COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2020.1815777

Taylor & Francis
Taylor & Francis Group

Check for updates

Ratio estimators for ranked set sampling in the presence of tie information

Eda Gizem Koçyiğit[a] and Cem Kadilar[b]

[a]Department of Statistics, Dokuz Eylul University, Izmir, Turkey; [b]Department of Statistics, Hacettepe University, Ankara, Turkey

ABSTRACT

In this article, we study the situation of observations whose ranks cannot be determined for the auxiliary variable in the Ranked Set Sampling (RSS) method. Therefore, we examine the case of tie information for ratio estimators of the population mean. We propose a new exponential ratio estimator using the modified isotonic estimator for this situation. Simulation results show that the proposed estimator is more efficient than the other estimators in literature. In addition, when we examine the recent COVID-19 pandemic situation, we see that the data is suitable for this structure. We can also see from the real data that the proposed estimator gives better results.

ARTICLE HISTORY
Received 23 March 2020
Accepted 23 August 2020

KEYWORDS
Isotonic estimator; Ranked set sampling; Ratio estimator; Tie information

SUBJECT CLASSIFICATION CODES
62D05; 94A20; 62F07

1. Introduction

Ranked set sampling (RSS), which was proposed by McIntyre (1952), is an alternative sampling technique to the simple random sampling (SRS). McIntyre's work was motivated by practices in agriculture (Kvam 2003; Halls and Dell 1966), biometrics (Samawi and Al-Sagheer 2001; Mahdizadeh and Zamanzade 2019) and medicine (Chen, Stasny, and Wolfe 2005). Patil (1995) proposed a variety of applications of RSS in environmental monitoring. Then, many studies have focused on this area (Nussbaum and Sinha 1997; Kvam 2003; Ozturk, Bilgin, and Wolfe 2005). It is shown that RSS has some advantages against SRS.

The advantages of the method are feasibility for infinite populations and fewer data required in the sample for the estimation of the population parameters. In applications, the researcher is asked to rank all units in a set of m from the smallest to the largest without measurement, such as visual sorting. It may not be possible or reliable that the researcher is completely confident to order of all observations. Using that kind of ranking method having the researcher's personal judgements causes significant increase of the ranking error. However, a partial ranking can be provided for the units in the sets. Ozturk (2011) discussed the difficulties caused by random breaking of ties and suggested a number of methods in which the researchers use the tie information. Frey (2012) demonstrated that the tie information could also be used completely, instead of partial ranking. Recently, this structure has been used in the field of biostatistics. Zamanzade and Wang (2020b) obtained the area under the ROC curve using a partially ordered set.

In the RSS, different rank values are given for each unit in all sets by assuming that there is no tie information. When there is tie information, the sampling scheme of the RSS-t method (tie method in RSS) records the tie information and randomly selects the corresponding unit in the

CONTACT Eda Gizem Koçyiğit eda.kocyigit@deu.edu.tr Department of Statistics, Dokuz Eylul University, Izmir, Turkey.
Supplemental data for this article is available online at https://doi.org/10.1080/03610918.2020.1815777.
© 2020 Taylor & Francis Group, LLC

i-th set of the j-th cycle. Let the selected unit be $Y_{[i]:j}$. The observations having ties are recorded in the **T** matrix, whose size is mxn, as follows:

$$T^j = \begin{bmatrix} I^j_{[1,1]} & \cdots & I^j_{[1,m]} \\ . & . & . \\ . & . & . \\ . & . & . \\ I^j_{[m,1]} & \cdots & I^j_{[m,m]} \end{bmatrix}$$

where $I^j_{[i:k]}$ is an indicator variable that is one if the unit having rank i is tied with the unit having rank k in the j-th cycle and zero otherwise ($i = 1, \ldots, m$; $k = 1, \ldots, m$). Remark that $\sum_{K=1}^{m} I^j_{[i:k]} \geq 1$ and if rank i is assigned to only one unit in the set, "$\geq$" is replaced with "$=$". Also note that all data for RSS-t technique cover not only the exact measurement values $\{Y_{[i]:j}, i = 1, \ldots, m, j = 1, \ldots, m\}$ but also matrices containing tie information matrices $\mathbf{T}^1, \ldots, \mathbf{T}^n$.

2. Estimators in literature

When there is tie in the RSS, the standard RSS mean estimator simply ignores the matrices of tie structures given by

$$\hat{\mu}_1 = \frac{1}{m}\sum_{i=1}^{m} \bar{Y}_{[i]}, \tag{1}$$

where $\bar{Y}_{[i]} = \frac{1}{n}\sum_{j=1}^{n} Y_{[i]j}$ shows the sample mean of the i-th ordered statistics. It is simply the average of sample units collected by the RSS.

One way to include the ties in the estimators is to divide each connected unit between the layers corresponding to the order it belongs to. If a measured unit r is connected to different rows, then it should be assigned as $1/r$ in each r layer, instead of 1. This strategy is a special version of a more general method proposed by MacEachern, Stasny, and Wolfe (2004) and the estimator can be written as follows:

$$\hat{\mu}_2 = \frac{1}{m}\sum_{i=1}^{m} \bar{Y}'_{[i]}, \tag{2}$$

where $\bar{Y}'_{[i]} = \dfrac{\sum_{\ell=1}^{m}\sum_{j=1}^{n_\ell} Y_{[\ell]j}\frac{I_{[\ell]ji}}{\sum_{k=1}^{m} I_{[\ell]jk}}}{\sum_{\ell=1}^{m}\sum_{j=1}^{n_\ell}\frac{I_{[\ell]ji}}{\sum_{k=1}^{m} I_{[\ell]jk}}}$ is defined as the weighted sample average for the layers. Here, $I^j_{[i:k]}$ is the indicator variable that is in the **T** matrix.

The means of the ordered sampling statistics, $\bar{Y}_{[1]}, \ldots, \bar{Y}_{[m]}$ or $\bar{Y}'_{[1]}, \ldots, \bar{Y}'_{[m]}$, may not be ascending order as it should be, owing to the correlation between the auxiliary variable and the measured variable that is not equal to one and thus the auxiliary variables are ranked as judgmentally. In this case, $\bar{Y}_{[1]}, \ldots, \bar{Y}_{[m]}$ or $\bar{Y}'_{[1]}, \ldots, \bar{Y}'_{[m]}$ estimates can be isotonized using the sample sizes $n_1, \ldots, n_m$ or the following weighted sample sizes by

$$n'_i \equiv \sum_{\ell=1}^{m}\sum_{j=1}^{n_\ell} \frac{I_{[\ell]ji}}{\sum_{k=1}^{m} I_{[\ell]jk}}, i = 1, 2, \ldots, m.$$

Wang, Lim, and Stokes (2008) defined the isonotic estimators by

$$\bar{Y}_{[i],iso} = \max_{1\leq r\leq i}\min_{i\leq s\leq m} \frac{\sum_{\ell=r}^{s} n_\ell \bar{Y}_{[\ell]}}{\sum_{\ell=r}^{s} n_\ell}, i = \ 1, \ldots, m, \tag{3}$$

$$\bar{Y}'_{[i],iso} = \max_{1\le r\le i}\min_{i\le s\le m}\frac{\sum_{\ell=r}^{s} n'_\ell \bar{Y}'_{[\ell]}}{\sum_{\ell=r}^{s} n_\ell}, i = \ 1,\ldots,m \tag{4}$$

that are isotonic versions of the estimates $\bar{Y}_{[1]},\ldots,\bar{Y}_{[m]}$ and $\bar{Y}'_{[1]},\ldots,\bar{Y}'_{[m]}$, respectively. These isotonic means in-stratum can also be obtained by using the PAVA algorithm which was explained by Robertson (1988). The exact mean estimates will be the mean of these isotonic averages which are given in (3) and (4). Due to this isotonic strategy, the following non-parametric mean estimators were, respectively, proposed by Ozturk (2007) for the unbalanced RSS data and by Wang, Lim, and Stokes (2008) for the judgment post stratified (JPS) data

$$\hat{\mu}_3 = \frac{1}{m}\sum_{i=1}^{m}\bar{Y}_{[i],iso}, \tag{5}$$

$$\hat{\mu}_4 = \frac{1}{m}\sum_{i=1}^{m}\bar{Y}'_{[i],iso}, \tag{6}$$

where $\hat{\mu}_3$ and $\hat{\mu}_4$ are the isotonic versions of $\hat{\mu}_1$ and $\hat{\mu}_2$, respectively. These methods have been considered by researchers recently (Ozturk 2007, 2014; Wang, Lim, and Stokes 2008; Frey and Feeman 2012, 2013; Wang, Wang, and Lim 2012; Dümbgen and Zamanzade 2013; Zamanzade 2016; Dastbaravarde, Arghami, and Sarmad 2016) and explained step by step in detail by Zamanzade and Wang (2017).

The ratio estimator, proposed by Prasad (1989) for the simple random sampling, is defined by

$$\hat{\mu}_5 = \frac{\bar{y}}{\bar{x}}\bar{X}, \tag{7}$$

where $\bar{x}$ and $\bar{y}$ are the sample means of the auxiliary and study variables, respectively and $\bar{X}$ shows the population mean of the auxiliary variable. The estimator in (7) is adapted to the ranked set sampling by Samawi and Muttlak (1996) as follows:

$$\hat{\mu}_6 = \frac{\bar{y}_{[n]}}{\bar{x}_{[n]}}\bar{X}, \tag{8}$$

where $\bar{x}_{[n]} = \frac{1}{mn_i}\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{[n]ij}$ and $\bar{y}_{[n]} = \frac{1}{mn_i}\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{[n]ij}$ are the sample means of the auxiliary and study variables, respectively, obtained by the ranked set sampling.

3. Proposed estimators

We have seen that isotonic estimators provide good results from studies on this subject, such as Frey (2012) and Zamanzade and Wang (2018). In addition, we know that ratio estimators give better results than the sample mean estimator. Then, we define the first proposed ratio estimator for the case of tie information in the ranked set sampling, motivated by Kadilar, Unyazici, and Cingi (2009), as follows:

$$\hat{\mu}_7 = \frac{\frac{1}{m}\sum_{i=1}^{m}\bar{Y}'_{[i],iso}}{\frac{1}{m}\sum_{i=1}^{m}\bar{X}'_{[i],iso}}\overline{X_t} = \frac{\hat{\mu}_4}{\hat{\mu}_{4[xt]}}\overline{X_t}, \tag{9}$$

where $\hat{\mu}_4$ is the isotonic mean estimator proposed by Wang, Lim, and Stokes (2008), $\hat{\mu}_{4[xt]}$ is also the isotonic mean estimator of the auxiliary variable X_t and $\overline{X_t}$ is the population mean of X_t that includes tie information. The proposed estimator in (9) simply combines the ratio estimator with the isotonic estimators.

We define the second ratio estimator for the presence of tie information in the ranked set sampling, inspired by the exponential ratio estimator proposed by Bahl and Tuteja (1991) for the SRS and adapted by Singh, Tailor, and Jatwa (2011) for the systematic sampling, as follows:

Table 1. COVID-19 data set parameters.

$N = 34$	β_{Xt}=16.77	$\bar{X}_t = 0.9118$
$m = n = 5$	ρ_{XtY}=0.842 (*p-value* < 0.05)	$\bar{Y} = 93$

Table 2. Ranking information in a sample.

Set	Ranked Units	Auxiliary Value (X_T)	Precise Value (Y)
1	**Guangxi &** Sichuan **&** Hunan, Hubei	0	2
2	Inner Mongolia **&** Anhui **& Guangdong &** Ningxia	0	8
3	Hebei **&** Yunnan, **Heilongjiang,** Hong Kong SAR	1	13
4	Hainan & Anhui & Liaoning, **Shanghai**	2	3

$$\hat{\mu}_8 = \hat{\mu}_4 \exp\left(\frac{\overline{X_t} - \hat{\mu}_{4[xt]}}{\overline{X_t} + \hat{\mu}_{4[xt]}}\right). \tag{10}$$

We illustrate the sampling scheme and the calculation of the estimators with the aid of confirmed and suspected cases of COVID-19 acute respiratory disease reported by provinces, regions, and cities in China. Suppose that we are trying to estimate the total death of the disease in a given population and we know that there is a high correlation between the number of in last 24 hours confirmed cases and the total death. Information about the population is taken from "Data as of 11 March 2020 from Coronavirus disease 2019 (COVID-19) Situation Report − 51" of World Health Organization (2020) and that population is summarized in Table 1.

To draw a ranked set sample from $n = 4$ using the set size $m = 4$, firstly we should draw m^2=16 subjects from population size $N = 34$. Then, we divide them into 4 sets randomly. In sets, we try to rank each subject using the confirmed cases in the last 24 hours as an auxiliary variable. In Table 2, an example of the RSS-t scheme is studied in detail. We connect the units with ampersands that are declared tied and the unit that is selected for precise value, given with boldface.

Table 2 shows that there are ties in several sets. From Table 2, we see that just two of the cities selected for precise measurement were involved in ties. In set 1, Guangxi was tied for ranks 1 to 3 and in set 2, Guangdong was tied for ranks 1 to 4. We select one of the units between ties randomly for precise value also record this tie information for using in the estimation process. The matrix that contains the tie information for Table 2 as follows:

$$T^1 = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Since the number of repetitions is one, $\bar{Y}_1 = 2$, $\bar{Y}_2 = 8$, $\bar{Y}_3 = 13$, and $\bar{Y}_4 = 3$. Thus, $\hat{\mu}_1 = 6.5$. If we split the tied units among the strata for the ranks tied in the following **T** matrix

$$T^*_{n_{ij}} = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

we obtain the weighted sample sizes from the column sum of the matrix given above $(n'_1, ..., n'_4) = (0.58\bar{3}, 0.58\bar{3}, 1.58\bar{3}, 1.25)$. Computing the weighted sample means as $\bar{Y}_1{}' = 4.571$, $\bar{Y}_2{}' = 4.571$, $\bar{Y}_3{}' = 9.894$, and $\bar{Y}_4{}' = 4$; thus, $\hat{\mu}_2 = 5.75$. Since $\bar{Y}_1 \leq \bar{Y}_2 \leq \bar{Y}_3 \leq \bar{Y}_4$ and $\bar{Y}_1{}' \leq \bar{Y}_2{}' \leq \bar{Y}_3{}' \leq \bar{Y}_4{}'$, conditions are not satisfied for this sample and we calculate isotonic sample means as $\bar{Y}_{1,iso} = 2$, $\bar{Y}_{2,iso} = 8$, $\bar{Y}_{3,iso} = 8$, $\bar{Y}_{4,iso} = 8$, and $\bar{Y}'_{1,iso} = 4.471$, $\bar{Y}'_{2,iso} = 4.571$, $\bar{Y}'_{3,iso} = 7.294$, $\bar{Y}'_{4,iso} = 7.294$. From these isotonic means, we calculate isotonized estimator $\hat{\mu}_3 = 6.5$ and

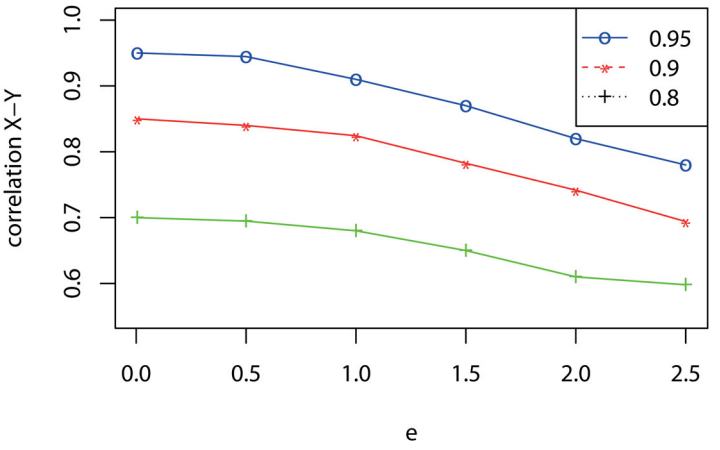


Figure 1. Changes of the correlation coefficient according to *e* values.

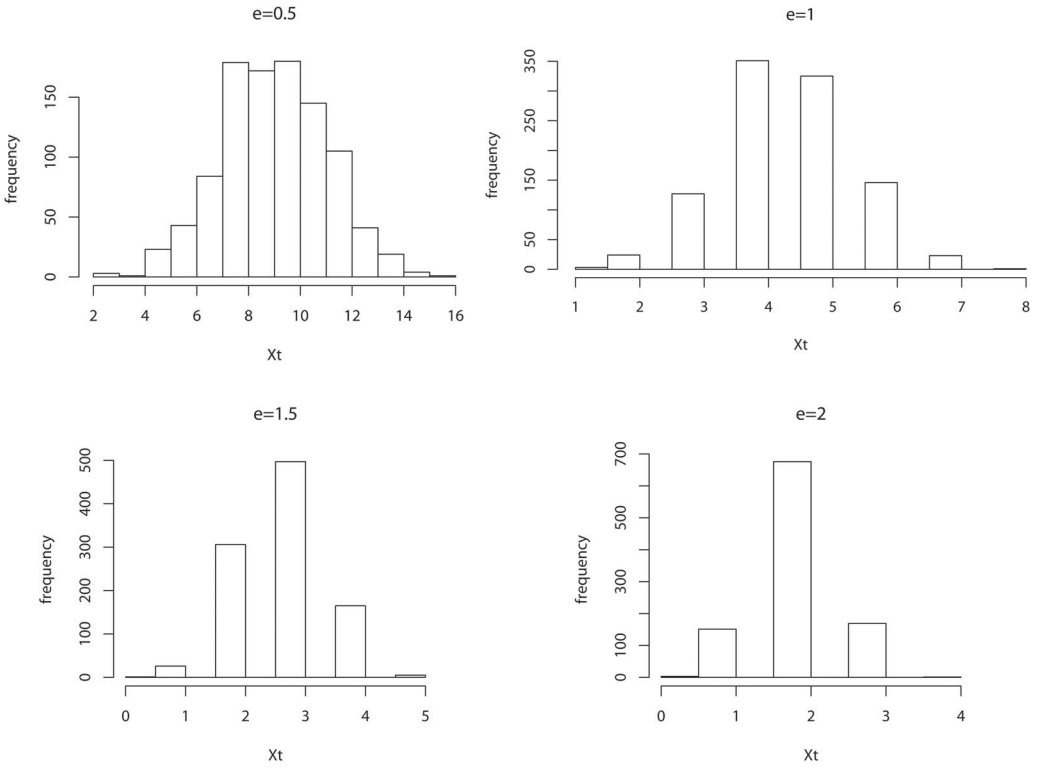


Figure 2. Histograms of X for different e values when N (1, 5) and $\rho = 0.7$.

$\hat{\mu}_4$ =5.932. To calculate the SRS estimator that is $\hat{\mu}_5$, it is enough to draw a size of 4 random subjects from the population. Consider the random sample as Ningxia, Macao SAR, Hong Kong SAR, and Anhui. Then, estimators can be calculated with arithmetic mean as $\hat{\mu}_5$=1.64. For calculating $\hat{\mu}_6$, **T** matrix is ignored and ties are broken randomly. Then, sample means for the auxiliary variable and the study variable as $\bar{x}_{RSS}$ =0.65 and $\bar{y}_{RSS} = \hat{\mu}_1$ =6.5, respectively; thus, $\hat{\mu}_6$=7.9. Isotonic means for the auxiliary variable X_t should be calculated for finding $\hat{\mu}_7$ and $\hat{\mu}_8$. It can be calculated with the same process of $\hat{\mu}_4$. Since $\bar{X}_1' \leq \bar{X}_2' \leq \bar{X}_3' \leq \bar{X}_4'$ for this case from Table 2, the isotonized estimators should be equal to $\hat{\mu}_{2[xt]}$. This estimator's calculation process is the

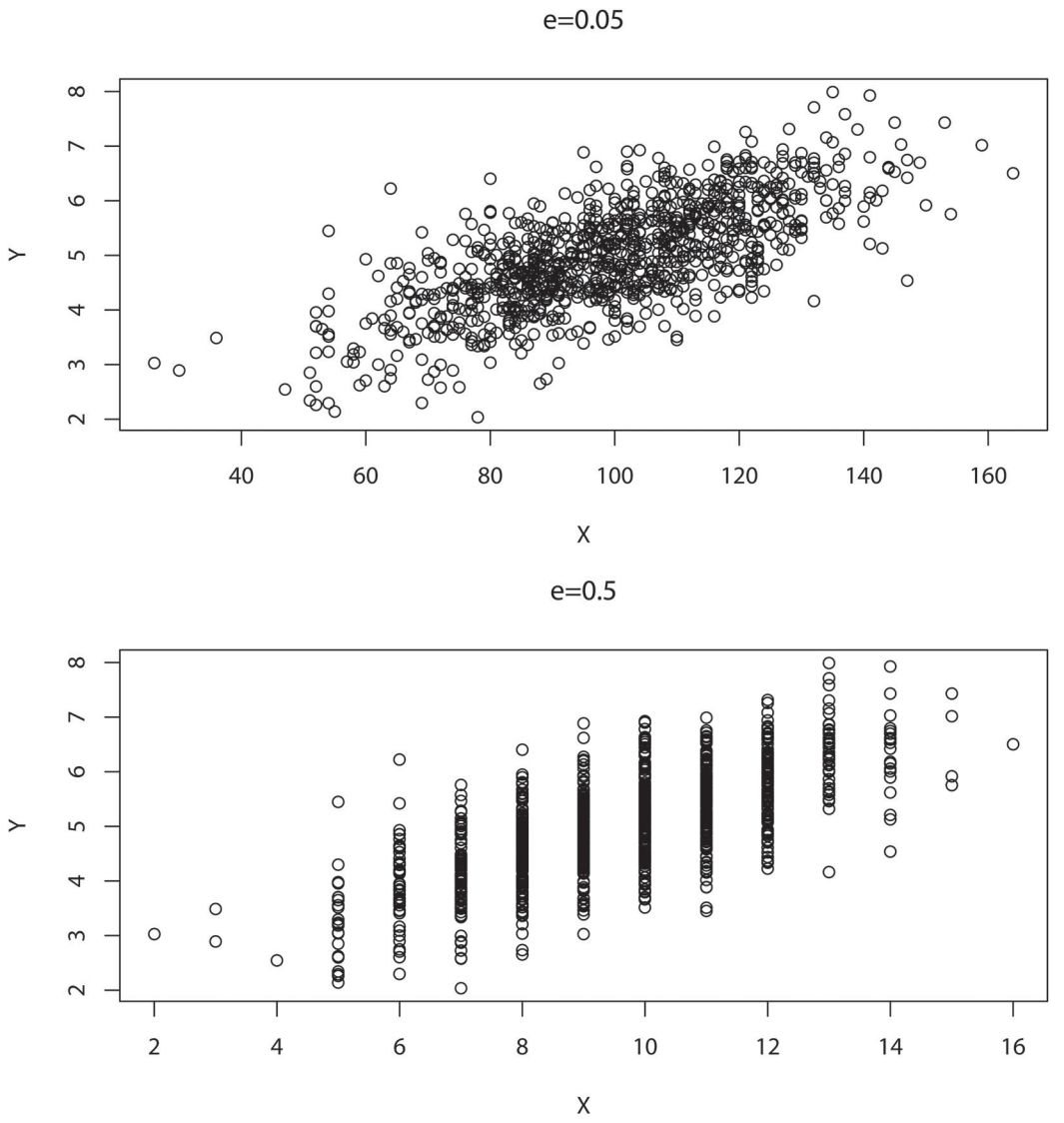


Figure 3. X-Y scatter plots for different e values when N (1, 5) and $\rho = 0.7$.

same as $\hat{\mu}_2$ with one difference that X sample data is used, instead of Y sample data. In this way, $\hat{\mu}_{4[xt]}$=$\hat{\mu}_{2[xt]}$= 0.558. Then, the proposed estimator can be calculated as $\hat{\mu}_7$ =9.693 and $\hat{\mu}_8$ =7.54, using $\hat{\mu}_4$ and $\bar{X}_t = 0.9118$, given in Table 1.

4. Simulation study

There are two methods in the case of a tie in the ranking in literature. The first method is the Discrete Perceived Size (DPS) that was developed by Fligner and MacEachern (2006) and the second one is the Tied-if-close (TIC) method that was proposed by Dell and Clutter (1972).

In the TIC method, if $|X_i - X_j| < e$, the condition is provided, which means that i-th and j-th units are connected. Here, e is a parameter that must be selected greater than zero. In the case of $| X_i - X_j | \geq e$, the units are considered to be sufficiently separated from each other and cannot be considered as tied.

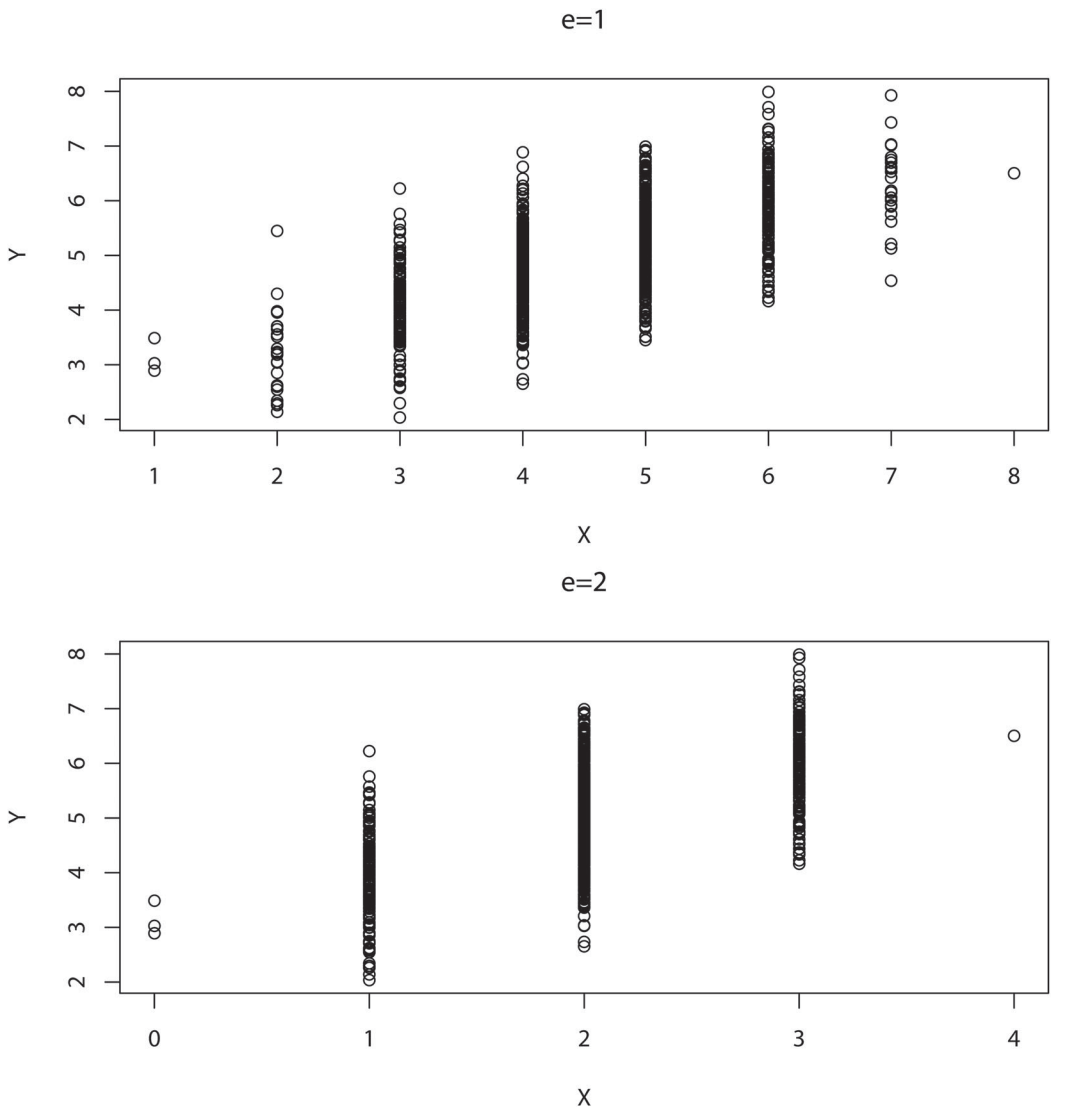


Figure 3. (Continued).

In the DPS method, the variable X is made discrete by rolling X/e down and $X_t = \lfloor X/e \rfloor$ is regained, where e is a parameter that must be selected to be greater than zero, as in the TIC method. In the case of $X_{ti} = X_{tj}$, i-th and j-th units are called tied.

In ratio estimators, since we need the population mean of the auxiliary variable, the DPS method has been used in the simulation, as ties cannot be produced on the population with the TIC method.

In DPS and TIC methods, distributions may show undesirable behavior when the e and m parameters are changed. This situation usually occurs when the e value increases. Increasing of e value is expected to lead to more ties, but the situation may not always produce the desired result. This was also demonstrated by Frey's theorem in 2012. In addition, as the e value increases, the X auxiliary variable becomes distant from its real values and the correlation between X and Y decreases. In Figure 1, the variation of the initial correlation coefficient determined by e value is shown.

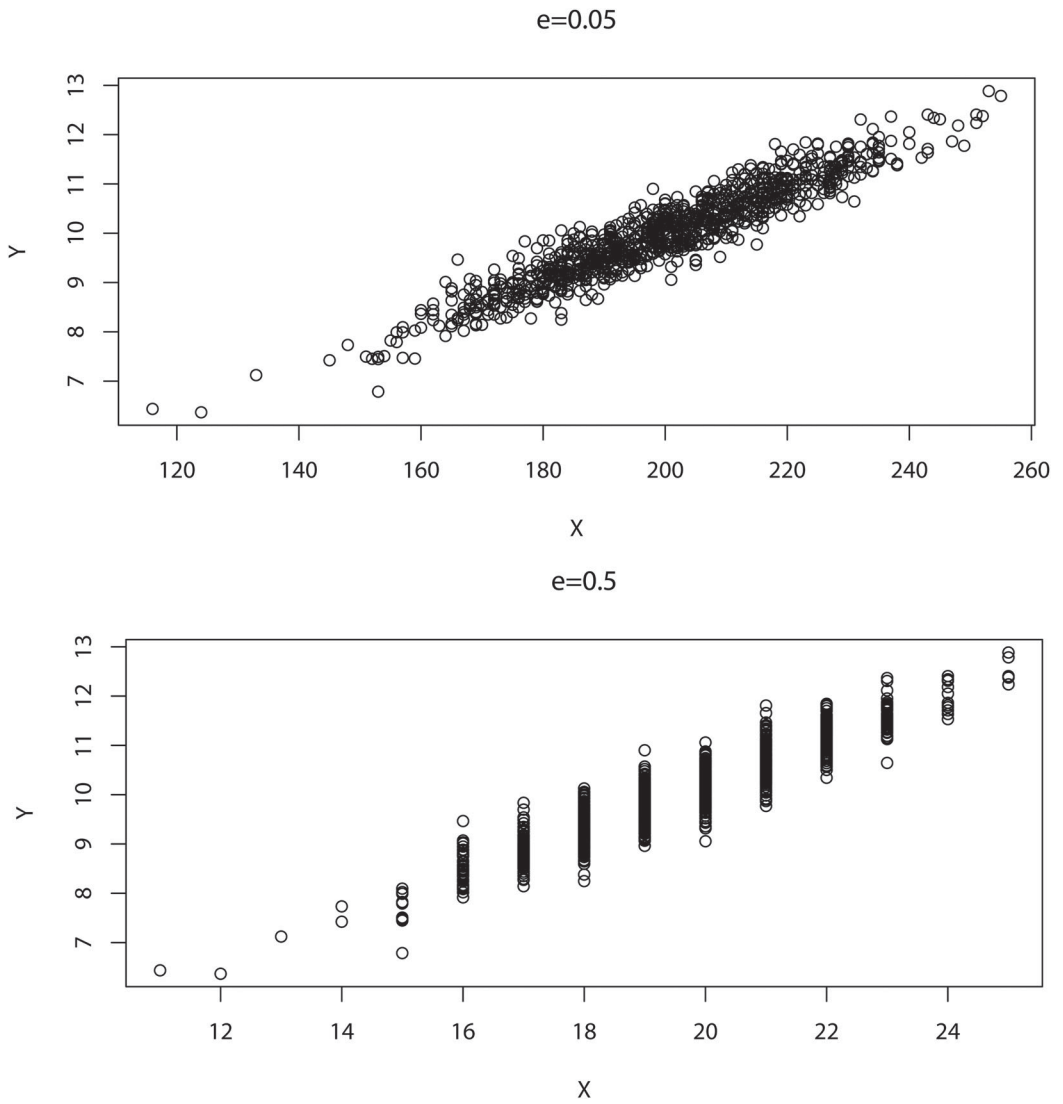


Figure 4. X-Y scatter plots for different e values when N (1, 5) and $\rho = 0.9$.

Histogram graphs of the auxiliary variable under the normal distribution for different e values are shown in Figure 2 and X-Y scatter plots under the normal distribution for different e values are shown in Figures 3 and 4.

In order to use the ratio estimators, a positive high correlation is expected between the interest and the auxiliary variables. For this reason, the e parameter required for the DPS model has been determined specifically for each distribution so as not to reduce the correlation coefficient between X and Y.

In the simulation coded by the R program, the population size of 1000 is derived from the bivariate normal distribution. Parameters are selected as $\rho = 0.7$, 0.8, and 0.9; $m = 3$, 4, and 5. The model was developed under the DPS method and the mean estimators were obtained for the number of trials of 100000.

The relative efficiency values, where the reference estimator is $\hat{\mu}_1$, are calculated by proportioning the mean square errors as follows:

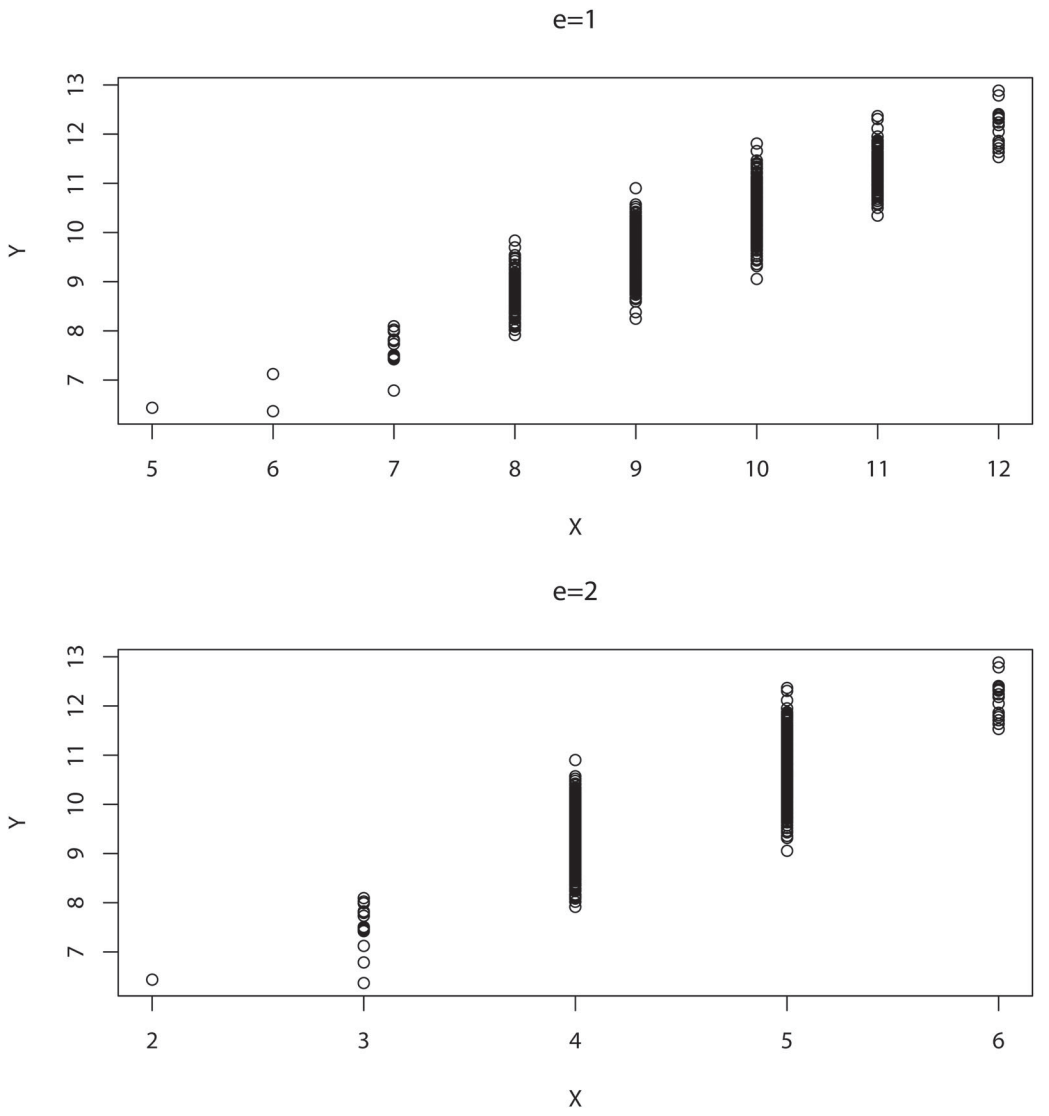


Figure 4. (Continued).

$$RE_i = \frac{MSE(\hat{\mu}_1)}{MSE(\hat{\mu}_i)}, \quad i = 2, 4, 5, 6, 7, 8 \tag{11}$$

Since $\hat{\mu}_1$ does not contain the tie information, all weight values are equal in the isotonic formula. Therefore, it is always equal to $\hat{\mu}_3$ that is an isotonic version of $\hat{\mu}_1$ and thus, $\mathrm{RE}_3 = 1$. For this reason, it is not included in Table 1. The brief information about the relative efficiencies is given in Table 3.

In the ranked set sampling, the auxiliary variable contains the tie information when the population is symmetrically distributed with zero-mean. Therefore, the values of the population means of $\overline{X_t}$ or $\hat{\mu}_{4[xt]}$ in the proposed estimator may be zero. For this reason, the parameters for the normal distribution are randomly selected as (1, 5) and the RE results are given in Table 2. It should be noted that the correlation values, given in Table 4, decrease as the e value increases, as given in Figure 1.

Table 3. Information about the relative efficiency values.

	Sampling Type	*Tie Info.*	*X Info.*	*Isotonic*
RE_2	RSS	+	-	-
RE_4	RSS	+	-	+
RE_5	RSS	-	+	-
RE_6	SRS	-	+	-
RE_7	RSS	+	+	+
RE_8	RSS	+	+	+

Table 4. Relative efficiencies for normal distribution with parameters μ=5, σ^2=1.

	ρ_{XY}	0.70			0.75			0.85			0.95		
DPS	*m*	3	4	5	3	4	5	3	4	5	3	4	5
***e* = 0.5**	RE_2	0.9895	0.9795	0.9752	0.9949	0.9836	0.9785	1.0037	1.001	0.9978	1.0434	1.0466	1.0487
	RE_4	1.0207	1.0209	1.0208	1.0242	1.0221	1.0227	1.0341	1.0390	1.0429	1.0676	1.0795	1.0842
	RE_5	1.2063	1.1354	1.0974	1.2845	1.2096	1.1498	1.9018	1.7431	1.6082	4.0425	3.5336	3.146
	RE_6	1.3542	1.3038	1.2605	1.4422	1.3772	1.3227	**2.0580**	**1.8920**	**1.7650**	**4.4**	**3.8446**	**3.4771**
	RE_7	1.3306	1.2758	1.2301	1.418	1.348	1.2925	2.0203	1.8452	1.7154	4.3165	3.7348	3.3659
	RE_8	**1.4371**	**1.3547**	**1.2951**	**1.5539**	**1.4512**	**1.3815**	1.8681	1.7283	1.6288	2.8541	2.6492	2.489
***e* = 1**	RE_2	0.9987	1	1.0061	1.0106	1.0155	1.0207	1.0328	1.044	1.0551	1.1017	1.1288	1.1511
	RE_4	1.0421	1.0473	1.0533	1.0532	1.0644	1.0671	1.0764	1.0921	1.1031	1.14	1.1715	1.1913
	RE_5	0.9687	0.9503	0.9155	1.1857	1.1314	1.0202	1.5351	1.4398	1.3746	2.5841	2.3153	2.1356
	RE_6	1.1644	1.1480	1.1425	1.3794	1.3589	1.2543	1.7871	1.6914	1.6239	**3.0415**	**2.7636**	**2.6011**
	RE_7	1.16	1.1430	1.1389	1.3462	1.3247	1.2407	1.7551	1.66	1.5917	3.0011	2.7144	2.5555
	RE_8	**1.3991**	**1.3252**	**1.2865**	**1.5445**	**1.4663**	**1.39**	**1.8749**	**1.7416**	**1.6565**	2.7675	2.5491	2.3981
***e* = 1.5**	RE_2	1.014	1.0253	1.0338	1.0337	1.0502	1.036	1.0534	1.0710	1.0996	1.1291	1.1689	1.2027
	RE_4	1.0601	1.0698	1.0737	1.0792	1.0953	1.0803	1.0975	1.1156	1.1408	1.17	1.208	1.2426
	RE_5	0.7471	0.7350	0.7331	0.9631	0.9334	0.746	1.0649	1.0377	1.0186	1.425	1.3452	1.2759
	RE_6	0.9597	0.9809	0.9899	1.1684	1.1803	1.0132	1.3478	1.3258	1.3252	1.8521	1.8058	1.7667
	RE_7	0.9842	1.0165	1.0353	1.1786	1.1946	1.0586	1.3820	1.3637	1.3683	1.9253	1.883	1.853
	RE_8	**1.3611**	**1.3089**	**1.27**	**1.5276**	**1.4603**	**1.3325**	**1.7691**	**1.6601**	**1.6069**	**2.4944**	**2.3028**	**2.1939**
***e* = 2**	RE_2	1.0443	1.0453	1.0490	1.0614	1.0685	1.0702	1.0876	1.0933	1.0991	1.1906	1.2001	1.206
	RE_4	1.0843	1.0838	1.0853	1.1019	1.1108	1.1104	1.1239	1.1322	1.1383	1.2208	1.2343	1.2423
	RE_5	0.5908	0.5934	0.5979	0.6606	0.6841	0.6832	0.7869	0.7909	0.7827	1.0739	1.0564	1.0191
	RE_6	0.7624	0.8011	0.8162	0.885	0.9221	0.9388	1.0312	1.0544	1.0713	1.4009	1.4133	1.3797
	RE_7	0.871	0.9001	0.9125	0.9867	1.0201	1.028	1.1318	1.1382	1.1521	1.5698	1.5629	1.5105
	RE_8	**1.2903**	**1.2561**	**1.2382**	**1.4537**	**1.4153**	**1.3786**	**1.6554**	**1.5715**	**1.5355**	**2.2523**	**2.1339**	**2.0314**
***e* = 2.5**	RE_2	0.9823	1.0237	1.0495	0.9915	1.033	1.064	1.0063	1.0713	1.1219	1.0367	1.1212	1.182
	RE_4	1.037	1.0631	1.0763	1.0468	1.0741	1.093	1.0661	1.1148	1.1511	1.0983	1.168	1.2141
	RE_5	0.3429	0.3494	0.3559	0.3637	0.377	0.3834	0.4716	0.4778	0.4742	0.4362	0.4416	0.431
	RE_6	0.5033	0.5482	0.5839	0.5382	0.5927	0.6345	0.6816	0.7348	0.7572	0.6502	0.7098	0.747
	RE_7	0.4963	0.6053	0.712	0.5419	0.6674	0.7771	0.6707	0.8078	0.9121	0.6479	0.799	0.9255
	RE_8	**1.1224**	**1.1341**	**1.1479**	**1.1952**	**1.2021**	**1.2072**	**1.4968**	**1.4587**	**1.4305**	**1.7179**	**1.6822**	**1.6354**

According to the simulation results, given in Figure 5, while the correlation coefficient increases and while the set sizes of *m* and *e* values decrease, the relative efficiency values increase.

5. Real COVID-19 data results

In Sec. 3, we calculated the estimator based on one sample from COVID-19 data. In this section, we consider the same variables as auxiliary variable and study variable from the same population, mentioned in Sec. 3. We are trying to estimate the total death of the disease in a given population and we know that there is a high correlation between the number of confirmed cases in last 24 hours and the total death. The population, is summarized in Table 1 in Sec. 3.

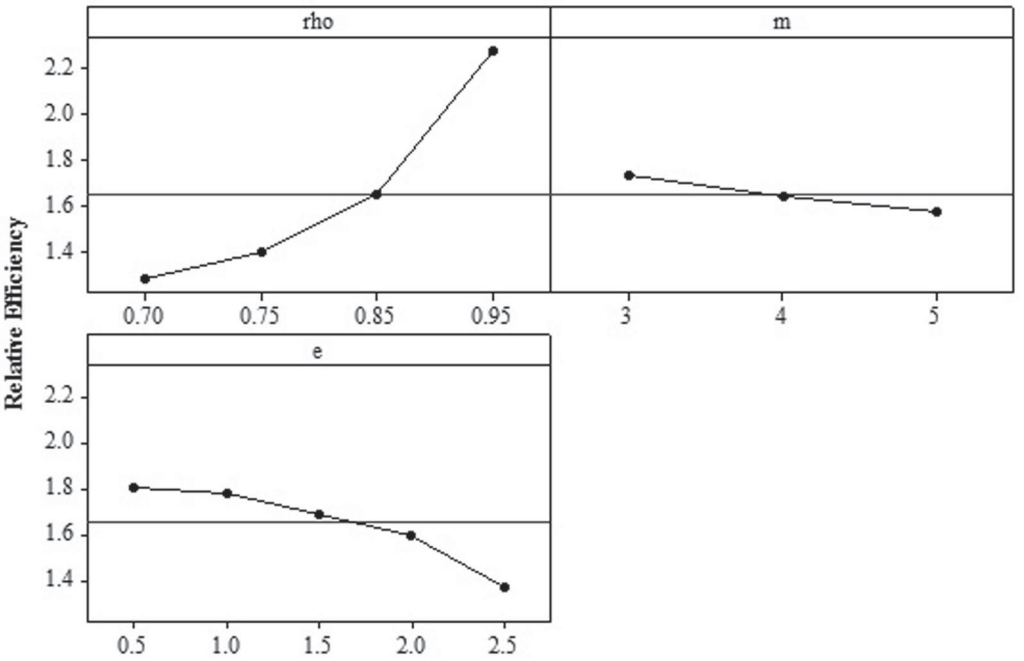


Figure 5. Graphs of RE values according to the correlation coefficients, set sizes, and *e* levels.

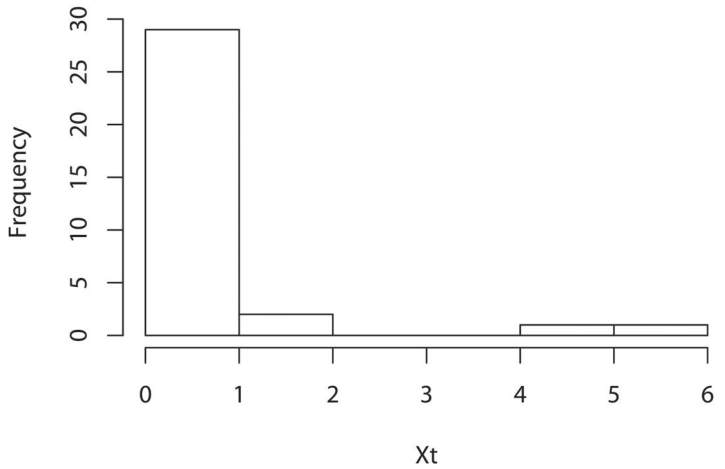


Figure 6. Histogram of the auxiliary variable confirmed cases having tie in COVID-19 data.

Table 5. Relative efficiencies for COVID-19 data set.

	$m=5$	$m=4$	$m=3$
RE_0	1.0726	1.1315	1.1894
RE_2	1.4653	1.5738	1.6583
RE_4	1.4655	1.5739	1.6582
RE_5	0	0	0
RE_6	0	0	0
RE_7	1.5345	1.6032	2.6654
RE_8	**2.5656**	**2.8715**	**3.6134**

We consider the number of confirmed cases in last 24 hours as the auxiliary variable and from Figure 6; we observe that there are ties for the number of confirmed cases in the last 24 hours. And since our variable is the number of patients, we know that there may be no patient, so the variable can take zero value.

In the simulation coded by the R program, the population parameters are given in Table 1 and set size is selected as $m = 3$, 4, and 5. The values of the mean estimators are obtained for the number of trials as 100000. The relative efficiency values are calculated by using (11). For the proposed estimator, $\hat{\mu}_7$, α and β are defined as kurtosis of auxiliary variable and 1, respectively. The RE results for COVID data are given in Table 5. It is clearly seen that the proposed estimators, $\hat{\mu}_7$ and $\hat{\mu}_8$, are quite efficient than the estimators in literature.

6. Conclusion

It is observed that the modified exponential ratio estimator is more efficient than the isotonic estimator proposed by Ozturk (2011). As expected, ratio estimators perform poorly when the correlation decreases and the number of ties increases. However, the proposed estimator gives better results under low correlations although it is a ratio estimator. In these conditions, where the efficiencies of other estimators decrease extremely, the isotonic part of the proposed exponential ratio estimator makes it more stable. In cases where ratio estimators cannot be calculated, such as working with zero-average variables or in cases where the auxiliary variable can be zero as COVID-19 data set, the proposed estimator gives lifelike results.

In cases where the correlation between Y and X_t is high and where the number of ties in X_t is low, it is seen that the simple ratio estimator under RSS gives better results than estimators using the tie information. In this condition, the $\hat{\mu}_7$ estimator may be preferred over the $\hat{\mu}_8$ estimator. Using the $\hat{\mu}_6$, which is a traditional ratio estimator, is not preferred under the RSS in the presence of the tie information, since the efficiency of the estimator decreases conspicuously when there are many ties in the population, as seen in Table 4.

For future studies on this subject, firstly it may be suggested to determine the number of e used in making ties with the DPS method. A formulation that would allow e to be selected according to the distribution parameters would be useful and interesting, instead of selecting e randomly. Secondly, it can also be considered if it is possible to obtain the mean square errors of estimators in theory. In addition, new methods proposed by Zamanzade and Wang (2020a) on this subject recently can be combined with the proposed estimator. The proposed estimator can be generalized with information about multi-concomitant variables. It can also be considered that tie method not only in the RSS method, but also in any case with ranking, such as operations research problems or nonparametric tests. Finally, for COVID-19 data of other countries, if there is a relationship between the number of patients diagnosed in the last 24 hours and the total number of deaths, the method can be tried on other countries' real data sets.

ORCID

Eda Gizem Koçyiğit http://orcid.org/0000-0002-0774-1376
Cem Kadilar http://orcid.org/0000-0003-4950-9660

References

Bahl, S., and R. Tuteja. 1991. Ratio and product type exponential estimators. *Journal of information and optimization sciences* 12 (1):159–64.

Chen, H., E. A. Stasny, and D. A. Wolfe. 2005. Ranked set sampling for efficient estimation of a population proportion. *Statistics in Medicine* 24 (21):3319–29. doi:10.1002/sim.2158.

Dastbaravarde, A., N. R. Arghami, and M. Sarmad. 2016. Some theoretical results concerning non-parametric estimation by using a judgment post-stratification sample. *Communications in Statistics - Theory and Methods* 45 (8):2181–203. doi:10.1080/03610926.2013.878355.

Dell, T. R., and J. L. Clutter. 1972. Ranked-set sampling theory with order statistics background. *Biometrics* 28 (2): 545–55. doi:10.2307/2556166.

Dümbgen, L., and E. Zamanzade. 2020. Inference on a distribution function from ranked set samples. *Annals of the Institute of Statistical Mathematics* 72 (1):157–85. doi:10.1007/s10463-018-0680-y.

Fligner, M. A., and S. N. MacEachern. 2006. Nonparametric two-sample methods for ranked-set sample data. *Journal of the American Statistical Association* 101 (475):1107–18. doi:10.1198/016214506000000410.

Frey, J. 2012. Nonparametric mean estimation using partially ordered sets. *Environmental and Ecological Statistics* 19 (3):309–26. doi:10.1007/s10651-012-0188-1.

Frey, J., and T. G. Feeman. 2012. An improved mean estimator for judgment post-stratification. *Computational Statistics & Data Analysis* 56 (2):418–26. doi:10.1016/j.csda.2011.08.006.

Frey, J., and T. G. Feeman. 2013. Variance estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics* 65 (3):551–69. doi:10.1007/s10463-012-0386-5.

Halls, L. K., and T. R. Dell. 1966. Trial of ranked-set sampling for forage yields. *Forest Science* 12:22–26.

Kadilar, C., Y. Unyazici, and H. Cingi. 2009. Ratio estimator for the population mean using ranked set Sampling. *Statistical Papers* 50 (2):301–309. doi:10.1007/s00362-007-0079-y.

Kvam, P. H. 2003. Ranked set sampling based on binary water quality data with covariates. *Journal of Agricultural, Biological, and Environmental Statistics* 8 (3):271–79. doi:10.1198/1085711032156.

MacEachern, S. N., E. A. Stasny, and D. A. Wolfe. 2004. Judgement post-stratification with imprecise rankings. *Biometrics* 60 (1):207–15. doi:10.1111/j.0006-341X.2004.00144.x.

Mahdizadeh, M., and E. Zamanzade. 2019. Efficient body fat estimation using multistage pair ranked set sampling. *Statistical Methods in Medical Research* 28 (1):223–34. doi:10.1177/0962280217720473.

McIntyre, G. A. 1952. A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research* 3 (4):385–90. doi:10.1071/AR9520385.

Nussbaum, B. D., and B. K. Sinha. 1997. Cost effective gasoline sampling using ranked set sampling. In Proceedings of the Section on Statistics and the Environment. American Statistical Association, 83–87.

Ozturk, O. 2007. Statistical inference under a stochastic ordering constraint in ranked set sampling. Journal of Nonparametric Statistics 19 (3):131–44. doi:10.1080/10485250701437232.

Ozturk, O. 2011. Sampling from partially rank-ordered sets. *Environmental and Ecological Statistics* 18 (4):757–79. doi:10.1007/s10651-010-0161-9.

Ozturk, O. 2014. Statistical inference for population quantiles and variance in judgment post-stratified samples. *Computational Statistics & Data Analysis* 77:188–205. doi:10.1016/j.csda.2014.02.021.

Ozturk, O., O. C. Bilgin, and D. A Wolfe. 2005. Estimation of population mean and variance in flock management: A ranked set sampling approach in a finite population setting. *Journal of Statistical Computation and Simulation* 75 (11):905–19.

Patil, G. P. 1995. Editorial: Ranked set sampling. *Environmental and Ecological Statistics* 2 (4):271–85. doi:10.1007/BF00569358.

Prasad, B. 1989. Some improved ratio type estimators of population mean and ratio in finite population sample surveys. *Communications in Statistics - Theory and Methods* 18 (1):379–92. doi:10.1080/03610928908829905.

Robertson, T. 1988. Order restricted statistical inference (No. 04; QA278. 7, R6.).

Samawi, H. M., and H. A. Muttlak. 1996. Estimation of ratio using rank set sampling. *Biometrical Journal* 38 (6): 753–64. doi:10.1002/bimj.4710380616.

Samawi, H. M., and O. A. M. Al-Sagheer. 2001. On the estimation of the distribution function using extreme and median ranked set sampling. *Biometrical Journal* 43 (3):357–73. doi:10.1002/1521-4036(200106)43:3<357::AID-BIMJ357>3.0.CO;2-Q.

Singh, H. P., R. Tailor, and N. K. Jatwa. 2011. Modified ratio and product estimators for population mean in systematic sampling. *Journal of Modern Applied Statistical Methods* 10 (2):424–35. doi:10.22237/jmasm/1320120180.

Wang, X., J. Lim, and S. L. Stokes. 2008. A nonparametric mean estimator for judgment poststratified data. *Biometrics* 64 (2):355–63. doi:10.1111/j.1541-0420.2007.00900.x.

Wang, X., K. Wang, and J. Lim. 2012. Isotonized CDF estimation from judgment poststratification data with empty strata. *Biometrics* 68 (1):194–202. doi:10.1111/j.1541-0420.2011.01655.x.

World Health Organization. 2020. Coronavirus disease 2019 (COVID-19) situation report–51. Geneva, Switzerland: World Health Organization. Accessed August 16, 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10

Zamanzade, E. 2016. An isotonized variance estimator for judgment post stratified data. *Journal of the Korean Statistical Society* 45 (1):25–32. doi:10.1016/j.jkss.2015.06.001.

Zamanzade, E., and X. Wang. 2017. Estimation of population proportion for judgment post-stratification. *Computational Statistics & Data Analysis* 112:257–69. doi:10.1016/j.csda.2017.03.016.

Zamanzade, E., and X. Wang. 2018. Proportion estimation in ranked set sampling in the presence of tie information. *Computational Statistics* 33 (3):1349–66. doi:10.1007/s00180-018-0807-x.

Zamanzade, E., and X. Wang. 2020a. Improved nonparametric estimation using partially ordered sets. In *Statistical methods and applications in forestry and environmental sciences*, Forum for Interdisciplinary Mathematics, ed. G. Chandra, R. Nautiyal, and H. Chandra, 57–77. Singapore: Springer.

Zamanzade, E., and X. Wang. 2020b. Estimating the area under a receiver operating characteristic curve using partially ordered sets. *International Journal of Bio-Statistics*. Advance online publication. doi:10.1515/ijb-2019-0127.