Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 48 (5) (2019), 1560 – 1569
DOI : 10.15672/hujms.466178

Research Article

Two-layer median ranked set sampling

Begum Kara Gulay*, Neslihan Demirel

Dokuz Eylul University, Faculty of Science, Department of Statistics, Izmir-Turkey

Abstract

In this article, we propose two-layer median ranked set sampling (TMRSS) design that combines median ranked set sampling (MRSS) and two-layer ranked set sampling (TRSS). Ranked set sampling (RSS) is an alternative sampling method that can improve the efficiency of estimators when exact measurement of response variable is either difficult, time consuming or expensive. Evaluation of the TMRSS performance for different distributions, set, and cycle sizes regarding mean and regression coefficients estimators and mean square of the regression model are carried out using Monte Carlo simulation study and real data application. The results indicate that estimators of TMRSS yields are either equivalent to or better than MRSS.

Mathematics Subject Classification (2010). 62D05

Keywords. ranked set sampling, median ranked set sampling, two-layer ranked set sampling, concomitant variables, mean estimator, regression estimators, relative efficiency

1. Introduction

A sample is a subset of a population that allows the researchers to make statistical inferences about a population, without investigating every unit. However, representation of the entire population by the chosen units in a sample to be used for the analysis is crucial. Typically, a variety of sampling techniques are present in the literature. The most common sampling method is known as simple random sampling (SRS). Each member of the population has an equal chance of being selected as subject in SRS. Recently, ranked set sampling (RSS) has become popular as an alternative sampling method, as it improves the efficiency of estimators by providing more representative sample from the population, in which the measurement of response variable is either difficult, expensive or time consuming. Also, RSS does not require a full measurement, uses ranks of the units based on a visual inspection or concomitant variables. McIntyre [12], proposed RSS for the first time for estimating mean pasture yields, whereas Takahashi and Wakimoto [17] developed its mathematical theory. Further, they demonstrated that regardless of ranking error, the mean estimator of RSS is unbiased and it is more efficient than the mean estimator of SRS for the same sample size. The ranking error occurs when the visual inspection or concomitant variables cannot be assigned the correct rank number to the observations in

*Corresponding Author.
Email addresses: begum.kara@deu.edu.tr, neslihan.ortabas@deu.edu.tr
Received: 01.10.2018; Accepted: 02.04.2019

the set. Dell and Clutter [7] as well as David and Levine [6] reported similar results under imperfect ranking, which means there is a ranking error. For some recent bibliography on the RSS, see [8, 11, 18, 20].

In RSS, let k be a positive integer that denotes the set size, and randomly select k^2 units by SRS from the target population. The k^2 selected units are divided arbitrarily into k sets, each having size k. Further, the ranking of these units is carried out according to professional judgment (visual inspection) or a concomitant variable correlated with the variable of interest. However, in comparison to the variable of interest, the measurement of this concomitant variable is quite cheap or easy. For definite quantification, the selection of a sample is carried out by including the smallest ranked unit from the first set, the second smallest ranked unit from the second set, with the procedure continuing until the selection of k^{th} smallest ranked unit from the k^{th} set. These steps complete one cycle of a ranked set sample of size k. Moreover, to obtain a ranked set sample of size $n = km$, the above procedure can be repeated m times.

Let $Y_{[1]1}, ..., Y_{[k]1}; Y_{[1]2}, ..., Y_{[k]2}; ...; Y_{[1]m}, ..., Y_{[k]m}$ be a ranked set sample, where $Y_{[i]j}$ represents the i^{th} ranked unit in the j^{th} cycle and $\bar{Y}_{RSS}$ denotes the mean estimator of RSS in the Equation (1.1). Under consistent ranking process, it is an unbiased estimator for the population mean (μ). A ranking mechanism is said to be consistent if $F = \frac{1}{k}\sum_{i=1}^{k} F_{[i]}$ holds, where F denotes the distribution function of Y and $F_{[i]}$ denotes the distribution function of the i^{th} ordered statistics of a random sample of size k from F (see details in [4], pages 12-14).

$$\bar{Y}_{RSS} = \frac{1}{mk}\sum_{j=1}^{m}\sum_{i=1}^{k} Y_{[i]j} \tag{1.1}$$

Recently, there have been many contributions to RSS research. Muttlak [13] proposed median ranked set sampling (MRSS). The MRSS method is easy to apply since only the middle of the sample is considered. Muttlak [14] mentioned that the MRSS can be performed with less ranking error in performing the ranking of the units with respect to the variable of interest. Hajighorbi and Saba [9] also pointed out that the MRSS method can be effortlessly employed in the field while saving time in performing the ranking of the units concerning the variable of interest. The MRSS is useful for situations in which identifying a sample unit with intermediate rank is easier less prone to error than a sample unit with lower/upper rank and the researcher is interested about the center of the distribution. If the researcher is interested about upper/lower tail of the distribution, then usage of ranked set sampling with extreme ranks should be in priority. See Zamanzade and Mahdizadeh [21] for more details.

The MRSS method can be summarized as follows: Randomly select k^2 units by simple random sample from the population, the k^2 selected units are allocated as randomly into k sets, each of size k, and the units within each set are ranked with respect to a variable of interest. If the set size k is odd, from each set the $((k+1)/2)^{th}$ smallest ranked unit, which is the median of the set should be selected for the measurement. On the other hand, if the set size k is even, one should select the $(k/2)^{th}$ smallest ranked units from the first $k/2$ sets, and the $((k+2)/2)^{th}$ smallest ranked units from the second $k/2$ sets. Further, to obtain km units, the cycle may be repeated m times. The estimators of the population mean using MRSS for odd and even set sizes are given in Equation (1.2) and (1.3), respectively.

$$\bar{Y}_{MRSS} = \frac{1}{mk}\sum_{j=1}^{m}\sum_{i=1}^{k} Y_{[\frac{k+1}{2}(i)]j} \tag{1.2}$$

where $Y_{[\frac{k+1}{2}(i)]j}$ denotes the $((k+1)/2)^{th}$ ranked unit in the i^{th} set for the j^{th} cycle.

$$\bar{Y}_{MRSS} = \frac{1}{mk}\sum_{j=1}^{m}\left(\sum_{i=1}^{k/2} Y_{[\frac{k}{2}(i)]j} + \sum_{i=(k+2)/2}^{k} Y_{[\frac{k+2}{2}(i)]j}\right) \tag{1.3}$$

where $Y_{[\frac{k}{2}(i)]j}$ denotes the $(k/2)^{th}$ ranked unit and $Y_{[\frac{k+2}{2}(i)]j}$ denotes the $((k+2)/2)^{th}$ ranked unit in the i^{th} set for the j^{th} cycle.

Al-Saleh and Al-Omari [2] proposed the multistage RSS to increase the efficiency of estimator of population mean, thereby estimating the average olive yields in a field in West Jordan. Chen and Shen [5] proposed two-layer RSS, which has two concomitant variables. In the first and second layer of procedure, sampling units are ranked corresponding to one and other concomitant variables, respectively. Typically, all the features of the general RSS can be applied to the two-layer RSS, as it falls into the scheme of the general RSS. Let $X^{\{1\}}$ and $X^{\{2\}}$ denote the concomitant variables. The outline of the steps of two-layer RSS is as follow: Identify kl^2 sets from the target population, each of size k, where the units in each sets are ranked using $X^{\{1\}}$. For l^2 ranked sets, the units with $X^{\{1\}}$ - rank 1 are selected, followed by selection, the units with $X^{\{1\}}$ - rank 2, for another l^2 ranked sets. Further with all other sets, the same process is repeated, which in turn completes the first layer of the procedure. In the second layer, the units selected in first layer are separated randomly or systematically into kl subsets, each of size l. Further, the ranking of the units in each of these subsets are carried out according to $X^{\{2\}}$. For the first ranked subset, the unit with $X^{\{2\}}$ - rank 1 is picked up, and its value will be measured on Y. Similarly, for the second ranked subset, the unit with rank 2 is selected, its value will be measured on Y. The process is further repeated with all other subsets, thereby making these steps part of a cycle. Repeating the cycle m times yields, the data set

$$\{\ Y_{[r][s]j} : r = 1, ..., k;\ s = 1, ..., l;\ j = 1, ..., m\ \}$$

where, $Y_{[r][s]j}$ is the measurement of Y in the j^{th} cycle on the unit with $X^{\{1\}}$ - rank r and $X^{\{2\}}$ - rank s, sample of size $n = klm$ [5].

The estimator of population mean of TRSS is given in Equation (1.4).

$$\bar{Y}_{TRSS} = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{l}\sum_{s=1}^{l}\frac{1}{k}\sum_{r=1}^{k} Y_{[r][s]j} \tag{1.4}$$

Al-Omari and Bouza [1] demonstrated the superiority of the two-layer RSS over the RSS.

In this paper, two-layer median ranked set sampling (TMRSS) design and its mean estimator have been proposed. Further, the performance of estimators of MRSS and TMRSS designs are being compared. Section 2 introduces the methodology of TMRSS design. In Section 3, the biases and relative efficiencies of the estimators are obtained using Monte Carlo simulation studies. Section 4 represents a real data application and in Section 5 concluding remarks are provided.

2. Two-layer median ranked set sampling

Two-layer median ranked set sampling (TMRSS) was first introduced in Kara[10]. TMRSS design and its mean estimator have been given in this section. The aim of this study is to improve the MRSS by using more than one concomitant variable. The motivation of TMRSS is to get more homogenous sample by two-layer ranking and to reduce ranking error by using the median. TMRSS is described as follows:

Let $X^{\{1\}}$ and $X^{\{2\}}$ denote the concomitant variables.

Step 1: First, kl^2 independent sets, each of size k, are drawn from the target population.

Step 2: Ranking of the units in each of these sets are carried out according to $X^{\{1\}}$.

- When the set size k is odd, select from each set the $((k+1)/2)^{th}$ smallest rank, the median of the set.
- When the set size k is even, for each k ranked sets, select the $(k/2)^{th}$ smallest rank from the first $k/2$ sets, and the $((k+2)/2)^{th}$ smallest rank from the second $k/2$ sets, with repeating the process for all the sets.

Thus, the first layer of the procedure completes.

Step 3: In the second layer, the units selected from the first layer are divided randomly or systematically, into kl subsets, each of size l.

Step 4: The units in each of these subsets are ranked using $X^{\{2\}}$.

- When the set size l is odd, select from each sample the $((l+1)/2)^{th}$ smallest rank, which is the median of the set.
- When the set size l is even, for each l ranked sets, select the $(l/2)^{th}$ smallest rank from the first $l/2$ sets, and the $((l+2)/2)^{th}$ smallest rank from the second $l/2$ sets, followed by further repetition of the process for all subsets.

Further, carrying out the measurement of the values of the selected units. Thus, it completes one cycle of the procedure.

Step 5: The cycle may be repeated m times to get klm units, yielding the data set

$$\{\ Y_{[s]c} : s = 1, ..., kl;\ c = 1, ..., m\ \}$$

where $Y_{[s]c}$ is the corresponding value of $(X^{\{1\}}, X^{\{2\}})$ and denotes the measurement of Y on s^{th} unit in the c^{th} cycle.

Equation (2.1) gives the estimator of population mean using TMRSS method.

$$\bar{Y}_{TMRSS} = \frac{1}{m}\sum_{c=1}^{m}\frac{1}{kl}\sum_{s=1}^{kl} Y_{[s]c} \tag{2.1}$$

An example is considered, with the aim of illustrating the construction of a two-layer median ranked set sample.

1. Assume that $k = 4$, $l = 2$ and, $m = 1$. We have a random sample of size 64 units into $kl^2 = 16$ sets, each of size $k = 4$. Let $X_{i(j)}^{\{p\}}$ be the j^{th} smallest order statistic $(j = 1, 2, 3, 4)$ of the i^{th} set ($i = 1, 2, \cdots, 16$), which is ranked according to p^{th}

$(p = 1, 2)$ concomitant variable.

2. For $p = 1$, the ranking of the units in each of these sets are carried out according to $X^{\{1\}}$. Upon being ranked, the sets appear as shown below:

$$(X^{\{1\}}_{i(1)}, X^{\{1\}}_{i(2)}, X^{\{1\}}_{i(3)}, X^{\{1\}}_{i(4)}),\ (i = 1, 2, \cdots, 16)$$

   – Subsequently, for every four ranked sets, select the second smallest rank from the first 2 sets, and the third smallest rank from the second 2 sets, with repeating the process further. The selected units from the first layer are given below:

$$(X^{\{1\}}_{1(2)}, X^{\{1\}}_{2(2)}, X^{\{1\}}_{3(3)}, X^{\{1\}}_{4(3)}, X^{\{1\}}_{5(2)}, X^{\{1\}}_{6(2)}, X^{\{1\}}_{7(3)}, X^{\{1\}}_{8(3)}, \\ X^{\{1\}}_{9(2)}, X^{\{1\}}_{10(2)}, X^{\{1\}}_{11(3)}, X^{\{1\}}_{12(3)}, X^{\{1\}}_{13(2)}, X^{\{1\}}_{14(2)}, X^{\{1\}}_{15(3)}, X^{\{1\}}_{16(3)})$$

   Thus, it completes the first layer of the procedure.

3. In the second layer, the units selected from the first layer are divided, randomly or systematically, into $kl = 8$ subsets, each of size $l = 2$.

4. For $p = 2$, the units in each of these subsets are ranked using $X^{\{2\}}$.

$$[(X^{\{1\}}, X^{\{2\}}_{i(1)}), (X^{\{1\}}, X^{\{2\}}_{i(2)})],\ (i = 1, 2, \cdots, 8)$$

   – Next, for each two ranked subsets, select the first smallest rank from the first subset and the second smallest rank from the second subset, with repeating the process further. The selected units from the second layer are specified below:

$$[(X^{\{1\}}, X^{\{2\}}_{1(1)}), (X^{\{1\}}, X^{\{2\}}_{2(2)}), (X^{\{1\}}, X^{\{2\}}_{3(1)}), (X^{\{1\}}, X^{\{2\}}_{4(2)}), \\ (X^{\{1\}}, X^{\{2\}}_{5(1)}), (X^{\{1\}}, X^{\{2\}}_{6(2)}), (X^{\{1\}}, X^{\{2\}}_{7(1)}), (X^{\{1\}}, X^{\{2\}}_{8(2)})]$$

   Therefore, the second layer of the procedure gets completed.

5. The selected units are measured on Y. Hence, one cycle of the procedure is completed. The final set, for one cycle, is

$$\left\{Y_{[s]1} : s = 1, \cdots, 8\right\} .$$

3. Simulation study

In this section, the TMRSS and MRSS estimators, which are the relative efficiencies of mean and regression coefficients estimators, the relative efficiencies of mean squares of the regression models, and the amount of bias of regression coefficient estimators, are being compared. With the intention of comparing the methods, simulation with R statistical software version 3.5.1 was carried out. The steps of simulation study are specified below:

1. Generated 10,000 observations coming from some symmetric and asymmetric distributions: Normal (0,1), Uniform (0,1), Exponential (1), Gamma (5,1), and Lognormal (0,1).

2. Y_i is calculated using the following regression model

$$Y_i = \beta_0 + \beta_1 X_i^{\{1\}} + \beta_2 X_i^{\{2\}} + \varepsilon_i \ , \ i = 1, \cdots, N$$

   where $\beta_0 = \beta_1 = \beta_2 = 1$ and ε_i is distributed normal with mean zero and the variance $\sigma_\varepsilon^2 = 0.25$, 0.5, and 1.0.

3. $N = 10,000$ is the repetition number.

4. The performances of the estimators are investigated for $k = 6, 8$ for MRSS and $k = 3, 4$ for TMRSS where $l = 2$ and $m = 3, 5, 10$.

5. The estimators of TMRSS are compared with the estimators of MRSS using $X^{\{1\}}$ as the ranking variable.

6. Let θ denote the unkonown parameter, and $\hat{\theta}_i$ denote the corresponding estimator of i^{th} number of repetitions. The bias values of regression coefficient estimators are computed using Equation (3.1).

$$Bias(\hat{\theta}_i, \theta) = \frac{1}{N} \sum_{i=1}^{N} (\hat{\theta}_i - \theta) \tag{3.1}$$

7. The mean square error (MSE) of the estimators are then computed by Equation (3.2).

$$MSE(\hat{\theta}_i, \theta) = \frac{1}{N} \sum_{i=1}^{N} (\hat{\theta}_i - \theta)^2 \tag{3.2}$$

8. For the comparison of efficiencies of the estimators, relative efficiency (RE) values are calculated as below:

$$RE_1 = \frac{MSE_{\bar{Y}_{MRSS}}}{MSE_{\bar{Y}_{TMRSS}}}, \qquad RE_2 = \frac{MSE_{\hat{\beta}_{MRSS}}}{MSE_{\hat{\beta}_{TMRSS}}}, \qquad RE_3 = \frac{MSE_{\hat{\sigma}^2_{MRSS}}}{MSE_{\hat{\sigma}^2_{TMRSS}}} \tag{3.3}$$

Tables 1-4 depict the results of the simulation studies. Table 1 shows the relative efficiencies of the population mean estimator. From the table, it is evident that every value of RE_1 is higher than 1, except for some cases of uniform distribution, thereby indicating that the population mean estimator of TMRSS design is more efficient than that of MRSS design. Moreover, the population mean estimator of TMRSS is more effective in asymmetric distributions. Elaborately, in asymmetric distributions, when other simulation parameters are fixed, increasing set size k or cycle size m, increases the efficiency of TMRSS over MRSS.

Table 2 shows the biases of the regression coefficient estimator $\hat{\beta}_2$. The table indicates that the biases of the regression estimator of TMRSS are comparable magnitude with the biases of the regression estimator of MRSS. Further, the biases values are barely large at uniform distributions in both TMRSS and MRSS designs. The biases are generally smaller in asymmetric distributions than symmetric distributions. The bias of $\hat{\beta}_0$ and $\hat{\beta}_1$ are not

Table 1. Relative efficiencies of the population mean estimator (RE_1)

m	k	σ_ε^2	Normal (0,1)	Uniform (0,1)	Exponential (1)	Gamma (5,1)	Lognormal (0,1)
3	3	1.00	1.026	1.060	1.159	1.105	1.213
		0.50	1.056	1.006	1.200	1.147	1.220
		0.25	1.101	0.992	1.258	1.134	1.208
	4	1.00	1.066	1.025	1.256	1.162	1.253
		0.50	1.103	1.003	1.245	1.173	1.280
		0.25	1.056	1.010	1.305	1.178	1.280
5	3	1.00	1.077	0.970	1.317	1.210	1.338
		0.50	1.076	0.998	1.408	1.224	1.329
		0.25	1.080	0.999	1.429	1.207	1.360
	4	1.00	1.086	1.005	1.332	1.181	1.298
		0.50	1.123	0.966	1.404	1.222	1.345
		0.25	1.118	1.041	1.448	1.254	1.378
10	3	1.00	1.041	0.999	1.465	1.266	1.416
		0.50	1.103	1.048	1.534	1.326	1.456
		0.25	1.137	1.062	1.563	1.276	1.493
	4	1.00	1.109	1.029	1.612	1.364	1.563
		0.50	1.069	1.000	1.698	1.387	1.551
		0.25	1.142	1.052	1.711	1.403	1.603

Table 2. Biases of regression estimator $\hat{\beta}_2$ of MRSS and TMRSS

m	k	σ_ε^2	Normal (0,1)		Uniform (0,1)		Exponential (1)		Gamma (5,1)		Lognormal (0,1)	
			MRSS	TMRSS	MRSS	TMRSS	MRSS	TMRSS	MRSS	TMRSS	MRSS	TMRSS
3	3	1.00	0.007	0.006	-0.034	-0.008	0.015	0.011	0.001	0.000	-0.011	-0.009
		0.50	0.004	0.005	0.002	0.009	-0.007	-0.004	0.001	0.001	0.001	0.002
		0.25	-0.004	-0.004	0.005	-0.002	0.007	0.003	0.002	0.001	0.003	0.001
	4	1.00	0.002	0.003	-0.019	-0.009	0.006	0.006	0.001	0.000	-0.008	-0.007
		0.50	0.001	0.002	0.000	0.001	-0.003	-0.003	0.000	0.001	0.001	0.002
		0.25	-0.006	-0.005	0.006	-0.002	0.004	0.004	0.001	0.001	0.002	0.001
5	3	1.00	0.002	0.005	-0.017	-0.013	0.005	0.004	0.001	0.000	-0.004	-0.002
		0.50	0.003	0.002	-0.007	0.007	-0.004	-0.003	-0.001	0.000	0.001	0.002
		0.25	-0.005	-0.004	0.007	0.007	0.004	0.004	0.000	0.001	0.002	0.002
	4	1.00	0.003	0.004	-0.029	-0.016	0.006	0.012	0.000	-0.002	-0.010	-0.006
		0.50	0.005	0.002	-0.013	0.001	-0.004	-0.002	0.000	0.000	0.000	0.003
		0.25	-0.006	-0.005	0.006	0.006	0.006	0.004	0.002	0.001	0.001	0.001
10	3	1.00	-0.002	0.004	-0.023	-0.011	0.001	0.010	0.000	0.000	-0.008	-0.004
		0.50	0.002	0.002	-0.004	0.007	-0.001	-0.001	-0.001	0.002	0.001	0.000
		0.25	-0.005	-0.005	0.002	0.002	0.008	0.004	0.002	0.000	0.003	0.002
	4	1.00	0.001	0.001	-0.025	-0.011	0.000	0.006	0.001	0.001	-0.003	-0.002
		0.50	0.003	0.003	0.003	-0.004	-0.002	-0.002	0.000	0.000	0.002	0.002
		0.25	-0.006	-0.004	0.003	0.002	0.005	0.004	0.002	0.001	0.003	0.002

shared, as they have a similar pattern.

RE_2 values are obtained for $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$. For better clarity of the presentation, the relative efficiencies of $\hat{\beta}_1$ are tabulated in Table 3. In the present study, the regression coefficient estimator of the proposed design, for all the simulation parameters, always performs better than the regression coefficient estimator of MRSS. Especially in asymmetric distibutions, the results are more favourable and the efficiency mostly increase, when k or m increase for fixed σ_ε^2.

Table 4 displays the relative efficiencies of the mean squares of the regression models. The values are close to 1.0; hence, we can conclude that MRSS and TMRSS designs have equivalent MSE.

Table 3. Relative efficiencies of the regression estimator $\hat{\beta}_1$ (RE_2)

m	k	σ_ε^2	Normal (0,1)	Uniform (0,1)	Exponential (1)	Gamma (5,1)	Lognormal (0,1)
3	3	1.00	1.553	1.416	1.610	1.561	1.663
		0.50	1.553	1.399	1.616	1.545	1.701
		0.25	1.559	1.399	1.628	1.549	1.734
	4	1.00	1.581	1.415	2.656	1.602	1.788
		0.50	1.551	1.396	1.637	1.562	1.783
		0.25	1.556	1.402	1.634	1.587	1.760
5	3	1.00	1.560	1.411	1.668	1.596	1.905
		0.50	1.570	1.393	1.691	1.583	1.875
		0.25	1.556	1.403	1.676	1.585	1.851
	4	1.00	2.158	1.825	2.370	2.214	2.565
		0.50	2.176	1.823	2.396	2.150	2.584
		0.25	2.157	1.830	2.323	2.183	2.565
10	3	1.00	2.171	1.824	2.377	2.201	2.701
		0.50	2.161	1.797	2.376	2.170	2.719
		0.25	2.132	1.821	2.367	2.194	2.681
	4	1.00	2.156	1.827	2.417	2.181	2.895
		0.50	2.149	1.805	2.446	2.168	2.850
		0.25	2.141	1.816	2.430	2.195	2.806

Table 4. Relative efficiencies of the mean square erros of regression models (RE_3)

m	k	σ_ε^2	Normal (0,1)	Uniform (0,1)	Exponential (1)	Gamma (5,1)	Lognormal (0,1)
3	3	1.00	1.105	1.011	0.995	0.995	0.994
		0.50	0.992	1.001	1.015	0.996	0.993
		0.25	1.007	1.003	0.999	0.992	0.999
	4	1.00	1.006	1.002	1.002	1.007	0.999
		0.50	0.996	0.994	1.004	1.005	0.994
		0.25	0.997	1.000	0.989	1.004	0.997
5	3	1.00	1.002	1.007	1.002	1.004	1.002
		0.50	0.998	0.993	1.012	1.005	0.998
		0.25	0.998	1.000	0.998	1.000	0.996
	4	1.00	0.998	0.998	1.013	1.003	0.999
		0.50	1.002	0.987	1.020	1.004	1.001
		0.25	0.997	1.000	1.001	1.003	0.997
10	3	1.00	1.009	1.005	1.006	1.005	0.995
		0.50	0.995	0.993	1.014	1.002	1.000
		0.25	0.998	1.000	1.002	1.000	0.995
	4	1.00	1.006	1.006	1.006	0.998	1.002
		0.50	0.993	0.988	1.016	1.001	0.995
		0.25	0.995	1.003	0.998	0.998	0.996

4. Real data application

In this section, for comparing the methods, a real data analysis is employed. The data set is the abalone data provided by Nash et al. [15] used for prediction of the age of the abalone. Cetintav et al. [3] and Sevinc et al. [16] use this data set, for application of RSS. The abalone dataset comes with the goal of attempting to predict abalone age (through the number of rings on the shell) given various descriptive attributes of the abalone. For the determination of abalone age, the shell is sliced through the cone to form rings, which is further stained and counted through a microscope, a boring and time-consuming task. Other measurements, which are easier to obtain, are used to predict the age. The data set contains nine variables; (*i*) Sex, (*ii*) Length (mm), (*iii*) Diameter (mm), (*iv*) Height (mm), (*v*) Whole weight (grams), (*vi*) Shucked weight (grams), (*vii*) Viscera weight (grams), (*viii*) Shell weight (grams), and (*ix*) Rings (integer +1.5 gives the age in years).

Assuming that this study aims to obtain the mean and regression estimation of the viscera weight, Y, which is more time-consuming measurement process as compared to other physical measurements. Diameter and Height are selected as the concomitant variables, $X^{\{1\}}$ and $X^{\{2\}}$, respectively, for ranking the abalones in the random sets without actual measurement of their Y values. The correlation coefficients between $X^{\{1\}}$, Y,and $X^{\{2\}}$, Y are 0.902 and 0.877, respectively. For this analysis, $k = 5$, $l = 2$ and $m = 5$ are considered. The results are tabulated in Table 5. From the table, it is evident that all values of relative efficiencies are greater than 1.0, thereby indicating that the estimators of TMRSS are more efficient than the estimators of MRSS.

Table 5. Relative efficiencies for real data application

RE_1	$RE_2(\hat{\beta}_0)$	$RE_2(\hat{\beta}_1)$	$RE_2(\hat{\beta}_2)$	RE_3
1.071	1.288	1.995	1.573	1.768

5. Discussion

This study proposes the two-layer median RSS design which is an extension of MRSS and its mean estimator. According to the results of simulation studies and real data application, the efficiency of the mean estimator of TMRSS is always more as compared to the mean estimator of MRSS except some cases of uniform distribution. Further, the relative efficiency values are higher in asymmetric distributions than symmetric ones. On the other hand, the regression coefficient estimators of TMRSS are comparable in magnitude with the biases of the regression coefficient estimators of MRSS. Nevertheless, it appears that these two estimators have little large biases in uniform distribution. In comparison to, the regression coefficient estimator of MRSS, for all the simulation parameters, the performance of the regression coefficient estimator of TMRSS is always better. Moreover, the efficiencies of mean and regression estimators of TMRSS increases upon increasing the set and/or cycle size, especially in asymmetric disributions, while keeping the other simulation parameters fixed. Additionally, the mean square errors of the regression models of the MRSS and TMRSS are found to be almost equal. As a future study, TMRSS may be improved as multi-layer for multiple concomitant variables. Additionally, the different estimators such as proportion and variance can be developed for TMRSS. As another direction for future research, the other modified RSS methods like neoteric ranked set sampling (NRSS) [19], which is consistently superior to the usual RSS for estimating population mean and variance can be extended to a two-layer design.

Acknowledgment. The authors are sincerely grateful to the anonymous referees for their time and effort in providing valuable comments and suggestions that have led to a substantial improvement in the quality of the paper.

References

[1] A. I. Al-Omari and C. N. Bouza, *Review of ranked set sampling: modifications and applications*, Investigación Oper. **35** (3), 215-235, 2014.

[2] M. F. Al-Saleh and A. I. Al-Omari, *Multistage ranked set sampling*, J. Statist. Plann. Inference **102** (2), 273-286, 2002.

[3] B. Cetintav, G. Ulutagay, S. Gurler and N.Demirel, *Mean estimation based on fwa using ranked set sampling with single and multiple rankers*, in: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 790-797, Springer, Cham, 2016.

[4] Z. Chen, Z. Bai and B. Sinha, *Ranked set sampling: theory and applications*, Springer Science & Business Media, 2003.
[5] Z. Chen and L. Shen, *Two-layer ranked set sampling with concomitant variables*, J. Statist. Plann. Inference **115** (1), 45-57, 2003.
[6] H. A. David and D. N. Levine, *Ranked set sampling in the presence of judgment error*, Biometrics **28**, 553-555, 1972.
[7] T. R. Dell and J. L. Clutter, *Ranked set sampling theory with order statistics background*, Biometrics, 545-555, 1972.
[8] J. Frey, *Constrained nonparametric estimation of the mean and the CDF using ranked-set sampling with a covariate*, Ann. Inst. Statist. Math. **64** (2), 439-456, 2012.
[9] S. Hajighorbani and R. A. Saba, *Stratified median ranked set sampling: Optimum and proportional allocations*, Journal of Statistical Research of Iran **9**, 87–102, 2012.
[10] B. Kara, Comparison of Efficiency of Two-Layer Median Ranked Set Sampling Using Concomitant Variables. Master's thesis, Dokuz Eylul University, Turkey, 2015.
[11] L. Khan, J. Shabbir and S. Gupta, *Unbiased ratio estimators of the mean in stratified ranked set sampling.* Hacet. J. Math. Stat. **46** (6), 1151–1158, 2016.
[12] G. McIntyre, *A method for unbiased selective sampling, using ranked sets*, Australian Journal of Agricultural Research **3** (4), 385–390, 1952.
[13] H. A. Muttlak, *Median ranked set sampling*, J. Appl. Statist. Sci. **6**, 245–255, 1997.
[14] H. A. Muttlak, *Median ranked set sampling with concomitant variables and a comparison with ranked set sampling and regression estimators*, Environmetrics: The official journal of the International Environmetrics Society **9** (3), 255–267, 1998.
[15] W. J. Nash, T. L. Sellers, S. R. Talbot, A. J. Cawthorn and W. B.Ford, *The population biology of abalone (haliotis species) in tasmania. i. blacklip abalone (h. rubra) from the north coast and islands of bass strait*, Sea Fisheries Division, Technical Report (48), 1994.
[16] B. Sevinc, S. Gurler and B. Cetintav, *Partial groups ranked set sampling and mean estimation*, J. Stat. Comput. Simul. **88** (14), 2799-2810, 2018.
[17] K. Takahasi and K. Wakimoto, *On unbiased estimates of the population mean based on the sample stratified by means of ordering*, Ann. Inst. Statist. Math. **20** (1), 1–31, 1968.
[18] X. Wang, S. Ahn and J. Lim, *Unbalanced ranked set sampling in cluster randomized studies*, J. Statist. Plann. Inference **187**, 1–16, 2017.
[19] E. Zamanzade and A. I. Al-Omari, *New ranked set sampling for estimating the population mean and variance*, Hacet. J. Math. Stat. **45** (6), 1891-1905, 2016.
[20] E. Zamanzade and M. Mahdizadeh, *A more efficient proportion estimator in ranked set sampling*, Statist. Probab. Lett. **129**, 28–33, 2017.
[21] E. Zamanzade and M. Mahdizadeh, *Using ranked set sampling with extreme ranks in estimating the population proportion*, Stat. Methods Med. Res., 13 pages, 2019, https://doi.org/10.1177/0962280218823793.