Mathematics

Research article

Parameter estimation in the Farlie–Gumbel–Morgenstern bivariate Bilal distribution via multistage ranked set sampling

S. P. Arun[1,*], M. R. Irshad[2], R. Maya[2], Amer I. Al-Omari[3,*] and Shokrya S. Alshqaq[4]

[1] Kerala University Library, Research Centre, University of Kerala, Thiruvananthapuram 695034, India

[2] Department of Statistics, Cochin University of Science and Technology, Cochin 682 022, Kerala, India; irshadmr@cusat.ac.in, publicationsofmaya@gmail.com

[3] Department of Mathematics, Faculty of Science, Al al-Bayt University, Mafraq 251113, Jordan

[4] College of Science, Jazan University, P.O.Box. 114, Jazan 45142, Kingdom of Saudi Arabia; salshekak@jazanu.edu.sa

* **Correspondence:** Email: arunsptvm@gmail.com, alomari_amer@yahoo.com.

Abstract: Ranked set sampling is a well-known and efficient method compared to simple random sampling for estimating population parameters. In this study, we focus on the challenge of estimating the scale parameter of the primary variable Z using a multistage ranked set sample obtained by ordering the marginal observations of an auxiliary variable W, where the pair (W, Z) follows the Farlie–Gumbel–Morgenstern bivariate Bilal distribution. Assuming that the dependence parameter ϕ is known, we introduce the best linear unbiased estimator for the scale parameter of the primary variable, utilizing a multistage ranked set sample. We also compare the efficiency of the proposed estimator with that of the maximum likelihood estimator based on the same number of measured units. It is found that the suggested estimators are more efficient than the classical estimators considered in this study.

1. Introduction

In 2013, [1] introduced a new lifetime distribution of one parameter, known as the Bilal distribution. The Bilal distribution is derived as a member of the family of distributions for the median of a random sample drawn from an exponential distribution. The cumulative distribution function (cdf) of the Bilal

distribution with the scale parameter σ is given by:

$$F(x; \sigma) = 1 - e^{-\frac{2x}{\sigma}} \left(3 - 2e^{-\frac{x}{\sigma}}\right); \ \sigma > 0, \ x \geq 0. \tag{1.1}$$

The corresponding probability density function (pdf) is defined by:

$$f(x; \sigma) = \frac{6}{\sigma} e^{-\frac{2x}{\sigma}} \left(1 - e^{-\frac{x}{\sigma}}\right); \ \sigma > 0, \ x \geq 0, \tag{1.2}$$

with survival function given by:

$$S(x; \sigma) = e^{-\frac{2x}{\sigma}} \left(3 - 2e^{-\frac{x}{\sigma}}\right); \ \sigma > 0, \ x \geq 0.$$

The Bilal distribution exhibits lower skewness and kurtosis compared to the exponential distribution and belongs to the class of distributions characterized as new, better than used in failure rates. Despite having only one parameter, the Bilal distribution demonstrates a strong capacity to fit two distinct real-world data sets: the first comprising 30 consecutive precipitation measurements (in inches) provided by [2], and the second consisting of waiting times before service for 100 bank customers, as reported by [3]. The closed-form expressions for all its statistical properties, includes key functions such as the quantile function, the hazard rate function (HRF), and a simple expression for the moments attracted several researchers and developed its various extensions. Abd-Elrahman and Niazi [4] proposed various estimators of the parameter for the Bilal distribution based on using Type-2 censored sample. To address the unimodal HRF of the Bilal distribution, Abd-Elrahman [5] proposed a two-parameter generalization referred to as the general Bilal distribution. A three-parameter generalization, called the Harris extended Bilal distribution was introduced, by [6], and its various properties have been discussed. Irshad et al. [7] proposed the Marshal–Olkin Bilal distribution, its applications in statistical process control, and the associated minification process. The proficiency of the univariate Bilal distribution has been firmly established in the literature, demonstrating its superiority over competing models in both theoretical and applied perspectives.

Maya et al. [8] introduced a bivariate version of the one-parameter Bilal distribution using the Morgenstern framework, called the Farlie–Gumbel–Morgenstern bivariate Bilal distribution (FGMBBD), and studied its inferential aspects based on concomitants of order statistics (COS). The pdf of FGMBBD is expressed as follows:

$$f(w, z) = \begin{cases} \frac{36}{\sigma_1 \sigma_2} e^{-\frac{2w}{\sigma_1}} \left(1 - e^{-\frac{w}{\sigma_1}}\right) e^{-\frac{2z}{\sigma_2}} \left(1 - e^{-\frac{z}{\sigma_2}}\right) \\ \times \left[1 + \phi \left(2e^{-\frac{2w}{\sigma_1}} \left\{3 - 2e^{-\frac{w}{\sigma_1}}\right\} - 1\right) \left(2e^{-\frac{2z}{\sigma_2}} \left\{3 - 2e^{-\frac{z}{\sigma_2}}\right\} - 1\right)\right], \\ \qquad w > 0, z > 0; \sigma_1 > 0, \sigma_2 > 0; -1 \leq \phi \leq 1. \\ 0, \ \text{otherwise.} \end{cases} \tag{1.3}$$

Figure 1 presents 3D plots of the FGMBBD pdf for various parameter values. We can see that the distribution has many shapes depending on the parameters values.
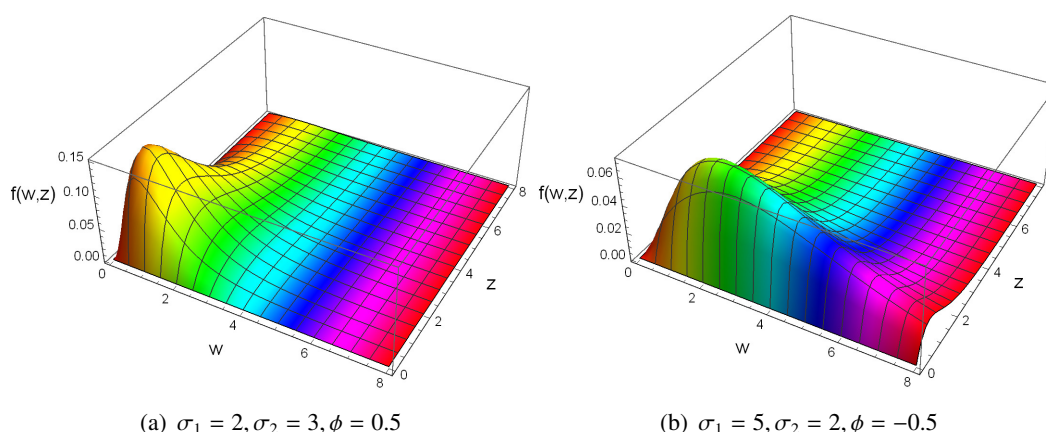
(a) $\sigma_1 = 2, \sigma_2 = 3, \phi = 0.5$ (b) $\sigma_1 = 5, \sigma_2 = 2, \phi = -0.5$

Figure 1. 3D plots of the FGMBBD pdf for several values of the parameters.

Arun et al. [9] evaluated the performance of the FGMBBD in comparison to the well-known FGM bivariate exponential distribution using two real-world data sets. The first data set, sourced from [10], consists of mineral content measurements in the dominant ulna (X) and non-dominant ulna (Y) of 25 elderly women. The second data set, also from [10], includes tail length (X) and wing length (Y) measurements in millimeters for a sample of 45 female hook-billed kites.

The marginal distributions of the variables W and Z are univariate Bilal distributions, with their respective pdf's given by:

$$f_W(w) = \frac{6}{\sigma_1} e^{-\frac{2w}{\sigma_1}} \left(1 - e^{-\frac{w}{\sigma_1}}\right); \quad if \ \ \sigma_1 > 0, \ w \geq 0,$$

and

$$f_Z(z) = \frac{6}{\sigma_2} e^{-\frac{2z}{\sigma_2}} \left(1 - e^{-\frac{z}{\sigma_2}}\right); \quad if \ \ \sigma_2 > 0, \ z \geq 0. \tag{1.4}$$

Clearly, $E(W) = \frac{5}{6}\sigma_1, Var(W) = \frac{13}{36}\sigma_1^2,$

$$E(Z) = \frac{5}{6}\sigma_2, \tag{1.5}$$

and

$$Var(Z) = \frac{13}{36}\sigma_2^2. \tag{1.6}$$

Note that the mean and variance values of Z depend on σ_2. The ranked set sampling (RSS) method, originally developed by [11], is designed to improve the precision of the sample mean as an estimator of the population mean. In this approach, a total of n sets of units, each containing n units, are selected. The units within each set are ordered using a judgmental method or a cost-effective technique that does not involve actual measurements of the selected observations. Subsequently, the unit ranked first in the first set is measured, followed by the unit ranked second in the second set, and this process continues until the unit ranked n largest in the n-th set is measured. The observations obtained through this process form a ranked set sample (rss) of size n. Hence, we only measured n rss units: $X_{1(1:n)}, X_{2(2:n)}, \ldots, X_{n(n:n)}$, which are obtained from n^2 units. For simplicity, we will use $X_{(i:n)}$ to denote the rss units.

Let the random variable (rv) X have a pdf $f(x)$ and cdf $F(x)$, with mean μ and variance σ^2. Consider a simple random sample (srs) of size n, denoted as $X_1, X_2, \ldots, X_n$. Let $X_{j(i:n)}$ represent the ith largest unit in a set of size n in the jth sample. Takahasi and Wakimoto [12] demonstrated that the rss estimator of the population mean, when based on perfect ranking, is unbiased and outperforms the estimator based on srs because it has a lower variance. The mean estimators based on srs and rss are, respectively given by: $\bar{X}_{SRS} = \dfrac{1}{n} \sum_{i=1}^{n} X_i$, and $\bar{X}_{RSS} = \dfrac{1}{n} \sum_{i=1}^{n} X_{(i:n)}$, with respective variances given by $\mathrm{Var}\left(\bar{X}_{SRS}\right) = \dfrac{\sigma^2}{n}$, and

$$\mathrm{Var}\left(\bar{X}_{RSS}\right) = \frac{\sigma^2}{n} - \frac{1}{n^2} \sum_{i=1}^{n} (\mu_{(i:n)} - \mu)^2, \text{ where } \mu_{(i:k)} = \int_{-\infty}^{\infty} x f_{(i:k)}(x) dx. \text{ Clearly, } \mathrm{Var}\left(\bar{X}_{RSS}\right) \le \mathrm{Var}\left(\bar{X}_{SRS}\right).$$

Stokes [13] introduced an alternative RSS scheme designed for situations where the primary variable of interest, denoted as Z, is challenging to measure directly, while an auxiliary variable, W, which is correlated with Z can be easily measured. The process involves selecting n independent bivariate sets, each containing n units. In the first set, the Z variable corresponding to the smallest ordered W is measured, followed by the Z variable associated with the second smallest W in the second set, and so on, until the Z variable corresponding to the largest W in the n-th set is measured. The resulting measurements on the Z variable from this new set of n units, selected according to the described method, constitute a rss, as proposed by [13]. Let $W_{(i:n)i}$ represent the observation on the auxiliary variable W from the unit obtained from the i-th set represent $Z_{[i:n]i}$ is used to denote the corresponding measurement of the study variable Z for that unit. Thus, $Z_{[i:n]i}$ for $i = 1, 2, \ldots, n$, collectively constitutes the rss of size n. David and Nagaraja [14] mentioned $Z_{[i:n]i}$ as the concomitant variable of the i-th order statistic obtain from the i-th sample. Using Stoke's procedure of RSS, [15] proposed estimators of the scale parameter associated with the Z variable, when (W, Z) follows the FGMBBD.

Suppose that the bivariate random vector (W, Z) follows the FGMBBD with pdf given in (1.3). Draw a rss using [13] scheme. Let $W_{(i:n)i}$ represent the observation obtained on the auxiliary variate W from the ith unit of the rss, and let $Z_{[i:n]i}$ denote the corresponding measurement of the variable related to $W_{(i:n)i}, i = 1, 2, \cdots, n$. Hence, $Z_{[i:n]i}$ is the ith COS of a random sample of size n drawn from the FGMBBD, with pdf given by (see, [8]) as:

$$f_{[i:n]}(z) = \frac{6}{\sigma_2} e^{-\frac{2z}{\sigma_2}} \left(1 - e^{-\frac{z}{\sigma_2}}\right) \left[1 + \phi \frac{(n - 2i + 1)}{(n + 1)} \left(2e^{-\frac{2z}{\sigma_2}} \left\{3 - 2e^{-\frac{z}{\sigma_2}}\right\} - 1\right)\right]. \tag{1.7}$$

Figure 2 shows 3D plots of the $f_{[i:n]}(z)$ for some choices of $n = 10, i = 1, 10, \phi = -0.9, 0.9$. The figure reflects the effect of these selections on the shape of this function.
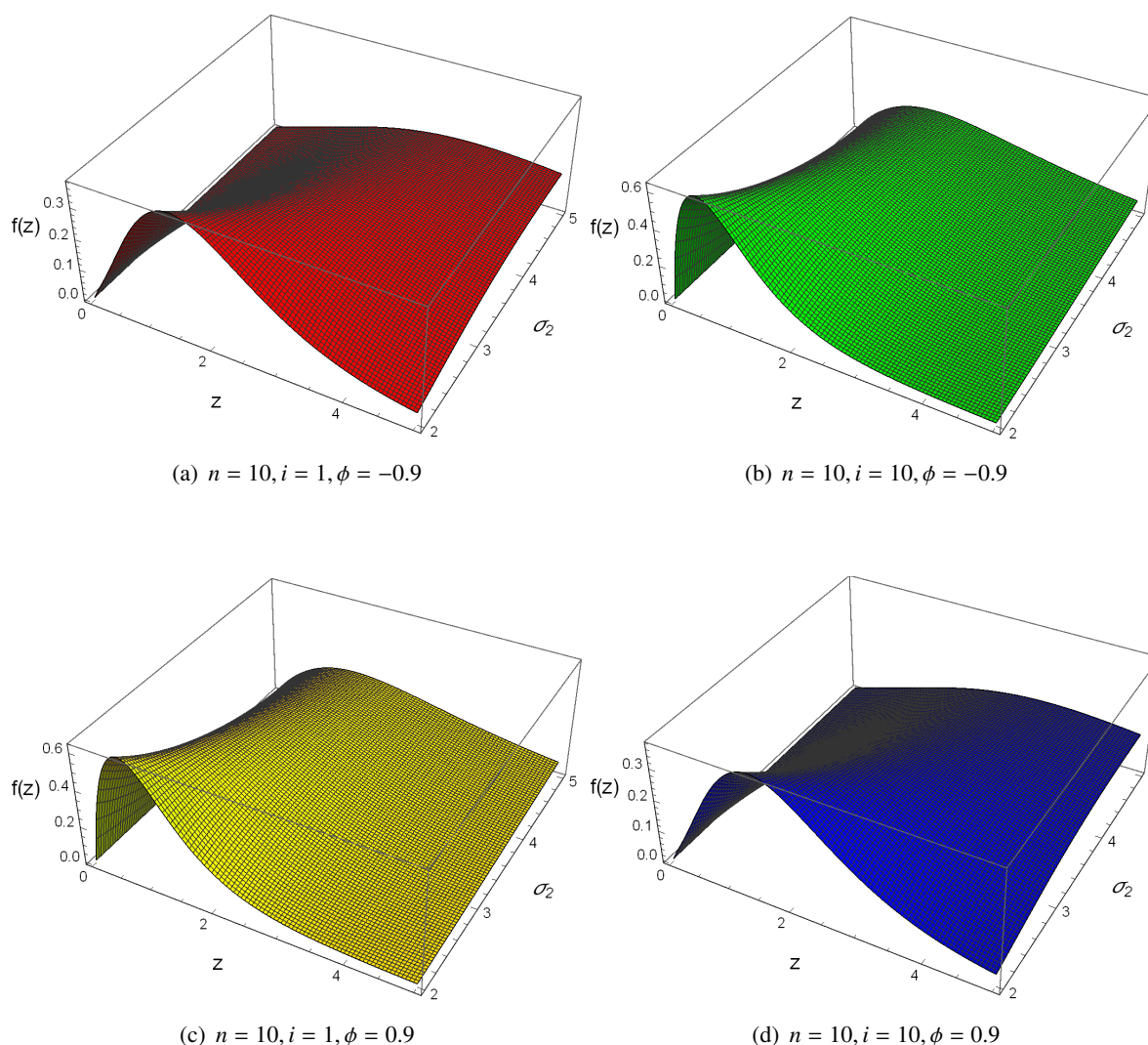
Figure 2. 3D plots of the function $f_{[i:n]}(z)$ of the FGMBBD with $2 \le \sigma_2 \le 5$.

The mean and variance of $Z_{[i:n]i}$ for $i = 1, 2, \cdots, n$, are obtained as

$$E[Z_{[i:n]i}] = \sigma_2 \left[\frac{5}{6} - \frac{19}{60} \phi \frac{(n - 2i + 1)}{(n + 1)} \right], \tag{1.8}$$

and

$$Var[Z_{[i:n]i}] = \sigma_2^2 \left[\frac{13}{36} - \frac{253}{1800} \phi \frac{(n - 2i + 1)}{(n + 1)} - \frac{361}{3600} \phi^2 \frac{(n - 2i + 1)^2}{(n + 1)^2} \right]. \tag{1.9}$$

Since, $Z_{[i:n]i}$ and $Z_{[j:n]j}$ for $i \ne j$ are arising from two independent samples, we obtain

$$Cov[Z_{[i:n]i}, Z_{[j:n]j}] = 0, i \ne j.$$

Al-Saleh and Al-kadiri [16] extended McIntyre's [11] concept of RSS, by introducing the double-stage ranked set sampling (DSRSS). Their findings affirmed that estimators based on the DSRSS scheme exhibit greater efficiency in estimating population parameters compared to those based on

both RSS and simple random sampling (SRS) schemes based on the same number of measured units. In a subsequent work, Al-Saleh and Al-Omari [17] extended the DSRSS to the multistage ranked set sampling (MSRSS) and demonstrated an increase in the precision of MSRSS estimators compared to both DSRSS and RSS estimators without needing to increase the sample size. Al-Saleh [18] introduced the concept of steady-state RSS and its corresponding parametric inference. Jemain and Al-Omari [19,20] proposed multistage percentile and quartile RSS methods for estimating the population mean. Jemain [21] suggested multistage median RSS for estimating the population median. For a deeper exploration of RSS and its variations, refer to [22], [23] for the varied L RSS method, [24] for an analysis of generalized robust-regression-type estimators under different RSS methods, and [25] for research on estimating a decreasing mean residual life based on RSS, with applications to survival analysis. Zamanzade et al. [26] considered nonparametric estimation of the mean residual lifetime in RSS with a concomitant variable. Zamanzade et al. [27] considered a ranked-based estimator of the mean past lifetime.

The aforementioned generalizations of RSS play a crucial role in developing inferential aspects of parent bivariate distributions in both theoretical and applied perspectives. Consequently, the objectives of this study are as follows:
(1) To identify the unit that has maximum Fisher information (FI) when ϕ is positive and negative.
(2) Building on the insights from the FI, we define the MSRSS and a steady-state RSS. Using these methodologies, we estimate the scale parameter associated with the study variable Z.
(3) To compare the efficiency of the proposed estimator based on MSRSS and steady-state RSS with the maximum likelihood estimator (MLE).

The rest of this paper is structured as follows. The FI included in the concomitant of a specific order statistic from a random sample arising from a distribution provides essential insight for selecting the most warranted unit(s) from a group of units to establish an appropriate RSS. Thus, the FI about the parameter σ_2, included in the concomitant of the ith order statistic from a random sample of size n drawn from the FGMBBD, has been derived and is presented in Section 2. As per the knowledge disseminated from the FI, we have identified that for the FGMBBD, the maximum amount of information about the parameter σ_2 is contained in the concomitant of the largest order statistic or the concomitant of the smallest order statistic, according as ϕ is positive or negative. Accordingly, we define an MSRSS in Section 2 and used its observations to introduce the best linear unbiased estimator (BLUE) of σ_2 when the dependence parameter ϕ is positive and negative. Here, we have additionally obtained the efficiency of the estimator of σ_2 based on the MSRSS and compared it with the MLE of σ_2. In Section 3, we proposed BLUE of σ_2 based on steady-state RSS and compared the efficiency of the same with respect to the MLE of σ_2. The article concludes in Section 4 with some suggestions for future works.

2. Estimation of σ_2 using MSRSS

Here, we define the MSRSS and propose the BLUE of σ_2 when ϕ is positive and negative based on observations generated by the MSRSS. The MSRSS at r stages (see, [17]) is defined as follows:
(1) Randomly select n^{r+1} sample units from the target population, where r denotes the number of stages in the MSRSS process. These selected units are then randomly divided into n^{r-1} sets, each containing n^2 units.

(2) For each set from Step 1, adopt Stoke's RSS scheme, to obtain an RSS of size n. This step results in n^{r-1} ranked sets, each of size n.

(3) Randomly organize the n^{r-1} ranked sets, each of size n obtained from Step 2, into n^{r-2} sets, each of size n^2. Without performing any actual measurements, apply the RSS scheme to each of the n^{r-2} sets to produce n^{r-2} second-stage ranked sets, each of size n.

(4) The process continues, without conducting any actual measurements, until we obtain an rth stage ranked set of size n.

(5) Finally, the n identified units from Step 4 are actually quantified for the variable of interest.

To identify the most suitable unit from a group of units for measurements, we derive the FI regarding the scale parameter σ_2 found in the concomitant of the ith order statistic from the FGMBBD.

The pdf of the concomitant of the ith order statistic given in (1.7) can be written as:

$$f_{[i:n]}(z) = \frac{6}{\sigma_2} e^{\frac{-2z}{\sigma_2}} (1 - e^{\frac{-z}{\sigma_2}}) \left[d + 2(1-d)e^{\frac{-2z}{\sigma_2}}(3 - 2e^{\frac{-z}{\sigma_2}}) \right],$$

where $d = d(\phi, i, n) = 1 - \phi \frac{(n-2i+1)}{(n+1)}$. By taking the logarithm, and then the first derivative partially with respect to σ_2, we obtain:

$$\frac{\partial \log f_{[i:n]}(z)}{\partial \sigma_2} = \frac{1}{\sigma_2} \left[-1 + \frac{2z}{\sigma_2} - \frac{ze^{\frac{-z}{\sigma_2}}}{\sigma_2(1 - e^{\frac{-z}{\sigma_2}})} + \frac{12(1-d)ze^{\frac{-2z}{\sigma_2}}(1 - e^{\frac{-z}{\sigma_2}})}{\sigma_2\{d + 2(1-d)e^{\frac{-2z}{\sigma_2}}(3 - 2e^{\frac{-z}{\sigma_2}})\}} \right].$$

Thus,

$$\left(\frac{\partial \log f_{[i:n]}(z)}{\partial \sigma_2} \right)^2 = \frac{1}{\sigma_2^2} \left[-1 + 2u - \frac{ue^{-u}}{(1 - e^{-u})} + \frac{12(1-d)ue^{-2u}(1 - e^{-u})}{\{d + 2(1-d)e^{-2u}(3 - 2e^{-u})\}} \right]^2,$$

where $u = \frac{z}{\sigma_2}$. The pdf of the transformed rv $U = \frac{Z}{\sigma_2}$ is obtained by

$$g(u) = 6e^{-2u}(1 - e^{-u}) \left[d + 2(1-d)e^{-2u}(3 - 2e^{-u}) \right].$$

Then, the FI about the parameter σ_2 contained in the concomitant of the ith order statistic is given by:

$$J_{\sigma_2}(Z_{[i:n]}, \phi) = E \left(\frac{\partial \log f_{[i:n]}(Z)}{\partial \sigma_2} \right)^2$$
$$= \sigma_2^{-2} q(d),$$

where

$$q(d) = \int_0^\infty \left[-1 + 2u - \frac{ue^{-u}}{(1 - e^{-u})} + \frac{12(1-d)ue^{-2u}(1 - e^{-u})}{\{d + 2(1-d)e^{-2u}(3 - 2e^{-u})\}} \right]^2$$
$$\times 6e^{-2u}(1 - e^{-u}) \left[d + 2(1-d)e^{-2u}(3 - 2e^{-u}) \right] du. \tag{2.1}$$

Now, fix $\phi > 0$. Since, $d = 1 - \phi \frac{(n-2i+1)}{(n+1)}$, $i = 1, 2, \cdots, n$, the values of d lie in the interval $0 \le 1 - \phi \le d \le 1 + \phi \le 2$. A graph of the $q(d)$ on the interval $0 \le 1 - \phi \le d \le 1 + \phi \le 2$ is presented in Figure 3.
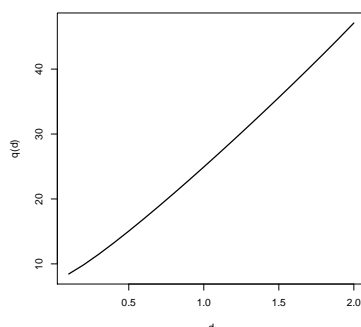
Figure 3. Graph of the function $q(d)$.

From the graph, it is easily observed that the maximum value of the $q(d)$ reaches at the upper extreme point of d, that is when $i = n$. Thus, one can conclude that when $\phi > 0$, the maximum FI is obtained in the concomitants of the largest order statistic. Hence, we consider the estimator based on the concomitants of the largest order statistic. Since $d(\phi, i, n) = d(-\phi, n - i + 1, n)$, we have $J_{\sigma_2}(Z_{[i:n]}, \phi) = J_{\sigma_2}(Z_{[n-i+1:n]}, -\phi)$. Hence $\phi < 0$, the maximum FI is obtained in the concomitants of the smallest order statistic.

By assuming that the rv (W, Z) has a FGMBBD with pdf given in (1.3), where Z is the variable of primary interest and W is an auxiliary variable, in this section first we consider $\phi > 0$ and carry out a MSRSS based on measurements made on an auxiliary variate to choose the rss to estimate σ_2 contained in FGMBBD based on the measurements made on the variable of primary interest. At each stage, from each set, we select the unit with the highest value on the auxiliary variable as the units of the ranked sets, aiming to maximize the FI about σ_2 for the final chosen rss.

Let $X_i^{(r)}, i = 1, 2, \cdots, n$ be the MSRSS units chosen at the rth stage. Since, the actual measurement of the auxiliary variable on each unit $X_i^{(r)}, i = 1, 2, \cdots, n$ has the largest value, we can write $Z_{[n:n]i}^{(r)}$ as the value measured on the variable of the primary interest to $X_i^{(r)}, i = 1, 2, \cdots, n$. Thus, easily one can verify that $Z_{[n:n]i}^{(r)}$ is the same as that of $Z_{[n^r:n^r]}$, the concomitant of the largest order statistic of n^r independent and identically distributed bivariate rvs from the FGMBBD. Also, $Z_{[n:n]i}^{(r)}, i = 1, 2, \cdots, n$ are independently distributed with a pdf given by:

$$f_{[n:n]i}^{(r)}(z; \phi) = \frac{6}{\sigma_2} e^{-\frac{2z}{\sigma_2}} \left(1 - e^{-\frac{z}{\sigma_2}}\right) \left[1 + \phi\left(\frac{n^r - 1}{n^r + 1}\right)\left(1 - 2e^{-\frac{2z}{\sigma_2}}\left\{3 - 2e^{-\frac{z}{\sigma_2}}\right\}\right)\right]. \tag{2.2}$$

Thus, the mean and variance of $Z_{[n:n]i}^{(r)}, i = 1, 2, \cdots, n$ are obtained as:

$$E[Z_{[n:n]i}^{(r)}] = \sigma_2 \left[\frac{5}{6} + \frac{19}{60}\phi\left(\frac{n^r - 1}{n^r + 1}\right)\right], \tag{2.3}$$

and

$$Var[Z_{[n:n]i}^{(r)}] = \sigma_2^2 \left[\frac{13}{36} + \frac{253}{1800}\phi\left(\frac{n^r - 1}{n^r + 1}\right) - \frac{361}{3600}\phi^2\left(\frac{n^r - 1}{n^r + 1}\right)^2\right]. \tag{2.4}$$

If we write

$$\zeta_{n^r} = \frac{5}{6} + \frac{19}{60}\phi\left(\frac{n^r - 1}{n^r + 1}\right), \tag{2.5}$$

and

$$\psi_{n^r} = \frac{13}{36} + \frac{253}{1800}\phi\left(\frac{n^r - 1}{n^r + 1}\right) - \frac{361}{3600}\phi^2\left(\frac{n^r - 1}{n^r + 1}\right)^2. \tag{2.6}$$

Then, for $i \leq i \leq n$, (2.3) and (2.4) can be written as:

$$E[Z^{(r)}_{[n:n]i}] = \sigma_2\zeta_{n^r}, \tag{2.7}$$

and

$$Var[Z^{(r)}_{[n:n]i}] = \sigma_2^2\psi_{n^r}. \tag{2.8}$$

Since $Z^{(r)}_{[n:n]i}$ and $Z^{(r)}_{[n:n]j}$ (for $i \neq j$) are measurements on Z made from two units involved in two independent samples, we have

$$Cov(Z^{(r)}_{[n:n]i}, Z^{(r)}_{[n:n]j}) = 0, \text{ for } i \neq j. \tag{2.9}$$

Assume that $\mathbf{Z}^{(r)}_{[n,n]} = (Z^{(r)}_{[n:n]1}, Z^{(r)}_{[n:n]2}, \cdots, Z^{(r)}_{[n:n]n})'$. Then, by using (2.7), we obtain the mean vector of $\mathbf{Z}^{(r)}_{[n,n]}$ as:

$$E[\mathbf{Z}^{(r)}_{[n,n]}] = \sigma_2\zeta_{n^r}\mathbf{1}, \tag{2.10}$$

and by using (2.8) and (2.9), the dispersion matrix of $\mathbf{Z}^{(r)}_{[n,n]}$ can be obtained as:

$$D[\mathbf{Z}^{(r)}_{[n,n]}] = \sigma_2^2\psi_{n^r}\mathbf{I}, \tag{2.11}$$

where $\mathbf{1}$ is a column vector of n ones and $\mathbf{I}$ is a unit matrix of order n. If $\phi > 0$ is involved in ζ_{n^r} and ψ_{n^r}, then (2.10) and (2.11) together defines a generalized Gauss–Markov setup, and hence the BLUE of σ_2 is obtained as:

$$\tilde{\sigma}_2^{n(r)} = \frac{1}{n\zeta_{n^r}}\sum_{i=1}^{n}Z^{(r)}_{[n:n]i}, \tag{2.12}$$

and its variance is obtained as

$$Var(\tilde{\sigma}_2^{n(r)}) = \frac{\psi_{n^r}}{n(\zeta_{n^r})^2}\sigma_2^2. \tag{2.13}$$

If we take $r = 1$ in the MSRSS method elucidated above, in this case, we get the usual single-stage RSS. Hence, the BLUE $\tilde{\sigma}_2^{n(1)}$ of σ_2 is given by

$$\tilde{\sigma}_2^{n(1)} = \frac{1}{n\zeta_n}\sum_{i=1}^{n}Z_{[n:n]i}, \tag{2.14}$$

with variance given by

$$Var(\tilde{\sigma}_2^{n(1)}) = \frac{\psi_n}{n(\zeta_n)^2}\sigma_2^2, \tag{2.15}$$

where we write $Z_{[n:n]i}$ instead of $Z^{(1)}_{[n:n]i}$ that represents the measurement of the variable of primary interest of the unit selected by the RSS. Also, ζ_n and ψ_n obtained by putting $r = 1$ in (2.5) and (2.6), respectively. Irshad et al. [15] computed the asymptotic variance of the MLE $\hat{\sigma}_2$ of σ_2. Using those values, we have evaluated the ratio $e(\tilde{\sigma}_2^{n(1)}|\hat{\sigma}_2) = \frac{Var(\hat{\sigma}_2)}{Var(\tilde{\sigma}_2^{n(1)})}$ for $\phi = 0.25(0.25)1$; $n = 2(2)20$ as a measure

of the efficiency of our estimator $\tilde{\sigma}_2^{n(1)}$ compared to the MLE $\hat{\sigma}_2$ based on n observations and the results are given in Table 1. From the table one can see that the efficiency increases with increases in ϕ and n. For example, when $\phi = 0.5$ and $n = 2, 8$, we have $\tilde{\sigma}_2^{n(1)} = 1.0596$ and 1.1773, respectively. Also, for $n = 6$ and $\phi = 0.25, 1$, the efficiency values are $\tilde{\sigma}_2^{n(1)} = 1.0720$ and 1.3580, respectively. However, it is clear that there is not much difference in the efficiency values for negative and positive values of ϕ.

Table 1. Efficiencies of the estimators $\tilde{\sigma}_2^{n(1)}$ and $\tilde{\sigma}_2^{1(1)}$ relative to $\hat{\sigma}_2$ of σ_2.

n	ϕ	$e(\tilde{\sigma}_2^{n(1)}|\hat{\sigma}_2)$	ϕ	$e(\tilde{\sigma}_2^{1(1)}|\hat{\sigma}_2)$
	0.25	1.0302	-0.25	1.0302
2	0.50	1.0596	-0.50	1.0596
	0.75	1.0863	-0.75	1.0863
	1.00	1.1026	-1.00	1.1030
	0.25	1.0588	-0.25	1.0588
4	0.50	1.1259	-0.50	1.1259
	0.75	1.1989	-0.75	1.1989
	1.00	1.2702	-1.00	1.2710
	0.25	1.0720	-0.25	1.0720
6	0.50	1.1577	-0.50	1.1577
	0.75	1.2563	-0.75	1.2563
	1.00	1.3580	-1.00	1.3600
	0.25	1.0800	-0.25	1.0800
8	0.50	1.1773	-0.50	1.1773
	0.75	1.2901	-0.75	1.2901
	1.00	1.4123	-1.00	1.4150
	0.25	1.0858	-0.25	1.0858
10	0.50	1.3273	-0.50	1.3273
	0.75	1.3119	-0.75	1.3119
	1.00	1.4461	-1.00	1.4490
	0.25	1.0882	-0.25	1.0882
12	0.50	1.1950	-0.50	1.1950
	0.75	1.3292	-0.75	1.3292
	1.00	1.4733	-1.00	1.4733
	0.25	1.0882	-0.25	1.0882
14	0.50	1.2026	-0.50	1.2026
	0.75	1.3395	-0.75	1.3395
	1.00	1.4979	-1.00	1.4979
	0.25	1.0909	-0.25	1.0909
16	0.50	1.2105	-0.50	1.2105
	0.75	1.3475	-0.75	1.3475
	1.00	1.5122	-1.00	1.5122
	0.25	1.0951	-0.25	1.0951
18	0.50	1.2119	-0.50	1.2119
	0.75	1.3606	-0.75	1.3606
	1.00	1.5249	-1.00	1.5249
	0.25	1.0928	-0.25	1.0928
20	0.50	1.2170	-0.50	1.2170
	0.75	1.3583	-0.75	1.3583
	1.00	1.5309	-1.00	1.5309

3. Estimation of σ_2 based on steady-state ranked set sampling

Al-Saleh [18] has proposed the steady-state RSS by letting r go to $+\infty$. If we apply the steady state RSS, then the asymptotic distribution of $Z_{[n:n]i}^{(r)}$ is obtained from (2.2) with a pdf given by

$$f_{[n:n]i}^{(\infty)}(z; \phi) = \frac{6}{\sigma_2} e^{-\frac{2z}{\sigma_2}} \left(1 - e^{-\frac{z}{\sigma_2}}\right) \left[1 + \phi\left(1 - 2e^{-\frac{2z}{\sigma_2}}\left\{3 - 2e^{-\frac{z}{\sigma_2}}\right\}\right)\right]. \tag{3.1}$$

Clearly, $Z_{[n:n]i}^{(\infty)}, i = 1, 2, \cdots, n$ are independent and identically distributed (iid) rvs, each with pdf defined in (3.1). Then, $Z_{[n:n]i}^{(\infty)}, i = 1, 2, \cdots, n$, may be considered as steady-state rss of size n. Therefore, from (2.3) and (2.4), the mean and variance of $Z_{[n:n]i}^{(\infty)}$ for $i = 1, 2, \cdots, n$ are obtained as:

$$E[Z_{[n:n]i}^{(\infty)}] = \sigma_2 \left[\frac{5}{6} + \frac{19}{60}\phi\right], \tag{3.2}$$

and

$$Var[Z_{[n:n]i}^{(\infty)}] = \sigma_2^2 \left[\frac{13}{36} + \frac{253}{1800}\phi - \frac{361}{3600}\phi^2\right]. \tag{3.3}$$

Let $\mathbf{Z}_{[n:n]}^{(\infty)} = (Z_{[n:n]1}^{(\infty)}, Z_{[n:n]2}^{(\infty)}, \cdots, Z_{[n:n]n}^{(\infty)})'$. Then, the BLUE $\tilde{\sigma}_2^{n(\infty)}$ based on $\mathbf{Z}_{[n:n]}^{(\infty)}$ and the variance of $\tilde{\sigma}_2^{n(\infty)}$ is obtained by taking the limit as $r \to \infty$ in (2.12) and (2.13), respectively, and are given by

$$\tilde{\sigma}_2^{n(\infty)} = \frac{1}{n\left[\frac{5}{6} + \frac{19}{60}\phi\right]} \sum_{i=1}^{n} Z_{[n:n]i}^{(\infty)}, \tag{3.4}$$

and

$$Var(\tilde{\sigma}_2^{n(\infty)}) = \frac{\left[\frac{13}{36} + \frac{253}{1800}\phi - \frac{361}{3600}\phi^2\right]}{n\left[\frac{5}{6} + \frac{19}{60}\phi\right]^2}\sigma_2^2. \tag{3.5}$$

Using the results of [15], we obtain the efficiency of $\tilde{\sigma}_2^{n(\infty)}$ relative to $\hat{\sigma}_2$ by taking the ratio of $Var(\hat{\sigma}_2)$ with respect to $Var(\tilde{\sigma}_2^{n(\infty)})$ and is given by:

$$\begin{aligned}
E(\tilde{\sigma}_2^{n(\infty)}|\hat{\sigma}_2) &= \frac{Var(\hat{\sigma}_2)}{Var(\tilde{\sigma}_2^{n(\infty)})} \\
&= \frac{I_{\sigma_2}^{(-1)}(\phi)\left[\frac{5}{6} + \frac{19}{60}\phi\right]^2}{\left[\frac{13}{36} + \frac{253}{1800}\phi - \frac{361}{3600}\phi^2\right]}.
\end{aligned}$$

Thus, the efficiency $e(\tilde{\sigma}_2^{n(\infty)}|\hat{\sigma}_2)$ is independent of the sample size n, and that is for a fixed ϕ, the $e(\tilde{\sigma}_2^{n(\infty)}|\hat{\sigma}_2)$ has the same values for all n.

Additionally, we have computed the value of $e(\tilde{\sigma}_2^{n(\infty)}|\hat{\sigma}_2)$ for $\phi = 0.25(0.25)1$, and the results are given in Table 2. From the table one can observe that the efficiency of $\tilde{\sigma}_2^{n(\infty)}$ increases as ϕ increases. Moreover, the estimator $\tilde{\sigma}_2^{n(\infty)}$ possesses the highest efficiency among other estimators of σ_2 considered in this study, and the value of the efficiencies varies from 1.1073 to 1.6352.

Table 2. Efficiencies of the estimators $\tilde{\sigma}_2^{n(\infty)}$ and $\tilde{\sigma}_2^{1(\infty)}$ relative to $\hat{\sigma}_2$.

ϕ	$e(\tilde{\sigma}_2^{n(\infty)}\|\hat{\sigma}_2)$	ϕ	$e(\tilde{\sigma}_2^{1(\infty)}\|\hat{\sigma}_2)$
0.25	1.1073	-0.25	1.1073
0.50	1.2472	-0.50	1.2472
0.75	1.4223	-0.75	1.4226
1.00	1.6352	-1.00	1.6365

As mentioned earlier in the case of FGMBBD, when $\phi < 0$, the concomitant of the smallest order statistic contains the maximum FI about σ_2. Consequently, we explore an MSRSS in this scenario, where we select a unit of a sample with the smallest value on the auxiliary variable as the units of ranked sets at each step and from each set, with the goal of maximizing FI on the final selected rss.

Let $Z_{[1:n]i}^{(r)}, i = 1, 2, \cdots, n$, represents the value measured on the primary variable of interest for the units selected at the rth stage of the MSRSS. Then, it is simple to see that $Z_{[1:n]i}^{(r)}$, the concomitant of the least order statistic of n^r iid bivariate random variables with FGMBBD, has the same distribution as $Z_{[1:n^r]}$. Moreover, $Z_{[1:n]i}^{(r)}, i = 1, 2, \cdots, n$, are also independently distributed with pdf given by:

$$f_{[1:n]i}^{(r)}(z; \phi) = \frac{6}{\sigma_2} e^{-\frac{2z}{\sigma_2}} \left(1 - e^{-\frac{z}{\sigma_2}}\right) \left[1 - \phi\left(\frac{n^r - 1}{n^r + 1}\right)\left(1 - 2e^{-\frac{2z}{\sigma_2}}\left\{3 - 2e^{-\frac{z}{\sigma_2}}\right\}\right)\right]. \tag{3.6}$$

Clearly, from (2.2) and (3.6), we have:

$$f_{[1:n]i}^{(r)}(z; \phi) = f_{[n:n]i}^{(r)}(z; -\phi). \tag{3.7}$$

Hence, it is obvious that $E(Z_{[n:n]i}^{(r)})$ for $\phi > 0$ and $E(Z_{[1:n]i}^{(r)})$ for $\phi < 0$ are the same. Similarly, $Var(Z_{[n:n]i}^{(r)})$ for $\phi > 0$ and $Var(Z_{[1:n]i}^{(r)})$ for $\phi < 0$ are the same. Consequently, if $\tilde{\sigma}_2^{1(r)}$ is the BLUE of σ_2 involved in the FGMBBD for $\phi < 0$ based on the MSRSS observations $Z_{[1:n]i}^{(r)}, i = 1, 2, \cdots, n$, then the coefficients of $Z_{[1:n]i}^{(r)}, i = 1, 2, \cdots, n$ in the BLUE $\tilde{\sigma}_2^{1(r)}$ for $\phi < 0$ the same as the coefficients of $Z_{[n:n]i}^{(r)}, i = 1, 2, \cdots, n$ in the BLUE $\tilde{\sigma}_2^{n(r)}$ for $\phi > 0$.

Further, we have $Var(\tilde{\sigma}_2^{1(r)}) = Var(\tilde{\sigma}_2^{n(r)})$. Hence, $Var(\tilde{\sigma}_2^{1(1)}) = Var(\tilde{\sigma}_2^{n(1)})$ and $Var(\tilde{\sigma}_2^{1(\infty)}) = Var(\tilde{\sigma}_2^{n(\infty)})$, where $\tilde{\sigma}_2^{1(1)}$ is the BLUE of σ_2 for $\phi < 0$ is the usual single-stage rss observations $Z_{[1:n]i}^{(r)}, i = 1, 2, \cdots, n$ and $\tilde{\sigma}_2^{1(\infty)}$ is the BLUE of σ_2 for $\phi < 0$ based on the steady-state rss observations $Z_{[1:n]i}^{(\infty)}, i = 1, 2, \cdots, n$. We have computed the efficiency $e(\tilde{\sigma}_2^{1(\infty)}|\hat{\sigma}_2)$ of the BLUE $\tilde{\sigma}_2^{1(r)}$ relative to $\hat{\sigma}_2$, the MLE of σ_2 for $\phi = -1, -0.75, -0.50, -0.25; n = 2(2)20$ and the results are included in Table 1. Also, we have computed $e(\tilde{\sigma}_2^{1(\infty)}|\hat{\sigma}_2)$ for $\phi = -1, -0.75, -0.50, and -0.25$, and the results are given in Table 2. Based on Table 2, one can infer that the efficiency increases as $|\phi|$ increases, and the value of the efficiency varies from 1.1073 to 1.6365.

Remark 3.1. If (W, Z) follows a FGMBBD with pdf given in (1.3), then the correlation coefficient between W and Z is given by

$$Corr(W, Z) = \frac{361}{1300}\phi, \quad -1 \le \phi \le 1.$$

Obviously, the correlation coefficient is maximum when $|\phi|$ is as large as 1. From the tables, we can observe that the efficiencies of all estimators increase as $|\phi|$ increases for a given sample size. Consequently, we observe that more information about σ_2 can be extracted from the RSS when $|\phi|$ is large. Therefore, we conclude that concomitant ranking is more effective for estimating σ_2 when the modulus value of the dependence parameter ϕ is large (ϕ approaches to ± 1).

Remark 3.2. The assumption that ϕ is known may be considered unrealistic in certain real-life situations. [8] addressed this issue by proposing a moment-type estimator for ϕ, which is expressed as follows:

$$\hat{\phi} = \begin{cases} -1, & if \ \varkappa < \frac{-361}{1300} \\ \frac{1300}{361}\varkappa, & if \ \frac{-361}{1300} \leq \varkappa \leq \frac{361}{1300} \\ 1, & if \ \varkappa > \frac{361}{1300}, \end{cases}$$

where $\varkappa$ is the sample correlation from $(W_{(i:n)i}, Z_{[i:n]i})$ for $i = 1, 2, \cdots, n$.

4. Conclusions

This study focuses on estimating the scale parameter of the primary variable Z using the MSRSS, where the ranking is based on the marginal observations of an auxiliary variable W. It is assumed that the joint distribution of (W, Z) follows the FGMBBD. Considering the dependence parameter ϕ is known, the BLUE for the scale parameter of Z under MSRSS as well as steady-state RSS schemes is presented. The efficiency of the suggested estimators is compared to the MLE based on the same number of measured units. The results indicate that the proposed estimators are more efficient than the classical estimators considered in this study. For further studies, one can estimate other parameters or use different modifications of RSS to estimate the same parameter.

Author contributions

Conceptualization, S.P.A., M.R.I., R.M., S.S.A., and A.I.A; methodology, S.P.A., M.R.I., and R.M.; software, S.P.A., M.R.I., A.I.A., and R.M.; validation, S.P.A., M.R.I., R.M., S.S.A., and A.I.A.; formal analysis, M.R.I., S.P.A., R.M., and A.I.A; writing-original draft preparation, M.R.I., S.P.A., R.M., A.I.A., and S.S.A.; writing-review and editing, M.R.I., S.P.A., R.M., A.I.A., and S.S.A.; visualization, M.R.I., S.P.A., R.M., A.I.A., and S.S.A. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

There is no conflict of interest with the publication of this work.

References

1. A. M. Abd-Elrahman, Utilizing ordered statistics in lifetime distributions production: a new lifetime distribution and applications, *J. Probab. Stat. Sci.*, **11** (2013), 153–164.

2. D. Hinkley, On quick choice of power transformation, *Appl. Stat.*, **26** (1977), 67–69. https://doi.org/10.2307/2346869

3. M. E. Ghitany, B. Atieh, S. Nadarajah, Lindley distribution and its application, *Math. comput. simul.*, **78** (2008), 493–506. https://doi.org/10.1016/j.matcom.2007.06.007

4. A. M. Abd-Elrahman, S. F. Niazi, Approximate Bayes estimators applied to the Bilal model, *J. Egypt. Math. Soc.*, **25** (2017), 65–70. https://doi.org/10.1016/j.joems.2016.05.001

5. A. M. Abd-Elrahman, A new two-parameter lifetime distribution with decreasing, increasing or upside-down bathtub-shaped failure rate, *Commun. Stat.-Theory Methods*, **46** (2017), 8865–8880. https://doi.org/10.1080/03610926.2016.1193198

6. R. Maya, M. R. Irshad, M. Ahammed, C. Chesneau, The Harris extended Bilal distribution with applications in hydrology and quality control, *AppliedMath*, **3** (2023), 221–242. https://doi.org/10.3390/appliedmath3010013

7. M. R. Irshad, M. Ahammed, R. Maya, A. I. Al-Omari, Marshal-Olkin Bilal distribution with associated minification process and accptance sampling plans, *Hacet. J. Math. Stat.*, **53** (2023), 201–229. https://doi.org/10.15672/hujms.1143156

8. R. Maya, M. R. Irshad, S. P. Arun, Farlie–Gumbel–Morgenstern bivariate Bilal distribution and its inferential aspects using concomitants of order statistics, *J. Probab. Stat. Sci.*, **19** (2021), 1–20.

9. S. P. Arun, C. Chesneau, R. Maya, M. R. Irshad, Farlie–Gumbel–Morgenstern bivariate moment exponential distribution and its inferences based on concomitants of order statistics, *Stats*, **6** (2023), 253–267. https://doi.org/10.3390/stats6010015

10. R. A. Johnson, D. W. Wichern, *Applied multivariate statistical analysis*, Pearson College Div, Subsequent edition, 2002.

11. G. A. McIntyre, A method for unbiased selective sampling using ranked sets, *Australian J. Agr. Res.*, **3** (1952), 385–390. https://doi.org/10.1071/AR9520385

12. K. Takahasi, K. Wakimoto, On unbiased estimates of the population mean based on the sample stratified by means of ordering, *Ann. Inst. Stat. Math.*, **20** (1968), 1–31.

13. S. L. Stokes, Ranked set sampling with concomitant variables, *Commun. Stat.-Theory Methods*, **6** (1977), 1207–1211. https://doi.org/10.1080/03610927708827563

14. H. A. David, H. N. Nagaraja, *Order statistics*, 3 Eds., John Wiley and Sons, New York, 2003.

15. M. R. Irshad, R. Maya, A. I. Al-Omari, A. A. Hanadeh, S. P. Arun, Estimation of a parameter of Farlie–Gumbel–Morgenstern bivariate Bilal distribution by ranked set sampling, *Reliability: Theory Appl.*, **18** (2023), 129–140.

16. M. F. Al-Saleh, M. Al-Kadiri, Double-ranked set sampling, *Stat. Probabil. Lett.*, **48** (2000), 205–212. https://doi.org/10.1016/S0167-7152(99)00206-0

17. M. F. Al-Saleh, A. I. Al-Omari, Multistage ranked set sampling, *J. Stat. plan. Infer.*, **102** (2002), 273–286. https://doi.org/10.1016/S0378-3758(01)00086-6

18. M. F. Al-Saleh, Steady-state ranked set sampling and parametric estimation, *J. Stat. plan. Infer.*, **123** (2004), 83–95. https://doi.org/10.1016/S0378-3758(03)00139-3

19. A. A. Jemain, A. I. Al-Omari, Multistage percentile ranked set samples, *Adv. Appl. Stat.*, **7** (2007), 127–139.

20. A. A. Jemain, A. I. Al-Omari, Multistage quartile ranked set samples, *Pakistan J. Stat.*, **23** (2007), 11–22.

21. A. A. Jemain, A. I. Al-Omari, K. Ibrahim, Multistage median ranked set sampling for estimating the population median, *J. Math. Stat.*, **3** (2007), 58–64.

22. A. I. Al-Omari, C. N. Bouza, Review of ranked set sampling: modifications and applications, *Revista Investigacion Operacional*, **35** (2014), 215–240.

23. A. Haq, J. Brown, E. Moltchanova, A. I. Al-Omari, Varied L ranked set sampling scheme, *J. Stat. Theory Pract.*, **9** (2015), 741–767. https://doi.org/10.1080/15598608.2015.1008606

24. N. Koyuncu, A. I. Al-Omari, Generalized robust-regression-type estimators under different ranked set sampling, *Math. Sci.*, **15** (2021), 29–40. https://doi.org/10.1007/s40096-020-00360-7

25. E. Zamanzade, E. Zamanzade, A. Parvardeh, Estimation of a decreasing mean residual life based on ranked set sampling with an application to survival analysis, *Int. J. Biostat.*, **20** (2024), 571–583. https://doi.org/10.1515/ijb-2023-0051

26. E. Zamanzade, M. Mahdizadeh, H. M. Samawi, Nonparametric estimation of mean residual lifetime in ranked set sampling with a concomitant variable, *J. Appl. Stat.*, **51** (2024), 2512–2528. https://doi.org/10.1080/02664763.2023.2301334

27. E. Zamanzade, M. Asadi, A. Parvardeh, E. Zamanzade, A ranked-based estimator of the mean past lifetime with an application, *Stat. Papers*, **64** (2023), 161–177. https://doi.org/10.1007/s00362-022-01314-y