5\overline{Y}^2 - 2w_6\overline{Y}\left(1 - \frac{5}{8}\epsilon_1^2\right) \\ + 2w_5w_6\left(1 - \frac{5}{8}\epsilon_1^2 + \frac{1}{2}\epsilon_0\epsilon_1\right)\end{aligned}$$

Taking expectations on both sides we get

$$\begin{aligned}MSE(t_{p5}^c) = \overline{Y}^2 + \overline{Y}^2 w_5^2(1+V_{20}) + w_6^2(1-V_{02}) - 2w_5\overline{Y}^2 \\ - 2w_6\overline{Y}\left(1 - \frac{5}{8}V_{02}\right) + 2w_5w_6\left(1 - \frac{5}{8}V_{02} + \frac{1}{2}V_{11}\right)\end{aligned}$$

$$MSE(t_{p5}^c) = C_3 + w_5^2 A_3 + w_6^2 B_3 - 2w_5 C_3 - 2w_6 D_3 + 2w_5 w_6 E_3 \tag{A7}$$

where

$$A_3 = \overline{Y}^2 (1 + V_{20})$$

$$B_3 = 1 - V_{02}$$

$$C_3 = \overline{Y}^2$$

$$D_3 = \overline{Y} \left(1 - \frac{5}{8} V_{02}\right)$$

$$E_3 = \overline{Y} \left(1 - \frac{5}{8} V_{02} + \frac{1}{2} V_{11}\right)$$

To find out the minimum MSE for the estimator t_{p5}^c, we partially differentiate equation (21) w.r.t. w_5 and w_6 and equating to zero we get

$$w_5^* = \frac{B_3 C_3 - D_3 E_3}{A_3 B_3 - E_3^2}$$

$$w_6^* = \frac{A_3 D_3 - C_3 E_3}{A_3 B_3 - E_3^2}$$

Putting the optimum values of w_5 and w_6 in the equation (21), we get a minimum MSE of t_{p5}^c as

$$MinMSE(t_{p5}^c) = C_3 + \frac{B_3 C_3^2 + A_3 D_3^2 - 2C_3 D_3 E_3}{E_3^2 - A_3 B_3}$$

$$t_{p6}^c - \overline{Y} = (W_7 - 1)\overline{Y} + w_7 \overline{Y} \epsilon_0 + w_8 \left(1 - \frac{3}{2}\epsilon_1 + \frac{15}{8}\epsilon_1^2\right)$$

Squaring on both sides we get

$$(t_{p6}^c - \overline{Y})^2 = \overline{Y}^2 + \overline{Y}^2 w_7^2 (1 + \epsilon_0^2) + w_8^2 (1 + 6\epsilon_1^2) - 2w_7 \overline{Y}^2 - 2w_8 \overline{Y} \left(1 - \frac{15}{8}\epsilon_1^2\right)$$
$$+ 2w_7 w_8 \left(1 + \frac{15}{8}\epsilon_1^2 - \frac{3}{2}\epsilon_0 \epsilon_1\right)$$

Taking expectations on both sides we get

$$MSE(t_{p6}^c) = \overline{Y}^2 + \overline{Y}^2 w_7^2 (1 + V_{20}) + w_8^2 (1 + 6V_{02}) - 2w_7 \overline{Y}^2 - 2w_8 \overline{Y} \left(1 + \frac{15}{8} V_{02}\right)$$
$$+ 2w_7 w_8 \left(1 + \frac{15}{8} V_{02} - \frac{3}{2} V_{11}\right)$$

$$MSE(t_{p6}^c) = C_4 + w_7^2 A_4 + w_8^2 B_4 - 2w_7 C_4 - 2w_8 D_4 + 2w_7 w_8 E_4 \tag{A8}$$

where

$$A_4 = \overline{Y}^2 (1 + V_{20})$$

$$B_4 = 1 + 6V_{02}$$
$$C_4 = \overline{Y}^2$$
$$D_4 = \overline{Y}\left(1 + \frac{15}{8}V_{02}\right)$$
$$E_4 = \overline{Y}\left(1 + \frac{15}{8}V_{02} - \frac{3}{2}V_{11}\right)$$

To find out the minimum MSE for the estimator t^c_{p6}, we partially differentiate equation (22) w.r.t. w_7 and w_8 and equating to zero we get

$$w^*_7 = \frac{B_4C_4 - D_4E_4}{A_4B_4 - E^2_4}$$
$$w^*_8 = \frac{A_4D_4 - C_4E_4}{A_4B_4 - E^2_4}$$

Putting the optimum values of w_7 and w_8 in the equation (22), we get a minimum MSE of t^c_{p6} as

$$minMSE(t^c_{p6}) = C_4 + \frac{B_4C^2_4 + A_4D^2_4 - 2C_4D_4E_4}{E^2_4 - A_4B_4}$$