Filomat 38:2 (2024), 599–608
https://doi.org/10.2298/FIL2402599C

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Calibration estimator of population mean in stratified extreme ranked set sampling with simulation study

Arzu Ece Çetin[a,c,*], **Nursel Koyuncu**[b]

[a]*Graduate School of Science and Engineering, Hacettepe University, Beytepe, Ankara, Turkey*
[b]*Department of Statistics, Hacettepe University, Beytepe, Ankara 06800, Turkey*
[c] *Department of Management, Gebze Technical University, Kocaeli 41400, Turkey*

Abstract. Calibration estimation sets the original weights to include the known population characteristics of auxiliary variables using constraints. In this article, we have proposed a new calibration estimator of the population mean in stratified extreme ranked set sampling design, which is more efficient and cost-effective design against other sampling designs in the literature. A detailed simulation study is carried out to observe the performance of proposed estimators. We have used the information of auxiliary variable to avoid ranking errors in our simulations. We have created samples from a bi-variate normal distribution with different values of ρ_{xy}. While one of these variables is taken as the variable of interest, the other is accepted as an auxiliary variable and used in ranking the sample units in each set. As a result of the simulation study using both synthetic and real data sets, we have found that our proposed estimators are more efficient than Sinha et al. [19] calibration estimator and classical stratified estimator.

1. Introduction

The calibration approach is a technique that uses auxiliary information to increase the accuracy of the estimators of population parameters of study variable in sampling literature. Calibration has gained importance in the field of sampling literature in recent years. The definition of calibration was introduced by Deville and Sarndal [3] in simple random sampling (SRS). After, many authors studied calibration estimation using different calibration constraints in survey sampling such as Kim and Park [4], Kim et al. [5], Mouhamed et al. [14], Koyuncu and Kadilar ([7], [8], [9]), Koyuncu [10], Sinha et al. [19] and Tracy et al. [22]. Singh et al. [20] studied calibration estimators in presence non-sampling errors. Zaman and Bulut [23] proposed robust calibration estimators. On the other side some authors extended calibration approach to the ranked set sampling (RSS) designs. Koyuncu [10] proposed calibration estimator of population mean under stratified ranked set sampling (SRSS) design. Shahzad et al. [17] have studied calibration estimators under median ranked set sampling. Shahzad et al. [18] proposed some new calibration estimators to estimate the population variance of Covid using L-moments under double-stratified random sampling. In sampling literature, authors also proposed new ranked set sampling designs and try to improve estimation methods. Al-Omari et al. [1] suggested stratified percentile ranked set sampling design for estimating the

2020 *Mathematics Subject Classification*. Primary 62D05.
Keywords. Calibration; Stratified extreme ranked set sampling; Auxiliary information; Efficiency; GREG estimator.
Received: 26 July 2023; Revised: 10 August 2023; Accepted: 18 August 2023
Communicated by Maria Alessandra Ragusa
* Corresponding author: Arzu Ece Çetin
Email addresses: arzuece.dogru@hacettepe.edu.tr (Arzu Ece Çetin), nkoyuncu@hacettepe.edu.tr (Nursel Koyuncu)

population mean. Mandowara and Mehta [12] studied modified ratio estimators in stratified ranked set sampling. Taconelli and Bonat [21] investigated the performance of six alternative estimation methods to MLE for parameter estimation under RSS.

Stratified extreme ranked set sampling design (SERSS) was proposed by Samawi and Saeid [16]. SERSS is more practical than ordinary RSS because we only need to successfully determine the first and/or last ordered unit in case of even sample size for each stratum or the median unit in case of odd sample size. In this study, moving direction from less error of ranking units and practicality we have studied calibration approach under SERSS. We have conducted a detailed simulation study with real and synthetic datasets to see the performance of the proposed estimators.

2. Proposed calibration estimators in SERSS

ERSS was investigated by Samawi et al. [15]. To predict the finite population mean μ_Y using ERSS, the operation can be explained briefly as follows:

1. The process includes drawing sets of each r units randomly from population for which the mean is to be predicted. The most important assumption is the smallest and the biggest units of the set can be fixed visually or with a little cost.
2. The lowest ranked unit is determined from the first r unit set. Then, the largest ranked unit is determined from the second r unit set. And the lowest ranked unit is determined from the third set of r units and so on. Thus, the first $(r-1)$ determined units is obtained using the first $(r-1)$ sets. The event of choosing the $r-th$ unit from the $r-th$ (i.e very last) set depends on whether r is odd or even.
3. When r is even, the measurement value of the largest unit ranked is measured.
4. Two options exist when r is odd:
   i. The average of the largest and lowest units in the $r-th$ set is measured for the measure of the $r-th$ unit.
   ii. The measure of the median for the measure of the $r-th$ unit is measured.
5. This procedure complete one cycle of ERSS. The period may be repeated m times until n elements of desired to obtain.

In stratified sampling the population of N units is first divided into L subpopulations of $N_1, N_2, \cdots, N_L$ units, respectively. These subpopulations are no overlapping and together they comprise the whole population, so that $N_1 + N_2 + \cdots + N_L = N$. The subpopulations are named strata. To obtain the full benefit from stratification, the values of the N_h $(h = 1, 2, \cdots, L)$ must be known. After the stratas have been determined, a sample is drawn from each, the drawings being made in different strata. The sample sizes within the strata are denoted by $n_1, n_2, \cdots, n_L$ respectively. If r_h is even then $X_{h1(1)}, X_{h2(r_h)}, X_{h3(1)} \cdots, X_{h\{r_h-1\}(1)}, X_{hr_h(r_h)}$ denotes the ERSSh(e) for the h-th stratum. If r_h is odd then $X_{h1(1)}, X_{h2(r_h)}, X_{h3(1)} \cdots, X_{h\{r_h-1\}(r_h)}, X_{hr_h(\frac{r_h+1}{2})}$ denotes the ERSSh(o) for the h-th stratum. The resulting L independent ERSSs comprise the SERSS. Assume that there is (a) strata with even set size and $(L-a)$ strata with odd set size SERSS of size n. For simplicity of notation, let $(m = 1)$ then $n_h = r_h$ and $(n = r)$, then the estimate of the mean using SERSS is given by

$$\bar{y}_{st(SERSS)} = \sum_{h=1}^{a} W_h \bar{y}_{(ERSS)h(e)} + \sum_{h=a+1}^{L} W_h \bar{y}_{(ERSS)h(o)}, \tag{1}$$

$$\bar{x}_{st[SERSS]} = \sum_{h=1}^{a} W_h \bar{x}_{[ERSS]h(e)} + \sum_{h=a+1}^{L} W_h \bar{x}_{[ERSS]h(o)} \tag{2}$$

where $W_h = \frac{N_h}{N}$ is the stratum weight and

$$\bar{y}_{(ERSS)h(e)} = \frac{1}{2}\left(\bar{Y}_{\mathrm{h(1)}} + \bar{Y}_{\mathrm{h(r_h)}}\right),$$

$$\bar{y}_{(ERSS)h(o)} = \frac{Y_{h1(1)} + Y_{h2(r_h)} + Y_{h3(1)} + ... + Y_{h\{r_h-1\}(r_h)} + Y_{hr_h\left(\frac{r_h+1}{2}\right)}}{r_h},$$

$$\bar{Y}_{h(1)} = 2\sum_{i=1}^{\frac{r_h}{2}} \frac{Y_{h\{2i-1\}(1)}}{r_h},$$

$$\bar{Y}_{h(r_h)} = 2\sum_{i=1}^{\frac{r_h}{2}} \frac{Y_{h(2i)(r_h)}}{r_h},$$

$$\bar{x}_{[ERSS]h(e)} = \frac{1}{2}\left(\bar{X}_{\mathrm{h[1]}} + \bar{X}_{\mathrm{h[r_h]}}\right),$$

$$\bar{x}_{[ERSS]h(o)} = \frac{X_{h1[1]} + X_{h2[r_h]} + X_{h3[1]} + ... + X_{h\{r_h-1\}[r_h]} + X_{hr_h\left[\frac{r_h+1}{2}\right]}}{r_h},$$

$$\bar{X}_{h[1]} = 2\sum_{i=1}^{\frac{r_h}{2}} \frac{X_{h\{2i-1\}[1]}}{r_h},$$

$$\bar{X}_{h[r_h]} = 2\sum_{i=1}^{\frac{r_h}{2}} \frac{X_{h\{2i\}[r_h]}}{r_h}.$$

It can be shown that

$$E\left(\bar{y}_{(ERSS)h(i)}\right) = \begin{cases} \frac{1}{2}\left(\mu_{Yh(1)} + \mu_{Yh(r_h)}\right), \ i = e \\ \left(\frac{r_h - 1}{2r_h}\right)\left(\mu_{Yh(1)} + \mu_{Yh(r_h)}\right) + \frac{1}{r_h}\mu_{Yh\left(\frac{r_h+1}{2}\right)}, \ i = o. \end{cases} \tag{3}$$

$$E\left(\bar{x}_{[ERSS]h(i)}\right) = \begin{cases} \frac{1}{2}\left(\mu_{Xh[1]} + \mu_{Xh[r_h]}\right), \ i = e \\ \left(\frac{r_h - 1}{2r_h}\right)\left(\mu_{Xh[1]} + \mu_{Xh[r_h]}\right) + \frac{1}{r_h}\mu_{Xh\left[\frac{r_h+1}{2}\right]}, \ i = o. \end{cases} \tag{4}$$

$$E\left(\bar{y}_{st(SERSS)}\right) = \frac{1}{2}\sum_{h=1}^{L} W_h\left(\mu_{Yh(1)} + \mu_{Yh(r_h)}\right) + \frac{1}{2}\sum_{h=a+1}^{L}\frac{W_h}{r_h}[2\mu_{Yh\left(\frac{r_h+1}{2}\right)} - \left(\mu_{Yh(1)} + \mu_{Yh(r_h)}\right)] \tag{5}$$

$$E\left(\bar{x}_{st[SERSS]}\right) = \frac{1}{2}\sum_{h=1}^{L} W_h\left(\mu_{Xh[1]} + \mu_{Xh[r_h}\right) + \frac{1}{2}\sum_{h=a+1}^{L}\frac{W_h}{r_h}[2\mu_{Xh\left[\frac{r_h+1}{2}\right]} - \left(\mu_{Xh[1]} + \mu_{Xh[r_h]}\right)] \tag{6}$$

Therefore, the mean and variance of are $\bar{y}_{st(SERSS)}$ and $\bar{x}_{st[SERSS]}$ given as

$$Var\left(\bar{y}_{st(SERSS)}\right) = \frac{1}{2}\sum_{h=1}^{L}\frac{W_h^2}{r_h}\left(S^2_{Yh(1)} + S^2_{Yh(r_h)}\right) + \frac{1}{2}\sum_{h=a+1}^{L}\frac{W_h^2}{r_h^2}\left[2S^2_{Yh\left(\frac{r_h+1}{2}\right)} - \left(S^2_{Yh(1)} + S^2_{Yh(r_h)}\right)\right], \tag{7}$$

$$Var\left(\bar{x}_{st[SERSS]}\right) = \frac{1}{2}\sum_{h=1}^{L}\frac{W_h^2}{r_h}\left(S^2_{Xh[1]} + S^2_{Xh[r_h]}\right) + \frac{1}{2}\sum_{h=a+1}^{L}\frac{W_h^2}{r_h^2}\left[2S^2_{Xh\left[\frac{r_h+1}{2}\right]} - \left(S^2_{Xh[1]} + S^2_{Xh[r_h]}\right)\right] \tag{8}$$

If the underlying distribution for each stratum is symmetric then it can be shown that $E\left(\bar{y}_{st(SERSS)}\right) = \mu_Y$ and $E\left(\bar{x}_{st[SERSS]}\right) = \mu_X$,

$$Var\left(\bar{y}_{st(SERSS)}\right) = \sum_{h=1}^{L} \frac{W_h^2}{r_h} S_{Yh(1)}^2 + \sum_{h=a+1}^{L} \frac{W_h^2}{r_h^2} \left[S^2_{Yh\left(\frac{r_h+1}{2}\right)} - S^2_{Yh(1)} \right], \tag{9}$$

$$Var\left(\bar{x}_{st[SERSS]}\right) = \sum_{h=1}^{L} \frac{W_h^2}{r_h} S_{Xh[1]}^2 + \sum_{h=a+1}^{L} \frac{W_h^2}{r_h^2} \left[S^2_{Xh\left[\frac{r_h+1}{2}\right]} - S^2_{Xh[1]} \right]. \tag{10}$$

We propose a new calibration estimator for the SERSS given by

$$\bar{y}_{st}\left(EN\right) = \sum_{h=1}^{a} \Omega_h \bar{y}_{ERSSh(e)} + \sum_{h=a+1}^{L} \Omega_h \bar{y}_{ERSSh(o)}, \tag{11}$$

where Ω_h are suitable calibrated weights under SERSS. The chi-square type distance function

$$\sum_{h=1}^{L} \frac{\left(\Omega_h - W_h\right)^2}{q_h W_h} \tag{12}$$

is minimized subject to two calibration constraints For more details kindly see Li and Ma [11])

$$\sum_{h=1}^{L} \Omega_h = \sum_{h=1}^{L} W_h = 1 \tag{13}$$

$$\sum_{h=1}^{L} \Omega_h \bar{x}_{[ERSS]h} = \sum_{h=1}^{L} W_h \mu_{xh} \tag{14}$$

The Lagrange function is given as

$$\Delta_{SERSS(i)} = \begin{cases} \sum_{h=1}^{a} \frac{\left(\Omega_h - W_h\right)^2}{q_h W_h} - 2\lambda_{0SERSS(i)} \left(\sum_{h=1}^{a} \Omega_h - \sum_{h=1}^{a} W_h \right) - 2\lambda_{1SERSS(i)} \left(\sum_{h=1}^{a} \Omega_h \bar{x}_{[ERSS]h(i)} - \sum_{h=1}^{a} W_h \mu_{xh} \right) \\ \sum_{h=a+1}^{L} \frac{\left(\Omega_h - W_h\right)^2}{q_h W_h} - 2\lambda_{0SERSS(i)} \left(\sum_{h=a+1}^{L} \Omega_h - \sum_{h=a+1}^{L} W_h \right) - 2\lambda_{1SERSS(i)} \left(\sum_{h=a+1}^{L} \Omega_h \bar{x}_{[ERSS]h(i)} - \sum_{h=a+1}^{L} W_h \mu_{xh} \right). \end{cases} \tag{15}$$

where λ_{0SERSS} and λ_{1SERSS} are Lagrange multiplier and $i = e, o$.

Setting $\frac{\partial \Delta_{SERSS(i)}}{\partial \Omega_h} = 0$

Differentiating Δ_{SERSS} according to calibration weights we get

$$\Omega_h = \begin{cases} \sum_{h=1}^{a} W_h + Q_h W_h \left(\lambda_{1SERSS(i)} \bar{x}_{[ERSS]h(i)} + \lambda_{0SERSS(i)} \right), \ i = e \\ \sum_{h=a+1}^{L} W_h + Q_h W_h \left(\lambda_{1SERSS(i)} \bar{x}_{[ERSS]h(i)} + \lambda_{0SERSS(i)} \right), \ i = o. \end{cases} \tag{16}$$

Putting these weights into equations (13) and (14) we have

$$\lambda_{0SERSS}=\begin{cases}-\dfrac{\left(\sum\limits_{h=1}^{a}W_h\left(\mu_{X_h}-\bar{x}_{[ERSS]h(i)}\right)\right)\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{x}_{[ERSS]h(i)}\right)}{\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{x}^2_{[ERSS]h(i)}\right)\left(\sum\limits_{h=1}^{a}Q_hW_h\right)-\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{x}_{[ERSS]h(i)}\right)^2},\ i=e\\ -\dfrac{\left(\sum\limits_{h=a+1}^{L}W_h\left(\mu_{X_h}-\bar{x}_{[ERSS]h(i)}\right)\right)\left(\sum\limits_{h=a+1}^{L}Q_hW_h\bar{x}_{[ERSS]h(i)}\right)}{\left(\sum\limits_{h=a+1}^{L}Q_hW_h\bar{x}^2_{[ERSS]h(i)}\right)\left(\sum\limits_{h=a+1}^{L}Q_hW_h\right)-\left(\sum\limits_{h=a+1}^{L}Q_hW_h\bar{x}_{[ERSS]h(i)}\right)^2},\ i=o.\end{cases}\tag{17}$$

$$\lambda_{1SERSS}=\begin{cases}\dfrac{\left(\sum\limits_{h=1}^{a}W_h\left(\mu_{X_h}-\bar{x}_{[ERSS]h(i)}\right)\right)\left(\sum\limits_{h=1}^{a}Q_hW_h\right)}{\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{x}^2_{[ERSS]h(i)}\right)\left(\sum\limits_{h=1}^{a}Q_hW_h\right)-\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{x}_{h[ERSS]h(i)}\right)^2},\ i=e\\ \dfrac{\left(\sum\limits_{h=a+1}^{L}W_h\left(\mu_{X_h}-\bar{x}_{[ERSS]h(i)}\right)\right)\left(\sum\limits_{h=a+1}^{L}Q_hW_h\right)}{\left(\sum\limits_{h=a+1}^{L}Q_hW_h\bar{x}^2_{[ERSS]h(i)}\right)\left(\sum\limits_{h=a+1}^{L}Q_hW_h\right)-\left(\sum\limits_{h=a+1}^{L}Q_hW_h\bar{x}_{h[ERSS]h(i)}\right)^2},\ i=o.\end{cases}\tag{18}$$

When we put these lambdas in equation (16) we get following calibration weight as

$$\Omega_h=\begin{cases}W_h+Q_hW_h\dfrac{\left(\sum\limits_{h=1}^{a}Q_hW_h\right)\bar{x}_{[ERSS]h(i)}-\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{x}_{[ERSS]h(i)}\right)}{\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{x}^2_{[ERSS]h(i)}\right)\left(\sum\limits_{h=1}^{a}Q_hW_h\right)-\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{x}_{[ERSS]h(i)}\right)^2}\left(\sum\limits_{h=1}^{a}W_h\left(\mu_{X_h}-\bar{x}_{[ERSS]h(i)}\right)\right),i=e\\ W_h+Q_hW_h\dfrac{\left(\sum\limits_{h=a+1}^{L}Q_hW_h\right)\bar{x}_{[ERSS]h(i)}-\left(\sum\limits_{h=a+1}^{L}Q_hW_h\bar{x}_{[ERSS]h(i)}\right)}{\left(\sum\limits_{h=a+1}^{L}Q_hW_h\bar{x}^2_{[ERSS]h(i)}\right)\left(\sum\limits_{h=a+1}^{L}Q_hW_h\right)-\left(\sum\limits_{h=a+1}^{L}Q_hW_h\bar{x}_{[ERSS]h(i)}\right)^2}\left(\sum\limits_{h=a+1}^{L}W_h\left(\mu_{X_h}-\bar{x}_{[ERSS]h(i)}\right)\right),i=o.\end{cases}\tag{19}$$

If we put these weights in equation (11) we have new calibration estimators under SERSS as follows:

$$\bar{y}_{st}\left(EN\right)=\bar{y}_{st}\left(EN\right)_{(e)}+\bar{y}_{st}\left(EN\right)_{(o)}\tag{20}$$

where

$$\begin{aligned}\bar{y}_{st}(EN)_{(e)}&=\sum_{h=1}^{a}W_h\bar{y}_{(ERSS)h(i)}\\&+\frac{\left(\sum\limits_{h=1}^{a}Q_hW_h\right)\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{y}_{(ERSS)h(i)}\bar{x}_{[ERSS]h(i)}\right)-\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{x}_{[ERSS]h(i)}\right)\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{y}_{(ERSS)h(i)}\right)}{\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{x}^2_{[ERSS]h(i)}\right)\left(\sum\limits_{h=1}^{a}Q_hW_h\right)-\left(\sum\limits_{h=1}^{a}Q_hW_h\bar{x}_{[ERSS]h(i)}\right)^2}\left(\sum_{h=1}^{a}W_h\left(\mu_{X_h}-\bar{x}_{[ERSS]h(i)}\right)\right)\end{aligned}$$

$$\bar{y}_{st}(EN)_{(o)} = \sum_{h=a+1}^{L} W_h \bar{y}_{(ERSS)h(i)}$$
$$+ \frac{\left(\sum\limits_{h=a+1}^{L} Q_h W_h\right)\left(\sum\limits_{h=a+1}^{L} Q_h W_h \bar{y}_{(ERSS)h(i)} \bar{x}_{[ERSS]h(i)}\right) - \left(\sum\limits_{h=a+1}^{L} Q_h W_h \bar{x}_{[ERSS]h(i)}\right)\left(\sum\limits_{h=1}^{L} Q_h W_h \bar{y}_{(ERSS)h(i)}\right)}{\left(\sum\limits_{h=a+1}^{L} Q_h W_h \bar{x}^2_{[ERSS]h(i)}\right)}$$
$$\times \left(\sum_{h=a+1}^{L} Q_h W_h\right) - \left(\sum_{h=a+1}^{L} Q_h W_h \bar{x}_{[ERSS]h(i)}\right)^2 \left(\sum_{h=a+1}^{L} W_h \left(\mu_{X_h} - \bar{x}_{[ERSS]h(i)}\right)\right)$$

This estimator can be rewritten as

$$\bar{y}_{st}\,(EN) = \bar{y}_{st(SERSS)} + \hat{\beta}_{st(SERSS)} \left(\mu_x - \bar{x}_{st(SERSS)}\right) \tag{21}$$

where

$$\hat{\beta}_{st(SERSS)} = \begin{cases} \dfrac{\left(\sum\limits_{h=1}^{a} Q_h W_h\right)\left(\sum\limits_{h=1}^{a} Q_h W_h \bar{y}_{h(ERSS)(i)} \bar{x}_{h[ERSS](i)}\right) - \left(\sum\limits_{h=1}^{a} Q_h W_h \bar{x}_{h[ERSS](i)}\right)\left(\sum\limits_{h=1}^{a} Q_h W_h \bar{y}_{h(ERSS)(i)}\right)}{\left(\sum\limits_{h=1}^{a} Q_h W_h \bar{x}^2_{h[ERSS](i)}\right)\left(\sum\limits_{h=1}^{a} Q_h W_h\right) - \left(\sum\limits_{h=1}^{a} Q_h W_h \bar{x}_{h[ERSS](i)}\right)^2}, & i = e \\ \dfrac{\left(\sum\limits_{h=a+1}^{L} Q_h W_h\right)\left(\sum\limits_{h=a+1}^{L} Q_h W_h \bar{y}_{h(ERSS)(i)} \bar{x}_{h[ERSS](i)}\right) - \left(\sum\limits_{h=a+1}^{L} Q_h W_h \bar{x}_{h[ERSS](i)}\right)\left(\sum\limits_{h=a+1}^{L} Q_h W_h \bar{y}_{h(ERSS)(i)}\right)}{\left(\sum\limits_{h=a+1}^{L} Q_h W_h \bar{x}^2_{h[ERSS](i)}\right)\left(\sum\limits_{h=a+1}^{L} Q_h W_h\right) - \left(\sum\limits_{h=a+1}^{L} Q_h W_h \bar{x}_{h[ERSS](i)}\right)^2}, & i = o. \end{cases} \tag{22}$$

The theoretical variance of regression estimator under ERSS was studied by Muttlak [13]. To simplify the formulas both odd and even cases; we have adapted Muttlak [13]'s notations to SERSS. The regression of Y on X under SERSS can be writen as

$$Y_{h(i:ex)j} = \mu_{Yh} + \frac{\rho_{xyh} S_{yh}}{S_{xh}} \left(X_{[i:ex]j} - \mu_{xh}\right) + \varepsilon_{ijh} \qquad i = 1, 2, ..., r_h;\ j = 1, 2, ..., m$$

Following these notations we can write the variance of $\bar{y}_{st}\,(EN)$ as

$$\text{Var}\left(\bar{y}_{st}\,(EN)\right) = \sum_{h=1}^{L} W_h^2 \frac{S^2_{Yh}}{r_h m} \left(1 - \rho^2_{xyh}\right)\left[1 + E\left(\frac{\bar{Z}^2_{ERSSh}}{S^2_{z3h}}\right)\right] \tag{23}$$

where $Z_{h(i:ex)j} = \frac{X_{h[i:ex]j} - \mu_{xh}}{S_{Xh}}$, $\bar{Z}_{ERSSh} = \frac{1}{r_h m} \sum\limits_{i=1}^{r_h} \sum\limits_{j=1}^{m} Z_{h(i:ex)j}$, $S^2_{z3h} = \frac{1}{r_h m} \sum\limits_{i=1}^{r_h} \sum\limits_{j=1}^{m} \left(Z_{h(i:ex)j} - \bar{Z}_{ERSSh}\right)^2$.
If we take $Q_h = 1$ in equation (19) we have following new calibration estimator

$$\bar{y}_{st(SERSS)}\,(A1) = \bar{y}_{st(SERSS)}\,(A1)_{(e)} + \bar{y}_{st(SERSS)}\,(A1)_{(o)} \tag{24}$$

$$\bar{y}_{st(SERSS)}\,(A1)_{(e)} = \bar{y}_{st(ERSS)h(e)} + \frac{\left(\sum\limits_{h=1}^{a} W_h \bar{y}_{(ERSS)h(e)} \bar{x}_{[ERSS]h(e)}\right) - \bar{x}_{st[ERSS](e)} \bar{y}_{st(ERSS)(e)}}{\left(\sum\limits_{h=1}^{a} W_h \bar{x}^2_{[ERSS]h(e)}\right) - \left(\sum\limits_{h=1}^{a} W_h \bar{x}_{[ERSS]h(e)}\right)^2} \left(\mu_X - \bar{x}_{st[ERSS](e)}\right),$$

$$\bar{y}_{st(SERSS)}\,(A1)_{(o)} = \bar{y}_{st(ERSS)h(o)} + \frac{\left(\sum\limits_{h=a+1}^{L} W_h \bar{y}_{(ERSS)h(o)} \bar{x}_{[ERSS]h(o)}\right) - \bar{x}_{st[ERSS](o)} \bar{y}_{st(ERSS)(o)}}{\left(\sum\limits_{h=a+1}^{L} W_h \bar{x}^2_{[ERSS]h(o)}\right) - \left(\sum\limits_{h=a+1}^{L} W_h \bar{x}_{[ERSS]h(o)}\right)^2} \left(\mu_X - \bar{x}_{st[ERSS](o)}\right)$$

If we take $Q_h = \dfrac{1}{\bar{x}_{[ERSS]h(i)}}$ in equation (19) we have following new calibration estimator

$$\bar{y}_{st(SERSS)}\,(A2) = \bar{y}_{st(SERSS)}\,(A2)_{(e)} + \bar{y}_{st(SERSS)}\,(A2)_{(o)} \tag{25}$$

$$\bar{y}_{st(SERSS)}(A2)_{(e)} = \bar{y}_{st(ERSS)h(e)} + \frac{\left(\sum\limits_{h=1}^{a} \frac{W_h}{\bar{x}_{[ERSS]h(e)}}\right) \bar{y}_{st(ERSS)h(e)} - \left(\sum\limits_{h=1}^{a} \frac{W_h \bar{y}_{(ERSS)h(e)}}{\bar{x}_{[ERSS]h(e)}}\right)}{\bar{x}_{st[ERSS](e)} \left(\sum\limits_{h=1}^{a} \frac{W_h}{\bar{x}_{[ERSS]h(e)}}\right) - 1} \left(\mu_X - \bar{x}_{st[ERSS](e)}\right),$$

$$\bar{y}_{st(SERSS)}(A2)_{(o)} = \bar{y}_{st(ERSS)h(o)} + \frac{\left(\sum\limits_{h=a+1}^{L} \frac{W_h}{\bar{x}_{[ERSS]h(o)}}\right) \bar{y}_{st(ERSS)h(o)} - \left(\sum\limits_{h=a+1}^{L} \frac{W_h \bar{y}_{(ERSS)h(o)}}{\bar{x}_{[ERSS]h(o)}}\right)}{\bar{x}_{st[ERSS](o)} \left(\sum\limits_{h=a+1}^{L} \frac{W_h}{\bar{x}_{[ERSS]h(o)}}\right) - 1} \left(\mu_X - \bar{x}_{st[ERSS](o)}\right).$$

3. Simulation Study

3.1. Using real data set

To see the performance of proposed calibration estimators in SERSS, we have conducted a simulation study in software R. To observe performances of the estimators, we use the real data concerning weight as a study variable and height as auxiliary variable in abalone data [2]. The data consist of $N = 4177$ observations with $\rho_{xy} = 0,82$, $\mu_y = 0,82$, $\mu_x = 0,13$, $C_x = 0,2955$ and $C_y = 0,5916$. We stratified the data set using gender in three strata. ($h = 1 : Female, 2 : Male, 3 : Infant$). The summary statistics of stratas are given in Table 1. In the simulation study, we have selected 10000 samples with different sample sizes $n = 3,5$ under stratified simple random sampling (SSRS) and ERSS designs using R software in Table 2. We have computed mean square errors (MSEs) and percent relative efficiencies (PREs) of estimators displayed in Table 2.

Table 1: The summary statistics of strata

Stratum-I	Stratum-II	Stratum-III
$N_{h1} = 1307$	$N_{h2} = 1342$	$N_{h3} = 1528$
$\mu_{yh1} = 1,046$	$\mu_{yh2} = 0,4313$	$\mu_{yh3} = 0,9914$
$\mu_{xh1} = 0,1580$	$\mu_{xh2} = 0,1079$	$\mu_{xh3} = 0,1513$
$C_{xh1} = 0,2525$	$C_{xh2} = 0,2956$	$C_{xh3} = 0,2300$
$C_{yh1} = 0,1768$	$C_{yh2} = 0,6635$	$C_{yh3} = 0,4745$
$W_{h1} = 0,3129$	$W_{h2} = 0,3212$	$W_{h3} = 0,3658$

Table 2: MSEs and PREs of estimators in Abalone data

n	MSE				
	$\bar{y}_{st}$	$\bar{y}_{st(sinha)1}$	$\bar{y}_{st(sinha)2}$	$\bar{y}_{st(SERSS)}$ $(A1)$	$\bar{y}_{st(SERSS)}$ $(A2)$
3	0.018607	0.010199	0.010254	0.006807	0.006891
5	0.011121	0.004694	0.004736	0.004480	0.004570
				PRE	
3	100	182.44	181.46	**273.36***	269.98
5	100	236.93	234.83	**248.21***	243.35

* indicates efficient estimator under different sample sizes.

3.2. *Using synthetic data set*

Taking motivation from Koyuncu [6] we have generated a synthetic data set of finite populations of size $N = 10000$ from a bivariate normal distribution $[N\left(\mu_x,\ \mu_y,\ \sigma_x^2,\ \sigma_y^2,\ \rho_{xy}\right)$ using mvrnorm function in R software. In the simulation, we considered $\mu_x = 2$, $\mu_y = 4$, $\sigma_x^2 = 2$, $\sigma_y^2 = 1$ and different values of ρ_{xy}. Bivariate Gaussian distributions for each stratum the variance–covariance matrix given, respectively, by

- Stratum 1

$$\Sigma = \begin{bmatrix} 1 & 0.90 \\ 0.90 & 1 \end{bmatrix}$$

- Stratum 2

$$\Sigma = \begin{bmatrix} 1 & 0.76 \\ 0.76 & 1 \end{bmatrix}$$

- Stratum 3

$$\Sigma = \begin{bmatrix} 1 & 0.55 \\ 0.55 & 1 \end{bmatrix}$$

- Stratum 4

$$\Sigma = \begin{bmatrix} 1 & 0.30 \\ 0.30 & 1 \end{bmatrix}$$

We have selected samples from the stratified population. We have computed MSEs and PREs of estimators with respect to $\bar{y}_{st}$, $\bar{y}_{st(sinha)1}$, $\bar{y}_{st(sinha)2}$, $\bar{y}_{st(SERSS)}$ $(A1)$ and $\bar{y}_{st(SERSS)}$ $(A2)$ for $n = 3, 5, 6, 7, 8, 10$ on the basis of 10000 replications and displayed in Table 3.

4. Conclusion

In this study, we have proposed new calibration estimators under SERSS. A simulation study, both real data (abalone) and synthetic data generated from Gaussian distribution were used to support the proposed work. From the simulation study, we saw that calibration estimators under SERSS perform better than SSRS and Shina et al. [19]. In future studies, new calibration estimators can be developed by applying new constraints to different sampling designs.

Acknowledgements

This study is based on Arzu Ece Çetin's Ph.D. thesis and supervised by Nursel Koyuncu.

Table 3: MSEs and PREs of estimators with X ∼ *GaussianDistrubition*

n	MSE				
	$\bar{y}_{st}$	$\bar{y}_{st(sinha)1}$	$\bar{y}_{st(sinha)2}$	$\bar{y}_{st(SERSS)}$ $(A1)$	$\bar{y}_{st(SERSS)}$ $(A2)$
3	0,082235	0,090002	0,090793	0,062997	0,063394
5	0,048043	0,062426	0,062527	0,030177	0,030161
6	0,040548	0,048479	0,048524	0,023840	0,023820
7	0,034205	0,043549	0,043536	0,017872	0,017875
8	0,030173	0,035637	0,035629	0,015867	0,015847
10	0,023958	0,033811	0,033840	0,011265	0,011242
	PRE				
3	100	91,370257	90,574574	**130,53789**	129,720720
5	100	76,959968	76,836061	159,20243	**159,288581***
6	100	83,641609	83,563078	170,08253	**170,225913***
7	100	78,545192	78,568194	**191,38813***	191,356982
8	100	84,667464	84,686237	190,15424	**190,400110***
10	100	70,860049	70,799303	212,67078	**213,103780***

* indicates efficient estimator under different sample sizes.

Funding

The authors took on no certain funding for this study.

References

[1] A.I. Al-Omari, I. Kamarulzaman, M.I. Syam,*Investigating the use of stratified percentile ranked set sampling method for estimating the population mean*, Proyecciones Journal of Mathematics. **30(3)** (2011), 351–368.
[2] C.L. Blake, C.J. Merz, *UCI repository of machine learning databases* , University of California. (1998), http://www.ics.uci.edu/ mlearn.
[3] J.C. Deville, C.E. Särndal, *Calibration estimators in survey sampling* , J. of the American Stat. Assoc. **87** (1972), 376–382.
[4] J.K. Kim, M. Park, *Calibration estimation in survey sampling*, Int. Stat. Review. **78** (2010), 21–29.
[5] J.M. Kim, E.A. Sungur, T.Y. Heo, *Calibration approach estimators in stratified sampling*, Stat. and Prob. Letters. **77** (2007), 99–103.
[6] N. Koyuncu, *New difference-cum-ratio and exponential type estimators in median ranked set sampling*, Hacettepe J. of Mathematics and Stat. **42** (2016), 207–225.
[7] N. Koyuncu, C. Kadilar, *Calibration estimator using different distance measures in stratified random sampling*, Int. J. of Modern Engineering Res. **3** (2013), 415–419.
[8] N. Koyuncu, C. Kadilar, *A new calibration estimator in stratified double sampling*, Hacettepe J. of Mathematics and Stat. **2** (2014), 331–340.
[9] N. Koyuncu, C. Kadilar, *Calibration weighting in stratified random sampling*, Communications in Stat.-Simulation and Com. **7** (2016), 2267–2275.
[10] N. Koyuncu, *Calibration estimator of population mean under stratified ranked set sampling design*, Communications in Stat.-Theory and Meth. **23** (2018), 5845–5853.
[11] C. Li, C. Ma, *The single-step preconditioned HSS method for singular linear systems*, Applied and Computational Mathematics. **20(2)** (2021), 247–256.
[12] V.L. Mandowara, N. Mehta, *Modified ratio estimators using stratified ranked set sampling*, Hacettepe Journal of Mathematics and Statistics. **43(3)** (2014), 461–471.
[13] H. A. Muttlak, *Regression estimators in extreme and median ranked set samples*, Journal of Applied Statistics, **28(8)** (2001), 1003-1017.
[14] A.M. Mouhamed, A.A. El-Sheikh, H.A. Mohamed, *A new calibration estimator of stratified random sample*, App. Mathematical Sci. **35** (2015), 1735–1744.
[15] H.M. Samawi, M.S. Ahmed, W. Abu-Dayyeh, *Estimating the population mean using extreme ranked set sampling*, Biometrical J. **5** (1972), 577–586.
[16] H.M. Samawi, L.J. Saeid, *Stratified extreme ranked set sample with application to ratio estimators*, J. of Modern App. Statistical Meth. **1** (1972), 12.
[17] U. Shahzad, F. Alshahrani, I.M. Almanjahie, S. Iftikhar, *Calibration-based mean estimators under stratified median ranked set sampling*, Mathematics, **8** (2023), 1825.
[18] U. Shahzad, I. Ahmad, I. Mufrah Almanjahie, M. Hanif, N.H. Al-Noor, *L-Moments and calibration-based variance estimators under double stratified random sampling scheme: Application of Covid-19 pandemic*, Scientia Iranica, **2** (2023), 814–821.
[19] N. Sinha, B.V.S. Sisodia, S. Singh, S.K. Singh, *Calibration approach estimation of the mean in stratified sampling and stratified double sampling*, Communications in Statistics-Theory and Methods, **46(10)** (2017), 4932–4942.

[20] G.N. Singh, D. Bhattacharyya, A. Bandyopadhyay, *A general class of calibration estimators under stratified random sampling in presence of various kinds of non-sampling errors*, Communications in Statistics-Simulation and Computation, **2** (2023), 320–333.
[21] C.A. Taconeli, W.H. Bonat, *On the performance of estimation methods under ranked set sampling*, Computational Statistics, **35**(2020), 1805–1826.
[22] D.C. Tracy, S. Singh, R. Arnab, *Note on calibration in stratified and double sampling*, Survey Methodology,**1** (2003), 99–104.
[23] T. Zaman, H. Bulut, *Robust calibration for estimating the population mean using stratified random sampling*, Scientia Iranica, (2023).