DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

RANKED SET SAMPLING FOR ENVIRONMENTAL RESEARCH IN SUSTAINABLE SMART CITIES: ESTIMATION OF DEPENDENCE MEASURES

by
Yusuf Can SEVİL

June, 2023
İZMİR

RANKED SET SAMPLING FOR ENVIRONMENTAL RESEARCH IN SUSTAINABLE SMART CITIES: ESTIMATION OF DEPENDENCE MEASURES

A Thesis Submitted to the
Graduate School of Natural And Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics

by
Yusuf Can SEVİL

June, 2023
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**RANKED SET SAMPLING FOR ENVIRONMENTAL RESEARCH IN SUSTAINABLE SMART CITIES: ESTIMATION OF DEPENDENCE MEASURES**" completed by **YUSUF CAN SEVİL** under supervision of **ASSOC. PROF. DR. TUĞBA YILDIZ** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

....

Assoc. Prof. Dr. Tuğba YILDIZ

Supervisor

....

Thesis Committee Member | Thesis Committee Member

....

Examining Committee Member | Examining Committee Member

Prof. Dr. Okan FISTIKOĞLU
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

I would like to express my deepest to my advisor, Assoc. Prof. Dr. Tuğba YILDIZ for her unwavering support, guidance, and invaluable insights throughout the course of this research. Her expertise and mentorship have been instrumental in shaping my understanding and fostering my academic growth. She has had a significant influence on my entire life, aside from being my thesis advisor. She has helped me finish my thesis by supporting me and keeping me motivated even when I falter. She provided me with some excellent advice that was helpful for my thesis as well as other aspects of my life. Being one of her students makes me feel privileged and honored.

Apart from my advisor, I would like to thank Prof. Dr. Burcu HÜDAVERDİ AKTAŞ for her unending encouragement, support and direction, which helped me see new possibilities. Also, I would like to express Asst. Prof. Dr. Celal Cem SARIOĞLU and Prof. Dr. Çağın KANDEMİR ÇAVAŞ for their contributions in the development of the thesis.

I am also indebted to my friend, Aylin GÖÇOĞLU for her continuous encouragement, simulating discussions, and assistance during this journey. She has always made my days better with kind gestures. I would also like to thank my brother from another mother, Tolga YAMUT for his academic suggestions that gave me a different perspective and most importantly for his social support. I would like to thank my friend, Ayça ÖLMEZ who is a Potterhead for her everlasting support, motivation and hours of conversations on interesting topic. On the other hand, I would like to express my deepest gratitude to my friend Sami AKDENİZ for his friendship. I also would like to thank Çağla ÇINAR for sharing my stress and inspiring me. A special thank goes to my friends at the Bookstore for their supports and encouragements. They are the unsung heroes of this thesis. Their camaraderie and shared experiences have made this research endeavor more enjoyable and rewarding.

I am indebted to family of the Department of Statistics at DEU, which I have always been honored to be a part of. I want to express my gratitude to my esteemed professors for all of their guidance.

I would like to express the Scientific and Technological Research Council of Turkey (TÜBİTAK) 2211/A National Ph.D. scholarship program and the Higher Education Council of Turkey (YÖK) 100/2000 Ph.D. scholarship program for supporting this thesis.

A special thank goes to my mother, Perihan SEVİL, my father, Hüseyin SEVİL, and my brother Ceyhun SEVİL for their endless support and patience. Their constant support and understanding have been a source of strength and motivation during both the highs and lows of this endeavor.

Yusuf Can SEVİL

RANKED SET SAMPLING FOR ENVIRONMENTAL RESEARCH IN SUSTAINABLE SMART CITIES: ESTIMATION OF DEPENDENCE MEASURES

ABSTRACT

In environmental problems, researchers face numerous obstacles and limitations. The main ones are encountered at the sampling step. For example, some of them are cost of sampling, measurement errors, the length of the sampling process, the inability to collect sampling units because of environmental factors, etc. Thus, it is important to select a sample that is representative of entire population. For this purpose, a single random sample can be selected. Here the issue is how well the population is represented by the sample that is selected at once. Against this problem, ranked set sampling (RSS) has been developed. RSS has been applied in many areas and has proven to be a cost effective sampling scheme.

Let (X, Y) be random vectors with the joint probability density function $f(x, y)$ and joint cumulative distribution $F(x, y)$. The aim of the dissertation is to estimate the dependence parameter of $f(x, y)$. In environmental studies, there are a number of issues in that should be examined by considering the correlation between random variables (X and Y) such as storage depth and translation time of water, various measures of groundwater quality (pH, nitrate, sulfate, and chloride), flood volume and its duration, flood peak and its volume, and so on. In this dissertation, maximum likelihood estimates based on SRS, RSS, generalized modified RSS (GMRSS) and extreme RSS (ERSS) are investigate for the dependence parameter of Gumbel (Type I) bivariate exponential and Farlie-Gumbel-Morgenstern (FGM) type bivarite gamma distribution. Also, likelihood ratio statistics are examined for testing the independence. An application is given for likelihood ratio test in Gumbel (Type I) bivariate exponential distribution. Furthermore, imperfect ranking case is examined in the RSS procedure. For this purpose, maximum pseudo-likelihood based on RSS is used in estimating the dependence parameter of bivariate normal distribution. On the other hand, some non-parametric bootstrap techniques for confidence interval of

dependence parameter are defined.

Keywords: Algorithm for sampling data, bivariate normal distribution, concomitants, dependence parameter, FGM type bivariate gamma distribution, Gumbel's (Type I) bivariate exponential distribution, likelihood ratio test, maximum likelihood, maximum pseudo-likelihood, order statistics, ranked set sampling

SÜRDÜRÜLEBILIR AKILLI KENTLERDE ÇEVRESEL ARAŞTIRMALAR IÇIN SIRALI KÜME ÖRNEKLEMESI: BAĞIMLILIK ÖLÇÜLERININ KESTIRIMI

ÖZ

Çevre problemlerinde, araştırmacılar birçok engelle ve kısıtlamayla karşılaşırlar. Bunların başlıcaları örnekleme aşamasında ortaya çıkar. Örneğin, örnekleme maliyeti, ölçüm hataları, örnekleme süresi, çevresel faktörler nedeniyle örnekleme birimlerinin toplanamaması gibi durumlar bunlardan bazılarıdır. Bu nedenle, bütün kitleyi temsil eden bir örneklem seçmek önemlidir. Bu amaçla, tek bir rastgele örneklem seçilebilir. Burada sorun, kitlenin bir seferde seçilen örneklem tarafından ne kadar iyi temsil edildiğidir. Bu soruna karşı, sıralı küme örneklemesi (SKÖ) geliştirilmiştir. SKÖ birçok alanda uygulanmış ve uygun maliyetli bir örnekleme yöntemi olduğu kanıtlamıştır.

(X,Y), bileşik olasılık yoğunluk fonksiyonunu $f(x,y)$ ve bileşik kümülatif dağılım fonksiyonunu $F(x,y)$ ile rastgele vektörler olsun. Tezin amacı, $f(x,y)$'nin bağımlılık parametresinin tahmin edilmesidir. Çevre çalışmalarında, suyun depolama derinliği ve aktarma süresi, yer altı suyu kalitesinin çeşitli ölçümleri (pH, nitrat, sülfat ve klorür), sel hacmi ve süresi, sel zirvesi ve sel hacmi gibi rastgele değişkenler (X ve Y) arasındaki korelasyonu dikkate alarak incelenmesi gereken birçok konu bulunmaktadır. Bu tezde, Gumbel (Tip I) iki değişkenli üstel ve Farlie-Gumbel-Morgenstern (FGM) tipi iki değişkenli gamma dağılımının bağımlılık parametresi için SKÖ, genelleştirilmiş geliştirilmiş SKÖ ve uç SKÖ'ye dayalı en çok olabilirlik tahminleri araştırılmıştır. Ayrıca, bağımsızlık testi için olabilirlik oranı istatistikleri incelenmektedir. Gumbel (Tip I) iki değişkenli üstel dağılımında olabilirlik oranı testi için bir uygulama sunulmuştur. Ayrıca, RSS prosedüründe kusurlu sıralama durumu incelenmiştir. Bu amaçla, iki değişkenli normal dağılımın bağımlılık parametresinin tahmininde SKÖ'ye dayalı en çok pseudo-olabilirlik yöntemi kullanılmıştır. Diğer taraftan, bağımlılık parametresinin güven aralığı için bazı parametrik olmayan bootstrap teknikleri tanımlanmıştır.

Anahtar kelimeler: Örnekleme verileri için algoritma, iki değişkenli normal dağılım, eşlenikler, bağımlılık parametresi, FGM tipi iki değişkenli gamma dağılımı, Gumbel’ın (Tip 1) iki değişkenli üstel dağılımı, en çok olabilirlik, en çok pseudo-olabilirlik, sıra istatistikleri, sıralı küme örneklemesi

CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF SYMBOLS

k	: Set size
m	: Number of cycles
$\mathbb{S}_{rj}$	: rth set in jth cycle
$Y_{(r)j}$	: rth ordered unit in jth cycle
$f_{r:k}(y)$	: PDF of rth order statistic
$N(\mu_x, \sigma_x^2, \mu_y, \sigma_y^2, \rho)$	: Normal distribution with parameters means (μ_x,μ_y), variances (σ_x^2, σ_y^2) and dependence parameter (ρ)
$f(x, y)$	: Joint PDF of (X, Y)
$F(x, y)$	: Joint CDF of (X, Y)
$(X_{(r)j}, Y_{[r]j})$	: rth X-ordered unit and its Y-concomitant in the jth cycle
$Cor(X, Y)$	: Correlation between X and Y
$f_{r:k}\,(x, y)$	: Joint PDF of $(X_{(r)j}, Y_{[r]j})$
$g_{[r:k]}\,(y)$	: PDF of Y-concomitant
θ	: Dependence parameter (Gumbel's (Type I) bivariate exponential distribution)
$\bar{F}(x, y)$	: Joint survival function
$f\,(y\|x)$	: Conditional PDF of Y given $X = x$
$RSS_1(m, k)$	: Ranked set sample with one variable
$RSS_2(m, k)$	: Ranked set sample with two variables
$ERSS(m, k)$	: Extreme ranked set sample with two variables
$MRSS(m, k)$	: Modified ranked set sample with two variables
$GMRSS(R = r)$	: Generalized modified ranked set sample in case of $R = r$
$\hat{\theta}_{SRS}$	: ML estimator based on SRS for θ
$\hat{\theta}_{RSS}$	: ML estimator based on RSS for θ
$\hat{\theta}_{GMRSS(R=r)}$	: ML estimator based on GMRSS(R=r) for θ
$\hat{\theta}_{ERSS}$	: ML estimator based on ERSS for θ
$\mathscr{L}$	: Likelihood function
L	: Log-likelihood function

$\lambda(x, y)$	: LRT statistic
$I(\ )$	: Indicator function
$c_u(v)$	: Conditional distribution function for U given $U = u$
λ	: Dependence parameter in FGM type bivariate gamma distribution
a, b	: Shape parameters in FGM type bivariate gamma distribution
$\Gamma(.)$	: Gamma function
$\gamma(., .)$	: Incomplete gamma function
$\hat{\lambda}_{SRS}$	: ML estimator based on SRS for λ
$\hat{\lambda}_{RSS}$	: ML estimator based on RSS for λ
$\hat{\lambda}_{GMRSS(R=r)}$	: ML estimator based on GMRSS(R=r) for λ
$\hat{H}(x)$, $\hat{G}(y)$	: EDFs for $H(x)$ and $G(y)$
Φ	: CDF of standard normal
$\hat{\rho}_{SRS}$	: Sample correlation coefficient based on SRS
$\hat{\rho}_{RSS}$	: Sample correlation coefficient based on RSS
$\hat{\rho}_{ZM}$	: Sample correlation coefficient based on RSS in Zheng & Modarres
$\hat{\rho}_{HMZ}$	: Sample correlation coefficient based on RSS in Hui, Zheng & Modarres
$\hat{\rho}_{MPL}$	: Sample correlation coefficient based on SRS using MPL
$\hat{\rho}_{YT}$	: Sample correlation coefficient based on RSS using MPL
$\left(X^*_{(r)j}, Y^*_{[r]j}\right)$	: rth X-ordered unit and its Y-concomitant in the jth cycle (for bootstrap methods)
$\hat{\varrho}^{\mathbb{N}}$	: Sample correlation coefficient based on Method $\mathbb{N}$

ABBREVIATIONS

RSS	: Ranked set sampling
SRS	: Simple random sampling
MRSS	: Modified ranked set sampling
GMRSS	: Generalized modified ranked set sampling
ERSS	: Extreme ranked set sampling
BRSS	: Bivariate ranked set sampling
QRSS	: Quasi-random ranked set sampling
MERSS	: Moving extreme ranked set sampling
FGM	: Farlie-Gumbel-Morgenstern
ML	: Maximum likelihood
PDF	: Probability density function
CDF	: Cumulative distribution function
EDF	: Empirical distribution function
FI	: Fisher information
CI	: Confidence interval
CPU	: Central Process Unit
MSE	: Mean square error
RE	: Relative efficiency
LRT	: Likelihood ratio test
MPL	: Maximum pseudo-likelihood
MML	: Modified maximum likelihood
PML	: Pseudo maximum likelihood
CBRSSR	: Concomitant based bootstrap ranked set sampling
CMRBRSS	: Concomitant based mixed row bootstrap ranked set sampling
CBRSSR	: Concomitant based bootstrap ranked set sampling by row

CHAPTER ONE
INTRODUCTION

In the last two centuries, many economic activities have been industrialized, which has had a significant impact on the environment and human health. According to the World Health Organization, protecting human and animal populations from diseases should be a top priority of a global health agenda, particularly in the case of degenerative pathologies. Thus, studies on environmental pollution have become more common, especially in recent years. Mostly, water and soil samples are obtained for use in pollution research. Studies on animal tissues, on the other hand, are a more pragmatic and indirect alternative than research on soil and water. However, studies in the literature indicate that some metallic and metalloid pollutants are present in mammals and quickly accumulate in a variety of tissues, including the liver, kidneys, and hair. Ceruti et al. (2002) looked into the amount of lead in rodent kidneys. Reynolds et al. (2006) investigated pocket gophers, a rodent species, to look for signs of environmental metal pollution. For an integrated health risk assessment at a mining site, Jasso-Pineda et al. (2007) examined soil and rodent-type heavy metal accumulations. To study the effects of prolonged human exposure to environmental pollutants, Minamia et al. (2009) measured basal levels of heavy metals in the tissue of rodent organs. In order to investigate environmental pollution, Bortey-Sam et al. (2016) examined the levels of heavy metals accumulated in the kidneys and livers of lemmings. According to Hazratian et al. (2017), the Norway rat (Rattus norvegicus) can be used as a bioindicator for lead and cadmium accumulation.

Collecting data is a critical step especially for environmental studies. At this step, researchers face many challenges and constraints. Let us use an illustration to clarify. Say our objective is to identify the source of pollution in a particular area. On a square, rectangular, or triangular grid placed over the area to be sampled, evenly spaced sampling locations are first created. This process is referred to as grid sampling or systematic sampling (Shtiza & Tashko, 2009). Then, traps are set up at each corner of the grid. According to biologists, the success rate of trapping rodents is between 3% and 10%. The contaminants are determined by examining the samples from the captured animals' hair and other tissues in a laboratory setting. When considering the entire process, difficulties such as high transportation costs, poor capture rates, drawn-out laboratory studies, and measuring errors are encountered. It is important to select a sample that is representative of the area for all these difficulties to make sense. Therefore, researchers should follow a sampling design.

Bhave & Sadhwani (2022) provided an excellent literature review on sampling in environmental studies. This study looked at a variety of sampling techniques, including judgmental sampling, simple random sampling (SRS), stratified sampling, RSS, systematic sampling and grid sampling, adaptive cluster sampling, and composite sampling. The authors also applied the sampling strategies to a real data issue. In our work, we focus on the ranked set sampling (RSS) design, and it will be discussed in the following section.

1.1 Ranked Set Sampling

McIntyre (1952, 2005) suggested the RSS procedure as a cost-effective sampling strategy. In the interest of estimating the pasture yield in Australia, he used RSS design. Because harvesting the crops is necessary for precise yield measurements, it is expensive. First, the interested pasture was divided into the 25 quadrats. The 25 quadrats were then randomly assigned into the five sets. After that, the rth judgement ranked quadrat is selected from the rth set where $r = 1, \cdots, 5$. Keep in mind that the rth judgement ranked quadrat is selected after a visual assessment of crop yield across all quadrats in the rth set. Figure 1.1 illustrates how the five selected quadrats represent the entire population. Instead of RSS, the five quadrats can be selected by using SRS. In this situation, the five quadrats could have been selected from the entire population (from different quantiles) or from the right tail (among maximum quantiles) of the population distribution. Hence, it can be conclude that RSS provides an estimator that is not only unbiased but also demonstrates equal or greater efficiency compared to the estimator based on SRS.

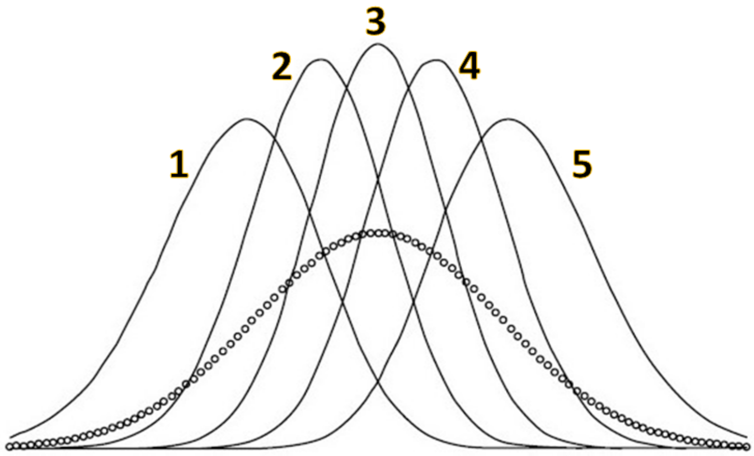


Figure 1.1 Dotted curve: The distribution of yields per quadrat and solid curves: distribution of samples from sets of five random samples that are of the same rank

Assume that k is the set size and m is the number of cycle to provide a general definition of RSS design. First, k^2 units are selected from the interested population. Then, these units are divided into the k sets at random. Let $\mathbb{S}_{rj} = \{Y_{r1j}, Y_{r2j}, \cdots, Y_{rkj}\}$ be the rth set in the jth cycle, $r = 1, 2, \cdots, k$ and $j = 1, 2, \cdots, m$. After that, without actually measuring, the units in each set are ranked from the smallest to the largest, $Y_{r(1)j} \leq Y_{r(2)j} \leq \cdots \leq Y_{r(k)j}$. Here, it is possible to rank the units in each set by using an auxiliary variable. This auxiliary variable could be the results of the preceding experiment or a different variable that is correlated with the variable of interest, Y. In addition, it is presumed that the measurement of the auxiliary variable is comparatively simpler and/or more cost-effective than that of the variable of interest. On the other hand, if an auxiliary variable is not available, the ranking procedure can be carried out by visual inspecting the units in each set. A scenario for the ranking process is provided in the example below.

Example 1 (Ozturk et al. (2005)) *Let's consider a scenario where we regularly monitor the biological growth of young sheep with the aim of enhancing meat quality and production. The problem is that young sheep require a lot of labor to restrain because they are very active animals during the measurement procedure. The weights at which the young sheep were born can be used as an auxiliary variable to rank the young sheep's current weights.*

Following the ranking process, $Y_{r(r)j}$ is selected for full measurement from the rth set. The first cycle is then complete and k measurements are obtained. If these steps are repeated m cycles, mk measurements are taken. Eventually, a ranked set sample is denoted by

$$RSS_1(k,m) = \begin{pmatrix} Y_{(1)1} & Y_{(2)1} & \cdots & Y_{(k)1} \\ Y_{(1)2} & Y_{(2)2} & \cdots & Y_{(k)2} \\ \vdots & \vdots & & \vdots \\ Y_{(1)m} & Y_{(2)m} & \cdots & Y_{(k)m} \end{pmatrix}$$

where RSS_1 stands for ranked set sample with one variable. Note that $Y_{(r)j}$ is used instead of $Y_{r(r)j}$ in the matrix for the sake of simplification. This procedure is also known as balanced RSS design since m order statistics are collected for each of the ranks $(1, 2, \cdots, k)$. We note that the parentheses () are utilized to denote the case of perfect ranking. If there is a potential for imperfect ranking, the bracket [] is employed in place of the parentheses (). In practice, the assumption of perfect ranking is unrealistic.

Let us discuss about the unbiasedness and efficiency properties of the mean estimator based on RSS in the example below.

Example 2 (Wolfe (2012)) *Let the set size is k and the number of cycle is m. A ranked set sample $RSS_1(k,m)$ is selected from a population with distribution function $F(y)$ and density $f(y)$. Under the perfect ranking case, the mean estimator based on RSS is*

$$\hat{\mu}_{RSS} = \frac{1}{mk}\sum_{j=1}^{m}\sum_{r=1}^{k} y_{(r)j}. \tag{1.1}$$

When the $m = 1$, $\hat{\mu}_{RSS} = \Sigma_{r=1}^{k} y_{(r)}/(mk)$. Let us prove the unbiasedness of the mean estimator.

$$E\left[\hat{\mu}_{RSS}\right] = \frac{1}{k}\sum_{r=1}^{k} E\left[Y_{(r)}\right] \tag{1.2}$$

where $Y_{(r)}$ has probability density function (PDF) of rth order statisic,

$$f_{r:k}(y) = \frac{k!}{(r-1)!(k-r)!} F^{r-1}(y)\left[1 - F(y)\right]^{k-r} f(y) \tag{1.3}$$

and

$$E\left[Y_{(r)}\right] = \int_{-\infty}^{\infty} y f_{r:k}(y)dy, \tag{1.4}$$

for $r = 1, 2, \cdots, k$. *Combining Eq. (1.2) and Eq. (1.3), the following equation is obtained.*

$$\begin{aligned} E\left[\hat{\mu}_{RSS}\right] &= \frac{1}{k}\sum_{r=1}^{k}\left\{\int_{-\infty}^{\infty} ky\binom{k-1}{r-1}\left[F(y)\right]^{r-1}\left[1-F(y)\right]^{k-r} f(y)dy\right\} \\ &= \int_{-\infty}^{\infty} yf(y)\left\{\sum_{r=1}^{k}\binom{k-1}{r-1}\left[F(y)\right]^{r-1}\left[1-F(y)\right]^{k-r}\right\}dy. \end{aligned} \tag{1.5}$$

Letting $i = r - 1$ *in (1.5),*

$$\sum_{r=1}^{k}\binom{k-1}{r-1}\left[F(y)\right]^{r-1}\left[1-F(y)\right]^{k-r} = \sum_{i=0}^{k-1}\binom{k-1}{i}\left[F(y)\right]^{i}\left[1-F(y)\right]^{(k-1)-i} = 1 \tag{1.6}$$

since the summation defined on i *covers the entire sample space for a binomial random variable with the parameters* $k-1$ *and* $p = F(y)$. *Then, it is seen that the* $\hat{\mu}_{RSS}$ *is unbiased.*

$$E\left[\hat{\mu}_{RSS}\right] = \int_{-\infty}^{\infty} kydy = \mu. \tag{1.7}$$

Now, we continue with the variance of $\hat{\mu}_{RSS}$.

$$V\left[\hat{\mu}_{RSS}\right] = \frac{1}{k^2}\sum_{r=1}^{k} V\left[Y_{(r)}\right] \tag{1.8}$$

Suppose that $\mu_{(r)} = E\left[Y_{(r)}\right]$ *for* $r = 1, 2, \cdots, k$ *and*

$$\begin{aligned} E\left[\left(Y_{(r)}-\mu\right)^2\right] &= E\left[\left(Y_{(r)}-\mu_{(r)}+\mu_{(r)}-\mu\right)^2\right] \\ &= E\left[\left(Y_{(r)}-\mu_{(r)}\right)^2\right] + \left(\mu_{(r)}-\mu\right)^2 \\ &= Var\left[Y_{(r)}\right] + \left(\mu_{(r)}-\mu\right)^2. \end{aligned} \tag{1.9}$$

The following equation is given by combining (1.8) and (1.9),

$$V\left[\hat{\mu}_{RSS}\right]=\frac{1}{k^{2}}\left\{\sum_{r=1}^{k}E\left[\left(Y_{(r)}-\mu\right)^{2}\right]-\sum_{r=1}^{k}\left(\mu_{(r)}-\mu\right)^{2}\right\} \tag{1.10}$$

$E\left[\left(Y_{(r)}-\mu\right)^{2}\right]$ *can be expressed as follows:*

$$\begin{aligned}\sum_{r=1}^{k}E\left[\left(Y_{(r)}-\mu\right)^{2}\right]&=\sum_{r=1}^{k}\int_{-\infty}^{\infty}k\left(y-\mu\right)^{2}\binom{k-1}{r-1}\left[F(y)\right]^{r-1}\left[1-F(y)\right]^{k-r}f(y)dy\\&=k\int_{-\infty}^{\infty}\left(y-\mu\right)^{2}f\left(y\right)\left\{\sum_{r=1}^{k}\binom{k-1}{r-1}\left[F(y)\right]^{r-1}\left[1-F(y)\right]^{k-r}\right\}dy.\end{aligned} \tag{1.11}$$

Applying the binomial distribution, we ascertain that the inner sum is equal to one.

$$\sum_{r=1}^{k}E\left[\left(Y_{(r)}-\mu\right)^{2}\right]=k\int_{-\infty}^{\infty}\left(y-\mu\right)^{2}f\left(y\right)dy=k\sigma^{2} \tag{1.12}$$

Using Eq. (1.10) and Eq. (1.12), it is evident that

$$V\left[\hat{\mu}_{RSS}\right]=\frac{1}{k^{2}}\left\{k\sigma^{2}-\sum_{r=1}^{k}\left(\mu_{(r)}-\mu\right)^{2}\right\}=\frac{\sigma^{2}}{k}-\frac{1}{k^{2}}\sum_{r=1}^{k}\left(\mu_{(r)}-\mu\right)^{2}. \tag{1.13}$$

We know that the variance of the mean estimator based on SRS is $V\left[\hat{\mu}_{SRS}\right]=\sigma^{2}/k$. *Thus, it is obvious that* $V\left[\hat{\mu}_{RSS}\right]\leq V\left[\hat{\mu}_{SRS}\right]$. *Each of the order statistics in* $RSS_1(k,m)$ *reduces to a random sample of size* m *as the ranking quality decrease, and as a result,* $\mu_{(r)}\rightarrow\mu$. *Under the imperfect ranking case,* $\hat{\mu}_{RSS}$ *is as efficient as* $\hat{\mu}_{SRS}$.

Following McIntyre (1952, 2005), Takahasi & Wakimoto (1968) examined the mean estimator based on RSS and demonstrated that $\hat{\mu}_{RSS}$ is an unbiased estimator with a smaller variance compared to its counterpart in SRS, irrespective of the ranking issue. Takahasi & Wakimoto (1968) clarified the relative efficiency (RE) of $\hat{\mu}_{RSS}$ with respect to $\hat{\mu}_{SRS}$ as follows:

$$1 \leq \frac{V\left[\hat{\mu}_{SRS}\right]}{V\left[\hat{\mu}_{RSS}\right]} \leq \frac{k+1}{2} \tag{1.14}$$

The impact of ranking errors on mean estimator based on RSS was examined by Dell & Clutter (1972). In this study, they defined a ranking error model $X = Y + \epsilon$ where X and ϵ are independent and $\epsilon \sim N(0, \sigma_\epsilon^2)$. By using the ranking error model, an auxiliary variable (X) can be generated. Following that, Ys are ranked by using the magnitudes of Xs. Ranking is perfect if $\sigma_\epsilon^2 = 0$; otherwise, is imperfect.

1.2 Ranked-based Sampling Designs

Developing ranked-based sampling designs is a remarkable topic in the literature. In this section, we focus on ranked-based sampling designs that use auxiliary variables. These designs are useful to obtain a sample from bivariate distributions. Also, these sampling designs are particularly preferred by authors studying on correlation estimation, ratio estimator, etc.

The RSS protocol with using auxiliary variable was first investigated by David & Levine (1972) and developed in detail by Stokes (1977). Supposed that (X, Y) are random variables with absolutely continuous joint CDF $F(x, y)$ and PDF $f(x, y)$. Assume that $\Omega = \{(u_1; X_1, Y_1), \cdots, (u_N; X_N, Y_N)\}$ is the population where u_c is the name (or an equivalent number) of the cth location and, where the variables X_c and Y_c are characteristics of the cth location, $c = 1, \cdots, N$. Suppose that the variable Y is difficult to measure and/or rank, but that the variable X can be at least easily and/or cheaply ranked in this population. RSS denotes ranked set sample with two variables and is obtained by using following steps:

(I) Select the k^2 locations from the Ω and these locations are divided into k sets at random;

$$\mathbb{S}_{rj} = \{u_{r1j}, u_{r2j}, \cdots, u_{rkj}\}$$

where $r = 1, \cdots, k$ and $j = 1, \cdots, m$.

Table 1.1 A brief review of the RSS literature

Subjects	Studies
Ranked-based sampling methods	Samawi et al. (1996), Muttlak (1997), Al-Saleh & Al-Kadiri (2000), Muttlak (2003), Al-Saleh & Zheng (2002), Gulay & Demirel (2019)
Bootstrap techniques	Chen et al. (2004), Hui et al. (2005), Alirezaei Dizicheh et al. (2021), Yamaguchi & Murakami (2021), Samawi & Chen (2021), Taconeli & de Lara (2022)
Estimating the population mean	Haq et al. (2014) and Singh et al. (2014)
Estimating the population variance	Bilgin et al. (2004), Chen & Lim (2011), Ozturk (2014), Ozturk & Demirel (2016) and Mahdizadeh & Zamanzade (2021)
Estimating the population proportion	Terpstra (2004), Terpstra & Liudahl (2004), Chen et al. (2007), Göçoğlu & Demirel (2019), Zamanzade & Mahdizadeh (2020) and Mahdizadeh & Zamanzade (2022)
Estimating the population distribution function	Nazari et al. (2014), Sevil (2017), Sevil & Yildiz (2017, 2020, 2021, 2023b, 2022a), Yildiz & Sevil (2018, 2019) and Zamanzade (2019)
Estimating the dependence parameter	Stokes (1980), Modarres & Zheng (2004), Al-Saleh & Samawi (2005), Zheng & Modarres (2006), Hanandeh & Al-Saleh (2013)
Regression	Patil et al. (1993), Samawi & Ababneh (2001), Chen & Wang (2004), Al-Nasser & Radaideh (2008)
Classification	Hatefi et al. (2014), Hatefi et al. (2015) and Hatefi et al. (2020)

(II) Rank the Xs from the smallest to the largest,

$$X_{r(1)j} \leq X_{r(1)j} \leq \cdots \leq X_{r(k)j}.$$

(III) Take measurements $X_{r(r)j}$ and its concomitant $Y_{r[r]j}$ from the $\mathbb{S}_{rj}$.

(IV) Repeat the steps (I)-(III), $j = 1, \cdots, m$,

$$RSS(k,m) = \begin{pmatrix} (X_{(1)1}, Y_{[1]1}) & (X_{(2)1}, Y_{[2]1}) & \cdots & (X_{(k)1}, Y_{[k]1}) \\ (X_{(1)2}, Y_{[1]2}) & (X_{(2)2}, Y_{[2]2}) & \cdots & (X_{(k)2}, Y_{[k]2}) \\ \vdots & \vdots & & \vdots \\ (X_{(1)m}, Y_{[1]m}) & (X_{(2)m}, Y_{[2]m}) & \cdots & (X_{(k)m}, Y_{[k]m}) \end{pmatrix}.$$

Stokes (1980) defined modified RSS (MRSS) which provides more efficient maximum likelihood (ML) estimator than ML estimators based on SRS and RSS. The first two steps in RSS and MRSS are identical. Following the second step,

(III) Take measurements $X_{r(1)j}$ and its concomitant $Y_{r[1]j}$ from the $\mathbb{S}_{rj}$ where $r = 1, \cdots, k$ and $j = 1, \cdots, m$.

(IV) Repeat the steps (I)-(III), $j = 1, \cdots, m$,

$$MRSS(k,m) = \begin{pmatrix} (X_{1(1)1}, Y_{1[1]1}) & (X_{2(1)1}, Y_{2[1]1}) & \cdots & (X_{k(1)1}, Y_{k[1]1}) \\ (X_{1(1)2}, Y_{1[1]2}) & (X_{2(1)2}, Y_{2[1]2}) & \cdots & (X_{k(1)2}, Y_{k[1]2}) \\ \vdots & \vdots & & \vdots \\ (X_{1(1)m}, Y_{1[1]m}) & (X_{2(1)m}, Y_{2[1]m}) & \cdots & (X_{k(1)m}, Y_{k[1]m}) \end{pmatrix}.$$

Also, maximum ranked pairs $X_{r(k)j}$ and $Y_{r[k]j}$ could be selected in step (III) of MRSS procedure. Sevil & Yildiz (2022b) established a method, named as generalized MRSS (GMRSS), in which any ranked pair can be selected. According to the definition of GMRSS, $X_{\tau(r)j}$ and its concomitant $Y_{r[\tau]j}$ are measured from the set $\mathbb{S}_{\tau j}$ in the step (III) of MRSS procedure where, $\tau = 1, \cdots, k$, $j = 1, \cdots, m$ and $r \in [1, k]$. Thus, it will be expressed as $GMRSS(R = r)$ in the remaining sections.

Also, Sevil & Yildiz (2022b) investigated ML estimator based on extreme RSS (ERSS) in case of even set size. This sampling procedure was suggested by Samawi et al. (1996) as a practical scheme. Due to the fact that this sampling strategy does not require for ranking units in the set. Both RSS and ERSS begin with the same two steps. After the second step,

(III) Take measurements $X_{r(1)j}$ and its concomitant $Y_{r[1]j}$ from the $\mathbb{S}_{rj}$ where $r = 1, \cdots, k/2$ and $j = 1, \cdots, m$.

(IV) Take measurements $X_{r(k)j}$ and its concomitant $Y_{r[k]j}$ from the $\mathbb{S}_{rj}$ where $r = (k/2) + 1, \cdots, k$ and $j = 1, \cdots, m$.

(V) Repeat the steps (I)-(IV), $j = 1, \cdots, m$,

$$ERSS(k,m) = \begin{pmatrix} (X_{1(1)1}, Y_{1[1]1}) & \cdots & (X_{k/2(1)1}, Y_{k/2[1]1}) & (X_{k/2+1(k)1}, Y_{k/2+1[k]1}) & \cdots & (X_{k(k)1}, Y_{k[k]1}) \\ (X_{1(1)2}, Y_{1[1]2}) & \cdots & (X_{k/2(1)2}, Y_{k/2[1]2}) & (X_{k/2+1(k)2}, Y_{k/2+1[k]2}) & \cdots & (X_{k(k)2}, Y_{k[k]2}) \\ \vdots & & \vdots & \vdots & & \vdots \\ (X_{1(1)m}, Y_{1[1]m}) & \cdots & (X_{k/2(1)m}, Y_{k/2[1]m}) & (X_{k/2+1(k)m}, Y_{k/2+1[k]m}) & \cdots & (X_{k(k)m}, Y_{k[k]m}) \end{pmatrix}.$$

As an extension of RSS, bivariate RSS (BVRSS) was developed by Al-Saleh & Zheng (2002). The steps of this sampling process are as follows:

(I) A random sample of size k^4 is selected from the population Ω for a given set size k, and it is divided into k^2 pools of size k^2 each. Here, each pool is a square matrix with k rows and k columns.

(II) For each k rows in the first pool, the minimum ranked value is determined by judgement with respect to the variable X.

(III) For each of the k minimum values in Step (II), the pair that corresponds to the smallest ranked value of the variable Y determined by judgement is selected. This pair is denoted by $(1, 1)$.

(IV) Following the step, the pair corresponding to the first minimum ranked value in the variable X and the second minimum ranked value in the variable Y is determined. These two values are represented by $(1, 2)$.

(V) The sampling procedure is repeated until the (k, k) is selected from the k^2th (last) pool.

By following these steps, k^2 pairs are selected for full measurement among the k^4 pairs in one cycle. It is possible to measure mk^2 pairs by repeating these steps k cycles. Let us give an illustration for these steps in the following table.

1.3 Applications of Ranked Set Sampling in Environmental Researches

We briefly mentioned a few of the difficulties that make the sampling process difficult at the beginning of the introduction. Despite these difficulties, RSS is a useful sampling technique which provides efficient estimators for parameters of environmental variables. In this section, we present some studies that have been given environmental data applications.

Halls & Dell (1966) used the RSS procedure to select a sample from a pine-hardwood forest. They observed that RSS provides considerably more efficient estimator than SRS at estimating the weights of browse and herbage. Due to its ease of collection, Cobby et al. (1985) recommended using the RSS process to increase the precision of estimates when sampling grass and grass clover swards. Martin et al. (1980) considered three different sampling procedure which are balanced RSS, unbalanced RSS and SRS in estimating the value of a heathland located in a forest. Lacayo et al. (2002) studied proportion estimator based on RSS and gave an example of determining the plutonium content of soil samples for the application of the proposed proportion estimator. They highlighted the existence of a robust correlation between soil levels of americium and plutonium concentrations. Additionally, they emphasized that measuring americium levels using a hand-held instrument is significantly more cost-effective compared to conventional plutonium soil analysis methods. Thus, plutonium concentration can be used as an auxiliary variable to rank the americium concentration in the RSS procedure. Barnett (1999) investigated two different positive skewed distributions which are lognormal and extreme value. The author pointed out the widespread application of positive skewed distributions in modelling environmental data as a reason for using these distributions. In this study, the author looked at the mean estimator and discussed the effectiveness of these estimators under skewed distributions. Murray et al. (2000) conducted a comprehensive investigation into the precision of the mean estimator based on RSS, considering both perfect and imperfect ranking scenarios across a range of univariate distributions. The authors also examined the total amount of deposit on the upper surface of the leaf, along with the mean percentage of the upper leaf surface covered by deposit. In order to estimate the mean and median of the crop production data from the United States Department of Agriculture, Husby et al. (2006) proposed estimators based on RSS. With and without replacement policies of RSS were examined by Deshpande et al. (2006). These sampling policies were used to obtain empirical data from the different finite populations which consists of measurements of height and diameter for long-leafed pines. Wolfe (2012) used the RSS process to collect a representative sample from gasoline in order to lower cost and improve the sampling process for further analysis. Yildiz & Sevil (2019) studied empirical distribution functions (EDFs) based on three different sampling designs in RSS. These EDFs have been demonstrated to be superior to the EDF based on SRS for a variety of symmetric and skewed distributions. These EDF estimators based on sampling designs were applied to air quality data for illustrative purposes. Younes (2020) developed a new decision-making process using median RSS. This process assisted Jordan in improving and marketing its decentralized wastewater treatment policy.

In these studies, it is supposed that environmental observations are independent. However, spatial structure could be seen among the observations. For this case, new sampling techniques have been developed by combining ranked set and geostatistical approaches. Robertson et al. (2021) introduced quasi-random RSS (QRSS). The authors obtained that QRSS with $k = 3$ is more efficient than RSS with $k = 20$. Also, the authors showed that QRSS performed well in comparison to balanced acceptance sampling, a geostatistical sampling design. Robertson et al. (2022) combined a spatially balanced sample with local ranking information. Also, the authors provided a modified QRSS design. Horvitz-Thompson estimator for total parameter was investigated for various spatial structures. The study demonstrated that local ranking is an useful strategy and attractive alternative to QRSS.

1.4 Literature Review on RSS-based Dependence Parameter Estimation

Some authors addressed the estimation of the correlation parameter using RSS and its modifications. Supposed that (X, Y) is a bivariate random vector with PDF $f(x, y)$ and cumulative distribution function (CDF) $F(x, y)$. The ML estimator of the dependence parameter is examined under the following three scenarios: (i) the parameters of the variables X and Y are known, (ii) the parameters of the variable X (which is more easily accessible) are known, but those of Y are not, and (iii) all parameters are unknown. In each of the three scenarios, Stokes (1980) looked into the ML estimator for dependence parameter of bivariate normal distribution using RSS. Stokes (1980) demonstrated that the ML estimator of dependence parameter using RSS in cases (i) is asymptotically as efficient as the ML estimator based on SRS. Therefore, Stokes (1980) showed that using MRSS, which only needs the maximum or minimum ranked units, can increase efficiency in three different cases. Modarres & Zheng (2004) examined fisher information (FI) about the ML estimator of dependence parameter. They considered the case (iii) and provided an explicit form of FI matrix when (X, Y) follows bivariate normal and extreme value distributions. According to their findings, it is seen that the ML estimator using RSS for the dependence parameter is as efficient as the ML estimator using SRS under cases (i) and (ii); however, it is more efficient than the ML estimator based on SRS under case (iii). In addition to FI, they obtained bootstrap and asymptotic confidence intervals (CIs) for the dependence parameter under the cases (i) and (ii). They established that the asymptotic CI provides better coverage probabilities for the dependence parameter, whereas all coverage probabilities of bootstrap CIs are 1s. Al-Saleh & Samawi (2005) considered BVRSS-based estimation of the dependence parameter of bivariate normal distribution under the cases (i) and (iii). The authors provided a non-parametric estimator based on BVRSS for the dependence parameter. Moreover, modified ML estimator was studied, and an example was given for bivariate normal distribution. Al-Saleh & Samawi (2005)'s research indicates that these estimators are unbiased and have less variance than their counterparts in bivariate SRS design. Zheng & Modarres (2006) proposed a robust correlation coefficient estimator based on RSS. They obtained the estimator by using the modified ML method. This estimator is robust to common ranking errors. They demonstrated that the suggested estimator is at least as efficient as the ML estimators based on SRS and RSS under the case (iii). They investigate the performance of the proposed estimator under imperfect and perfect ranking. Hanandeh & Al-Saleh (2013) looked at moving extreme RSS (MERSS) to develop estimators for the dependence parameter of Downton's Bivariate Exponential Distribution using under the cases (i) and (iii). The authors established that MERSS offers an unbiased estimator for the dependence parameter in case (i), but not in case (iii). Additionally, the case (iii) is seen to have a negative effect on the relative efficiencies.

Recently, the estimation of the dependence parameter for the Gumbel's (Type I) bivariate exponential distribution was taken into account by Sevil & Yildiz (2022b). Also, Sevil & Yildiz (2023a) dealt with estimating the dependence parameter of Farlie-Gumbel-Morgenstern (FGM) type bivariate gamma distribution. In subsequent chapters, we will discuss the works in more detail.

1.5 Motivation and Outline of the Dissertation

In the environmental studies that should be investigated by taking into account the correlation between random variables (X and Y), such as water storage depth and water translation time, various groundwater quality measures pH, nitrate, sulfate, and chloride), flood volume and its duration, flood peak and its volume, and so on. It is evident from the studies cited in the previous section that the bivariate normal distribution was extensively studied. Therefore, some studies have recommended looking into other bivariate distributions, particularly those where the two random variables may have a conditionally nonlinear relationship, for example Stokes (1980) and Al-Saleh & Samawi (2005). Because environmental quantities frequently have a skewed distribution with a positive skewness rather than a normal distribution. Let us provide some motivating examples.

Example 3 (Moore & Clarke (1981)) *A problem with rainfall runoff modelling was discussed. They defined two variables called the storage depth s and the translation time t in the section titled "A Bivariate Exponential Storage-Translation Model". In this section, they stated "a basin with thin soils in the higher altitude areas that are furthest from the basin outfall is likely to have s and t negatively correlated". They established that the Gumbel's (Type I) bivariate exponential distribution is the proper distribution to use when modelling rainfall data.*

Example 4 (Bárdossy (2006)) *The quality of groundwater was examined using four different parameters: pH, nitrate, sulfate, and chloride. This paper was presented two theoretical copula-based models: Gaussian and non-Gaussian. Because of its asymmetrical dependence, which is better suited to describing the spatial dependence of groundwater quality parameters, this paper recommended the non-Gaussian copula. Also, Bárdossy (2006) showed that the suggested copula-based models can be applied with ease for geostatistical simulation.*

Example 5 (Balakrishnan & Lai (2009)) *Multivariate hydrological phenomena, such as floods and storms, can be effectively modeled using a bivariate gamma distribution, where the marginal distributions exhibit distinct scale and shape parameters. Yue et al. (2001) examined bivariate gamma models to capture hydrological events such as flood peaks and volumes, as well as flood volumes and durations. A similar experiment was also conducted by Long & Krzysztofowicz (1992).*

However only a few studies have taken into account the estimation of the dependence parameter in the scenario where the random variables have a conditionally nonlinear relationship. In the dissertation, we investigate ML estimators under Gumbel's (Type I) bivariate exponential and FGM type bivariate gamma distributions.

The ranked-based sampling designs are presented in Section 1.2 under the assumption that the Xs are perfectly ranked. Many authors in the literature offered

ML estimators based on RSS and its modifications under the assumption. However, perfect ranking assumption is not possible in practice. As a solution to the problem, the dissertation provides a robust estimate based on RSS for correlation coefficient of bivariate normal distribution.

The remaining of the dissertation is organized as follows. In the Chapter 2, we present some approaches and algorithms to generate rank-based sampling designs from a bivariate PDF. We consider the dependence parameters of Gumbel's (Type I) bivariate exponential and FGM type bivariate gamma distributions. ML estimators based RSS, $GMRSS(R = r)$ and ERSS are examined under perfect ranking. In Gumbel's (Type I) bivariate exponential distribution, likelihood ratio test (LRT) statistic based on RSS, $GMRSS(R = r)$ and ERSS are studied for testing $H_0 : \theta = 0$. An extensive Monte-Carlo simulation results are presented. In Chapter 3, we take into account the imperfect ranking case. We investigate maximum pseudo-likelihood (MPL) estimate based on RSS for the correlation coefficient of bivariate normal distribution. A Monte-Carlo simulation is performed for making comparison between MPL estimator based on RSS and its counterparts in SRS and RSS. In Chapter 4, we define some concomitant based non-parametric bootstrap techniques for CI of correlation coefficient. Finally, general conclusions and final remarks are given in Chapter 5.

CHAPTER TWO
STATISTICAL INFERENCE ON DEPENDENCE PARAMETER UNDER PERFECT RANKING CASE

Let X and Y are correlated continuous random variables with the CDF $F(x, y)$ and PDF $f(x, y)$. The CDFs and PDFs of marginals are notated by $H(x)$, $h(x)$, $G(y)$ and $g(y)$. One of the random variables, let's say X, is regarded as ranking criterion. In order to assign ranks to Ys, ranking information from Xs are used. In this procedure, Xs are deemed to be ranked perfectly. However, there may be an error in the ranking of Ys since ranking quality depends on the magnitude of the correlation between X and Y that is denoted by $Cor(X, Y)$. If $Cor(X, Y) = \pm 1$, then Ys are perfectly ranked; otherwise, they are not. On the other hand, $Cor(X, Y) = \pm 1$ is unrealistic. Thus, it is assumed that $X_{(r)}$ is the rth order statistic and $Y_{[r]}$ is its concomitant.

David (1973) coined the phrase "concomitant of order statistics". The asymptotic distribution of the concomitant was examined by David & Galambos (1974) under the assumption that (X, Y) follows bivariate normal distribution. Balasubramanian & Beg (1998) provided the properties of $Y_{[r]}$ for Gumbel's (Type I) bivariate exponential distribution and recurrence relations between moments of concomitants.Abo-Eleneen & Nagaraja (2002) studied FI in a single order statistic and its concomitant for FGM family. Also, they gave an application using Gumbel's (Type II) bivariate exponential distribution. Nagaraja (2003) discussed the fundamentals of the distribution theory as well as methods for simulating functions of the concomitants. Joint distribution of Y-order statistic and Y-concomitant of X-order statistic was developed by He (2007) and He & Nagaraja (2009). Moreover, He (2007) investigated FI matrix for Type II right censored samples that is selected from bivariate normal, Downton's bivariate exponential and Gumbel's (Type II) bivariate exponential distributions. Wang (2008) examined the distribution of concomitants of order statistics under the assumption is that X is an equally correlated multivariate normal sample. The distribution of $Y_{[r]}$ was considered by Bairamov & Bekci (2012) for FGM family with uniform marginals. From Downton, Marshall-Olkin, Gumbel's and FGM bivariate exponential distributions, He et al. (2013) suggested algorithms for generating order statistics and their concomitants. Their algorithms, however, are not useful for the simulation of $RSS(k,m)$. In the dissertation, we investigate this issue and develop various algorithms. For FGM bivariate Lomax distribution, Philip & Thomas (2017) discussed the estimation of the parameters associated with the distribution of the concomitant variable based on RSS.

Suppose that $(X_{(r)j}, Y_{[r]j})$ is $RSS(k,m)$ where $r = 1, \cdots, k$ and $j = 1, \cdots, m$. Thus, there are m pairs for a fixed rank r. In the pairs, X has the same PDF of rth order statistic which is given by Eq. (1.3) and is denoted by $h_{r:k}(x)$. The joint PDF of the pairs $(X_{(r)j}, Y_{[r]j})$ can be written as following:

$$\begin{aligned} f_{r:k}(x,y) &= h_{r:k}(x) f(y|x) \\ &= \frac{k!}{(r-1)!(k-r)!} H^{r-1}(x) \left[1 - H(x)\right]^{k-r} f(x,y). \end{aligned} \tag{2.1}$$

On the other hand, the PDF of Y is expressed as follows:

$$\begin{aligned} g_{[r:k]}(y) &= \int_{-\infty}^{\infty} f_{r:k}(x,y)\, dx \\ &= \frac{k!}{(r-1)!(k-r)!} \int_{-\infty}^{\infty} H^{r-1}(x) \left[1 - H(x)\right]^{k-r} f(x,y) dx. \end{aligned} \tag{2.2}$$

If (X,Y) follows bivariate normal distribution with the parameters μ_x, μ_y, σ_x^2, σ_y^2 and ρ, Y_i can be expressed in the following form,

$$Y_i = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (X_i - \mu_x) + \epsilon_i, \tag{2.3}$$

where X_i and ϵ_i are mutually independent for $i = 1, \cdots, n$. Here, ϵ_i has mean $E[\epsilon_i] = 0$ and variance $V[\epsilon_i] = \sigma_y^2 (1 - \rho^2)$. Similar to Eq. (2.3), the form of $Y_{[r]}$ is given by

$$Y_{[r]} = \mu_y + \rho \frac{\sigma_y}{\sigma_x} \left(X_{(r)} - \mu_x\right) + \epsilon_{[r]}, \tag{2.4}$$

where $\epsilon_{[r]}$ is independent of the $X_{(r)}$ for $r = 1, \cdots, k$. Letting,

$$\alpha_r = E\left[\frac{X_{(r)} - \mu_x}{\sigma_x}\right], \quad \text{and} \quad \beta_{rs} = Cov\left[\frac{X_{(r)} - \mu_x}{\sigma_x}, \frac{X_{(s)} - \mu_x}{\sigma_x}\right] \tag{2.5}$$

the following formulas can be found, e.g. in David & Nagaraja (2004) (formulas (6.8.3a-d) p. 145),

$$\begin{aligned} E\left[Y_{[r]}\right] &= \mu_y + \rho\sigma_y\alpha_r \\ V\left[Y_{[r]}\right] &= \sigma_y^2\left(\rho^2\beta_{rr} + 1 - \rho^2\right) \\ Cov\left[X_{(r)}, Y_{[s]}\right] &= \rho\sigma_x\sigma_y\beta_{rs} \\ Cov\left[Y_{[r]}, Y_{[s]}\right] &= \rho^2\sigma_y^2\beta_{rs} \;\; r \neq s. \end{aligned} \tag{2.6}$$

In the rest of the chapter, we will deal with estimating the dependence parameter of the Gumbel's (Type I) bivariate exponential and Farlie-Gumbel-Morgenstern type bivariate gamma distribution using RSS and its modifieds.

2.1 Gumbel's (Type I) Bivariate Exponential Distribution

RSS, $GMRSS(R = r)$, and ERSS are the three sampling techniques that we take into consideration in this section where $r \in [1, k]$. This section first provides ranked-based algorithms to generate data from Gumbel's (Type I) bivariate exponential distribution. By using ML method, this section enhances estimators based on RSS, $GMRSS(R = r)$, and ERSS for the association parameter of Gumbel's (Type I) bivariate exponential distribution. Also, we examine LRTs based on RSS, $GMRSS(R = r)$, and ERSS to determine if there is a statistically significant association between the random variables.

2.1.1 Preliminaries

Bivariate exponential distributions are frequently used in reliability and survival analysis, Balakrishnan & Lai (2009). Numerous bivariate exponential distributions have been proposed in the literature to model the aging and failure processes of two components. Three different bivariate exponential distributions known as Type I, Type II and Type III were proposed by Gumbel (1960).

Suppose that (X, Y) be a bivariate random vector with absolutely continuous joint CDF $F(x, y)$,

$$F(x,y) = 1 - \exp\{-x\} - \exp\{-y\} + \exp\{-x - y - \theta xy\}, \quad 0 < x, y < \infty, \tag{2.7}$$

and the joint PDF $f(x, y)$

$$f(x,y) = [(1+\theta x)(1+\theta y) - \theta] \exp\{-x - y - \theta xy\}, \quad 0 < x, y < \infty \tag{2.8}$$

where marginals follow standard exponential distributions and $0 \leq \theta \leq 1$. Figure 2.1 presents illustrations for the Eqs. (2.7) and (2.8).

The joint survival function is given by

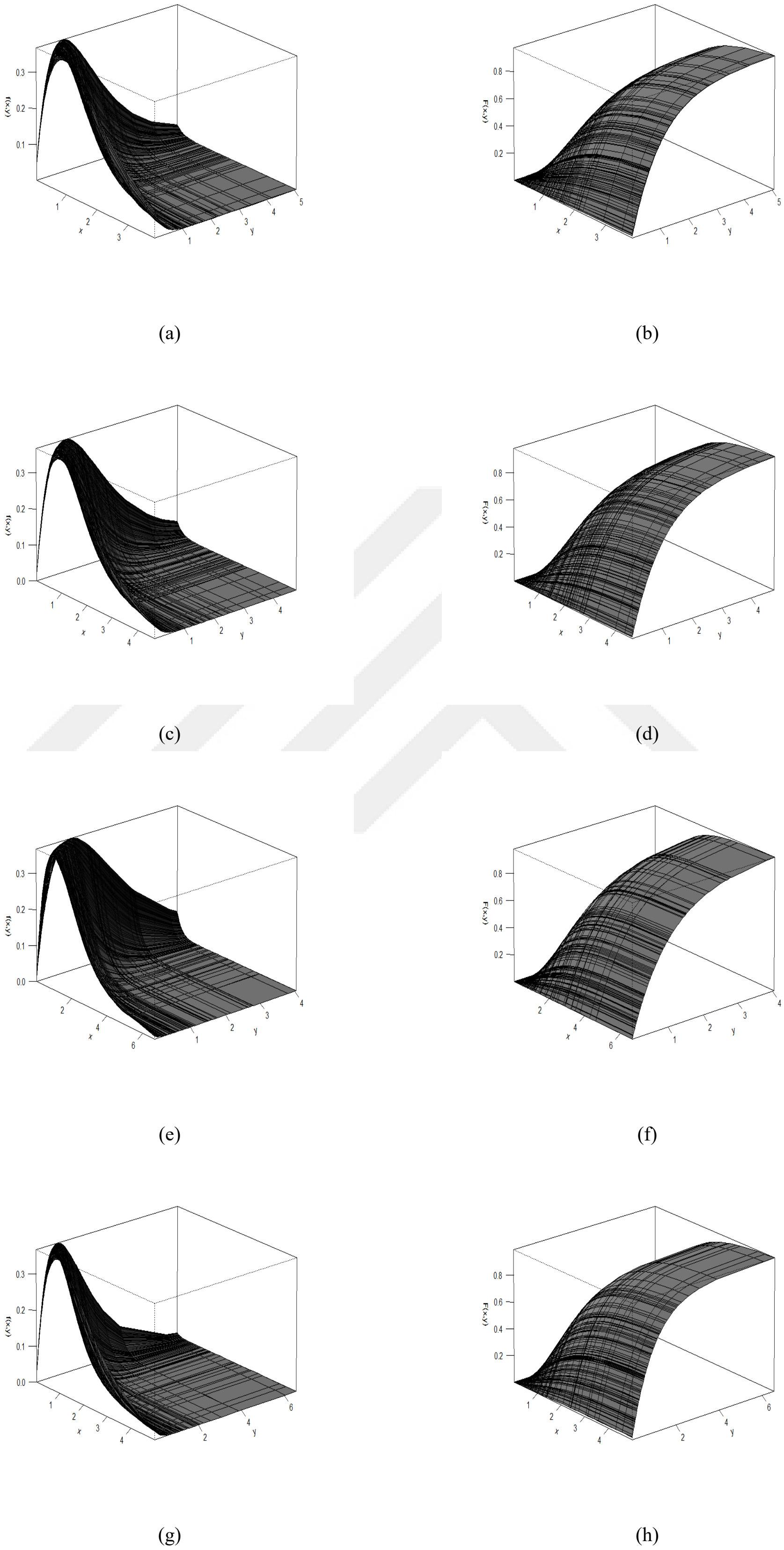
$$\bar{F}(x,y) = \exp\{-x - y - \theta xy\}, \quad 0 < x, y < \infty \tag{2.9}$$

The conditional PDF of Y given $X = x$ is

$$\begin{aligned} f(y|x) &= \frac{f(x,y)}{f(x)} \\ &= [(1+\theta x)(1+\theta y) - \theta] \exp\{-y - \theta xy\} \end{aligned} \tag{2.10}$$

The conditional mean and variance of Y given $X = x$ are expressed as follows:

$$E[Y|X=x] = \frac{1+\theta+\theta x}{(1+\theta x)^2}, \tag{2.11}$$



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 2.1 Gumbel's (Type I) bivariate exponential distribution; (a): Joint PDF for $\theta = 0$, (b): Joint CDF for $\theta = 0$, (c): Joint PDF for $\theta = 0.4$, (d): Joint CDF for $\theta = 0.4$, (c): Joint PDF for $\theta = 0.6$, (d): Joint CDF for $\theta = 0.6$, (c): Joint PDF for $\theta = 1$, and (d): Joint CDF for $\theta = 1$

and

$$V[Y|X=x] = \frac{(1+\theta+\theta x)^2 - 2\theta^2}{(1+\theta x)^4}. \tag{2.12}$$

The following equation gives the joint expectation of XY,

$$E[XY] = \frac{1}{\theta} e^{1/\theta} E_1\left(\frac{1}{\theta}\right) \tag{2.13}$$

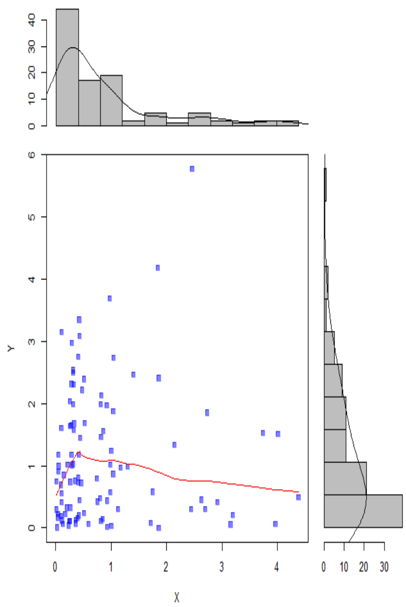
where

$$E_i(\omega) = \int_1^\infty \xi^{-i} \exp\{-\omega\xi\} \, d\xi \tag{2.14}$$

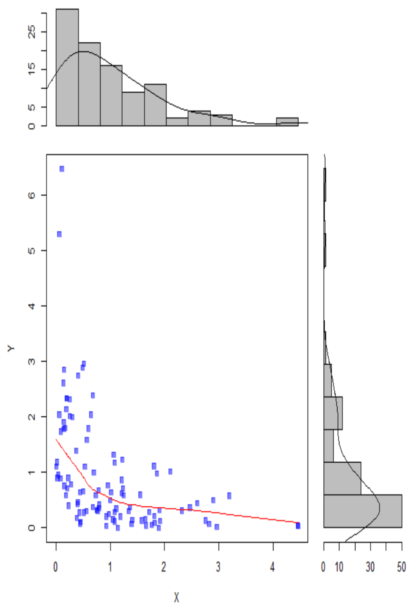
Thus, the correlation coefficient between X and Y is,

$$\begin{aligned} Cor(x,y) = \rho &= \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}} \\ &= \frac{1}{\theta} e^{\theta^{-1}} E_1\left(\frac{1}{\theta}\right) - 1 \end{aligned} \tag{2.15}$$

where ρ ranges from 0 when $\theta = 0$ to -0.40365 when $\theta = 1$. The correlation coefficient between the variables is drawn and given by Figure 2.2.



(a)



(b)

Figure 2.2 Correlation between X and Y; (a): $\theta = 0$ and (b): $\theta = 1$

Barnett (1985) investigated estimating the association parameter, θ, based on SRS for Gumbel's (Type I) bivariate exponential distribution. Balasubramanian & Beg (1998) provided the properties of $Y_{[r]}$ for $r = 1, \cdots, k$. In this study, the PDF of $Y_{[1]}$ was obtained using Eq. (2.2).

$$\begin{aligned} g_{[1:k]}(y) &= \int_0^\infty \exp\{-y(1+\theta x)\}\left[(1+\theta x)(1+\theta y)-\theta\right] n \exp\{-nx\}\, dx \\ &= n \exp\{-y\} \int_0^\infty \exp\{-(\theta y + n)x\}\left[\theta x(1+\theta x)+(1+\theta y)-\theta\right] dx \\ &= \left[n - \frac{n(n-1)}{\theta(y+n/\theta)} - \frac{n(n-1)}{\theta(y+n/\theta)^2}\right] \exp\{-y\}, \quad y \geq 0. \end{aligned} \tag{2.16}$$

Then, the PDF of $Y_{[r]}$ was defined by using the connections of PDFs of order statistics,

$$\begin{aligned} g_{[r:k]}(y) &= \sum_{i=k-r+1}^{k} (-1)^{i-k+r-1} \binom{i-1}{k-r} \binom{k}{i} g_{[1:i]}(y) \\ &= \sum_{i=k-r+1}^{k} (-1)^{i-k+r-1} \binom{i-1}{k-r} \binom{k}{i} \\ &\qquad \left[i - \frac{i(i-1)}{\theta(y+i/\theta)} - \frac{i(i-1)}{\theta(y+i/\theta)^2}\right] \exp\{-y\} \end{aligned} \tag{2.17}$$

2.1.2 Algorithms to Generate Ranked-based Samples

To generate dependent random variables (X, Y) from Gumbel's (type I) bivariate exponential distribution, we consider the conditional distribution approach.

Approach I (Conditional Distribution Approach): This notion is frequently credited to Rosenblatt (1952). In this approach, X is first simulated from $f(x)$, and Y is subsequently simulated from the conditional PDF $f(y|x)$. Which one should be the initial variable, and which one should be obtained conditionally on the initial variable are possible questions. Rubinstein (1981) straightforwardly acknowledged that there is no known method for determining the optimal order of variable representation in a vector to minimize Central Process Unit (CPU) time. See Balakrishnan & Lai (2009) and Devroye (1986) for examples of how this approach has been used with various bivariate probability distributions. The conditional distribution in Eq. (2.10) can be expressed as follows after some algebraic manipulation:

$$\begin{aligned} f\left(y|x\right) &= \left(\theta\left(1+\theta x\right)y+1+\theta x-\theta\right)\exp\left\{-\left(1+\theta x\right)y\right\} \\ &= \frac{\theta}{1+\theta x}\left(\left(1+\theta x\right)^{2}y\exp\left\{-\left(1+\theta x\right)y\right\}\right) \\ &+ \frac{1+\theta x-\theta}{1+\theta x}\left(\left(1+\theta x\right)\exp\left\{-\left(1+\theta x\right)y\right\}\right) \end{aligned} \tag{2.18}$$

The first part of the Eq. (2.18) is $gamma\left(2,\left(1+\theta x\right)\right)$ and the second part is $Exp\left(1+\theta x\right)$. Based on the Eq. (2.18), Devroye (1986) defined the following algorithm. We define an R function which is named as *BiExp(n,theta)* where n is the

Algorithm 1: Generating data from Gumbel's (type I) BED

Step I: Generate $X, T \sim Exp(1)$;
Step II: Generate $U \sim Unif(0,1)$;
Step III: IF $U \leq \frac{\theta}{1+\theta X}$, THEN;
  Generate $E \sim Exp(1)$;
  $T \leftarrow T+E$;
Step IV: RETURN (X,$Y=\frac{T}{1+\theta X}$). THEN;

number of observations and theta controls the magnitude of the dependence between random variables X and Y.

To generate complete and censored samples including order statistic and its concomitant, there are some approaches which are investigated by He et al. (2013). However, since different stochastic mechanisms are used in RSS, these approaches are not applicable to generate RSS data from bivariate distributions. The ranked-based algorithms that we suggest in this section use the subsequent approach.

Approach II (Ranked Sets): In this approach, the rth set of size k is generated from the CDF $F(x,y)$ for $r=1\cdots,k$. Following that, the set is ranked in accordance with its X-values. From the rth set, the pairs $(X_{(r)j}, Y_{[r]j})$ are identified for $r=1,\cdots,k$. The other pairs are used in the ranking process and then are discarded. This approach is simple to use with various bivariate probability distributions. On the other hand, we can say that time lost in ranking is this approach's disadvantage. Parallel processing and vectorized simulation can, however, overcome it. Also, we note that Y-values can be selected as ranking criteria instead of X-values. This case has no impact on CPU usage or outputs. Based on this approach, we define the *Algorithm 2*. We

Algorithm 2: Generating the $RSS(k, m)$ from Gumbel's (type I) BED
Step I: Generate the pairs (X_i, Y_i) by using *Algorithm 1* where $i = 1, \cdots, k$;
Step II: Rank the pairs according to their X-values;
Step III: Select $X_{(r)j}$ and its concomitant $Y_{[r]j}$;
Step IV: Repeat (1)-(3), $r = 1, \cdots, k$ and $j = 1, \cdots, m$;
Step V: RETURN ($X_{(r)j}$,$Y_{[r]j}$).

provide an R function *RSS(k,m,theta)* where k is the set size, m is the number of cycle and theta is the value of the dependence parameter θ.

The another ranked-based algorithm is given to generate $GMRSS(R = r)$ from the Gumbel's (type I) bivariate exponential distribution. Based on the following approach, *Algorithm 3* is built.

Approach III (Basic Approach): This is a basic simulation approach for generating order statistics, see He et al. (2013). First, n sets are constructed from $F(x, y)$. After ranking the sets according to their X-values, the rth ranked pairs $(X_{(r)}, Y_{[r]})$ are selected and the others are discarded. In the *Algorithm 3*, we have to choose a rank value $r \in [1, k]$. Then, the algorithm provides pairs ($X_{\tau(r)j}$,$Y_{\tau[r]j}$) from $f_{r:k}(x, y)$. For

Algorithm 3: Generating the $GRSS(R = r)$) from Gumbel's (type I) BED
Step I: Generate the pairs (X_i, Y_i) by using *Algorithm 1* where $i = 1, \cdots, k$;
Step II: Rank the pairs according to their X-values;
Step III: Select $X_{\tau(r)j}$ and its concomitant $Y_{\tau[r]j}$;
Step IV: Repeat (1)-(3), $\tau = 1, \cdots, k$ and $j = 1, \cdots, m$;
Step V: RETURN ($X_{\tau(r)j}$,$Y_{\tau[r]j}$).

the *Algorithm 3*, we use the R function $GMRSS(k, m, R, theta)$ where k is the set size, m is the number of cycle, R is the rank value r and theta is the value of the dependence parameter.

We take into account that $n/2$ pairs are selected for each minimum and maximum as we investigated the effects of $f_{1:k}(x, y)$ and $f_{k:k}(x, y)$ in combination on ML estimator for the dependence parameter. This sampling procedure is known as ERSS in case of even set size which is developed by Samawi et al. (1996). The following approach is given for generating ERSS.

Approach IV (Extreme Ranks): The n sets of size k are generated from $F(x, y)$. After ranking the sets by using their X-values, the minimum ranked pairs $(X_{r(1)j}$,$Y_{r[1]j})$ from the $n/2$ sets and the maximum ranked pairs $(X_{r(k)j}$,$Y_{r[k]j})$ are selected from the others. To achieve this, at least one of the set size k and the number of cycles m have to be even. For example, $k/2$ pairs are selected for each extreme rank when the set size is even, and the number of cycles is odd. On the other hand, minimum ranked pairs are selected in the first $m/2$ cycles while maximum ranked pairs are selected from all sets in the remaining cycles. To obtain the $ERSS(k, m)$, we follow the steps in *Algorithm 4*. Note that the *Algorithm 4* is used under the case where the set size is even. With

Algorithm 4: Generating $ERSS(k, m)$ from Gumbel's (type I) BED

Step I: Generate the pairs (X_i, Y_i) by using *Algorithm 1* where $i = 1, \cdots, k$;
Step II: Rank the pairs according to their X-values;
Step III: Select $X_{\tau(1)j}$ and its concomitant $Y_{\tau[1]j}$ for $1 \leq \tau \leq k/2$;
Step IV: Select $X_{\tau(k)j}$ and its concomitant $Y_{\tau[k]j}$ for $k/2 < \tau \leq k$;
Step V: Repeat (1)-(3), $\tau = 1, \cdots, k$ and $j = 1, \cdots, m$;
Step VI: RETURN $(X_{\tau(r)j}$,$Y_{\tau[r]j})$ for $r \in \{1, k\}$.

a small adjustment to the algorithm, it can be used when the set size is odd. Also, we construct an R function *ERSS(k,m,theta)* where k is the set size, m is the number of cycle and theta is the value of the dependence parameter.

These R functions mentioned in this section can are available at https://github.com/YCS92/GenerateRankedBasedSamples.

2.1.3 Maximum Likelihood Estimates

2.1.3.1 Simple Random Sampling

Let (X_i, Y_i) be a simple random sample which is generated using *Algorithm 1* (or is selected from a population). Here, the issue is estimating the correlation coefficient which is presented by Eq. (2.15). For this issue, the following estimator can be used,

$$\hat{\rho} = \frac{1}{\hat{\theta}} \exp(\hat{\theta}^{-1}) E_1 \left(\frac{1}{\hat{\theta}} \right) - 1. \tag{2.19}$$

This issue establishes into the problem of estimating θ, as evident in Eq. (2.19). Barnett (1985) introduced the ML and method of moment estimators of θ. Barnett (1985) demonstrated in this study that the ML estimator based on SRS is more efficient than the method of moment estimator based on SRS. In this section, we discuss the ML estimator based on SRS which is denoted by $\hat{\theta}_{SRS}$. Let us define the likelihood function $\mathscr{L}_{SRS} = \prod_{i=1}^{n} f(x_i, y_i)$. Then, log-likelihood function is given by

$$L_{SRS}(\theta) = \sum_{i=1}^{n} \ln \left[(1 + \theta x_i)(1 + \theta y_i) - \theta \right] - \sum_{i=1}^{n} (x_i + y_i + \theta x_i y_i). \tag{2.20}$$

The first derivative of the $L_{SRS}(\theta)$ with respect to θ is

$$\begin{aligned} \frac{\partial L_{SRS}(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{x_i(1+\theta y_i) + y_i(1+\theta x_i) - 1}{(1+\theta x_i)(1+\theta y_i) - \theta} - \sum_{i=1}^{n} x_i y_i = 0 \\ \sum_{i=1}^{n} \frac{x_i(1+\theta y_i) + y_i(1+\theta x_i) - 1}{(1+\theta x_i)(1+\theta y_i) - \theta} = \sum_{i=1}^{n} x_i y_i \end{aligned} \tag{2.21}$$

Barnett (1985) obtained the approximate value of $\hat{\theta}_{SRS}$ by using a computer algorithm.

2.1.3.2 Ranked Set Sampling

Assume that $RSS(k, m)$ is generated (or is selected from a population) by using *Algorithm 2*. The rth ranked pairs $\left(X_{(r)j}, Y_{[r]j}\right)$ in the sample $RSS(k, m)$ follows the

PDF, $f_{r:k}(x,y)$, which is given by Eq. (2.1) where $r=1,\cdots,k$ and $j=1,\cdots,m$. The likelihood function is $\mathscr{L}_{RSS}=\prod_{j=1}^{m}\prod_{r=1}^{k}f_{r:k}\left(x_{(r)j},y_{[r]j}\right)$. The log-likelihood is presented by the following equation.

$$
\begin{aligned}
L_{RSS}(\theta) &= \sum_{j=1}^{m}\sum_{r=1}^{k}\ln\left(f_{(r:k)}\left(x_{(r)j},y_{[r]j}\right)\right) \\
&= \sum_{j=1}^{m}\sum_{r=1}^{k}\ln\left(\frac{k!}{(r-1)!\,(k-r)!}\right) \\
&+ \sum_{j=1}^{m}\sum_{r=1}^{k}(r-1)\ln\left(F\left(x_{(r)j}\right)\right) \\
&+ \sum_{j=1}^{m}\sum_{r=1}^{k}(k-r)\ln\left(1-F\left(x_{(r)j}\right)\right) \\
&+ \sum_{j=1}^{m}\sum_{r=1}^{k}\ln\left(f\left(x_{(r)j},y_{[r]j}\right)\right).
\end{aligned}
\tag{2.22}
$$

Given that θ appears in $f(x,y)$ of $f_{(r:k)}(x,y)$, the first derivative of the last term of $L_{RSS}(\theta)$ with respect to θ yields the subsequent equation.

$$
\begin{aligned}
\frac{\partial L_{RSS}(\theta)}{\partial\theta} = \sum_{j=1}^{m}\sum_{r=1}^{k}&\frac{y_{[r]j}\left(1+\theta y_{r(r)j}\right)+x_{(r)j}\left(1+\theta y_{[r]j}\right)-1}{\left(1+\theta y_{[r]j}\right)\left(1+\theta x_{(r)j}\right)-\theta} \\
&-\sum_{j=1}^{m}\sum_{r=1}^{k}x_{(r)j}y_{[r]j}=0
\end{aligned}
\tag{2.23}
$$

It is possible to determine the approximate value of the ML estimator based on RSS ($\hat{\theta}_{RSS}$) by solving the one of the Eqs. (2.22) and (2.23).

2.1.3.3 Generalized Modified Ranked Set Sampling

Now, we suppose that $GMRSS(R=r)$ is generated (or is selected from a population) by using the *Algorithm 3*. The pairs $\left(X_{\tau(r)j},Y_{\tau[r]j}\right)$ in $GMRSS(R=r)$ have the PDF, $f_{r:k}(x,y)$, which is provided by Eq. (2.1). The likelihood function is

$\mathscr{L}_{GMRSS(R=r)} = \prod_{j=1}^{m}\prod_{\tau=1}^{k} f_{r:k}\left(x_{\tau(r)j}, y_{\tau[r]j}\right)$ and the log-likelihood is

$$\begin{aligned} L_{GMRSS(R=r)}\left(\theta\right) &= \sum_{j=1}^{m}\sum_{\tau=1}^{k} \ln\left(f_{(r:k)}\left(x_{\tau(r)j}, y_{\tau[r]j}\right)\right) \\ &= mk\ln\left(\frac{k!}{(r-1)!\,(k-r)!}\right) \\ &+ \sum_{j=1}^{m}\sum_{\tau=1}^{k}(r-1)\ln\left(F\left(x_{\tau(r)j}\right)\right) \\ &+ \sum_{j=1}^{m}\sum_{\tau=1}^{k}(k-r)\ln\left(1-F\left(x_{\tau(r)j}\right)\right) \\ &+ \sum_{j=1}^{m}\sum_{\tau=1}^{k}\ln\left(f\left(x_{\tau(r)j}, y_{\tau[r]j}\right)\right). \end{aligned} \tag{2.24}$$

The following equation is obtained by taking the first derivative of $L_{GMRSS(R=r)}\left(\theta\right)$ with respect to θ.

$$\begin{aligned} \frac{\partial L_{GMRSS(R=r)}(\theta)}{\partial\theta} = \sum_{j=1}^{m}\sum_{\tau=1}^{k} \frac{y_{\tau[r]j}\left(1+\theta x_{\tau(r)j}\right) + x_{\tau(r)j}\left(1+\theta y_{\tau[r]j}\right) - 1}{\left(1+\theta y_{\tau[r]j}\right)\left(1+\theta x_{\tau(r)j}\right) - \theta} \\ - \sum_{j=1}^{m}\sum_{t=1}^{k} x_{\tau(r)j} y_{r\tau[r]j} = 0, \end{aligned} \tag{2.25}$$

By solving equation (2.24) or (2.25), it is possible to calculate the approximate value of the ML estimator based on $GMRSS(R=r)$ ($\hat{\theta}_{GMRSS(R=r)}$).

2.1.3.4 Extreme Ranked Set Sampling

Let the pairs $\left(X_{\tau(r)j}, Y_{\tau[r]j}\right)$ are generated (or is selected from a population) by using *Algorithm 4* where $r \in \{1, k\}$. Recall that half of the sample is drawn from $f_{1:k}(x, y)$ and the others are drawn from $f_{k:k}(x, y)$. Thus, the likelihood function is $\mathscr{L}_{ERSS} = \prod_{j=1}^{m}\prod_{r=1}^{k/2}\left(f_{1:k}\left(x_{r(1)j}, y_{r[1]j}\right)\right)\prod_{j=1}^{m}\prod_{r=(k/2)+1}^{k}\left(f_{k:k}\left(x_{r(k)j}, y_{r[k]j}\right)\right)$. Following that, the log-likelihood function is

$$
\begin{aligned}
L_{ERSS}(\theta) &= \sum_{j=1}^{m}\sum_{r=1}^{k/2} \ln\left(f_{(1:k)}\left(x_{r(1)j}, y_{r[1]j}\right)\right) \\
&+ \sum_{j=1}^{m}\sum_{r=(k/2)+1}^{k} \ln\left(f_{(k:k)}\left(x_{r(k)j}, y_{r[k]j}\right)\right) \\
&= A + B,
\end{aligned}
\tag{2.26}
$$

where

$$
\begin{aligned}
A &= \frac{mk}{2}\ln(k) + (k-1)\sum_{j=1}^{m}\sum_{r=1}^{k/2} \ln\left(1 - F\left(x_{r(1)j}\right)\right) \\
&+ \sum_{j=1}^{m}\sum_{r=1}^{k/2} f\left(x_{r(1)j}, y_{r[1]j}\right)
\end{aligned}
\tag{2.27}
$$

and

$$
\begin{aligned}
B &= \frac{mk}{2}\ln(k) + (k-1)\sum_{j=1}^{m}\sum_{r=(k/2)+1}^{k} F\left(x_{r(k)j}\right) \\
&+ \sum_{j=1}^{m}\sum_{r=(k/2)+1}^{k} f\left(x_{r(k)j}, y_{r[k]j}\right).
\end{aligned}
\tag{2.28}
$$

A' and B' represent the first derivatives of A and B with respect to θ, respectively.

$$
\begin{aligned}
A' &= \sum_{j=1}^{m}\sum_{r=1}^{k/2} \frac{y_{r[1]j}\left(1+\theta x_{r(1)j}\right) + \left(1+\theta y_{r[1]j}\right)x_{r(1)j} - 1}{\left(1+\theta y_{r[1]j}\right)\left(1+\theta x_{r(1)j}\right) - \theta} \\
&- \sum_{j=1}^{m}\sum_{r=1}^{k/2} x_{r(1)j} y_{r[1]j},
\end{aligned}
\tag{2.29}
$$

and

$$
\begin{aligned}
B' &= \sum_{j=1}^{m}\sum_{r=(k/2)+1}^{k} \frac{y_{r[k]j}\left(1+\theta x_{r(k)j}\right) + \left(1+\theta y_{r[k]j}\right)x_{r(k)j} - 1}{\left(1+\theta y_{r[k]j}\right)\left(1+\theta x_{r(k)j}\right) - \theta} \\
&- \sum_{j=1}^{m}\sum_{r=(k/2)+1}^{k} x_{r(k)j} y_{r[k]j}.
\end{aligned}
\tag{2.30}
$$

Thus,

$$\frac{\partial L_{ERSS}(\theta)}{\partial\theta} = A' + B' = 0 \tag{2.31}$$

An approximate value of $\hat{\theta}_{ERSS}$ can be obtained by solving the Eqs. (2.18) or (2.21).

2.1.4 Simulation Results

By building a comprehensive Monte Carlo simulation, the given ML estimators from the previous sections are examined in terms of biases and relative efficiencies. Since there are no closed forms of $\hat{\theta}_{SRS}$, $\hat{\theta}_{RSS}$, $\hat{\theta}_{GMRSS(R=r)}$ and $\hat{\theta}_{ERSS}$, we use *optimize* function (Brent, 2013) in R statistical programming language. In the Monte Carlo simulation, we use this R function to determine the global maximum points of the log-likelihood functions which are given in Eqs. (2.20), (2.22), (2.24) and (2.26). Throughout the Monte Carlo simulation, $10,000$ samples are generated by using R functions in Section 2.1.2. It is supposed that the values of dependence parameter are $\boldsymbol{\theta} = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$, the set sizes are $\boldsymbol{k} = \{3, 4, 5, 6\}$ and the number of cycles are $\boldsymbol{m} = \{2, 5, 10, 15\}$. Remember that the extreme ranked set sample contains both $mk/2$ minimum and $mk/2$ maximum ranked units. Therefore, the set sizes are considered in ERSS as $\boldsymbol{k} = \{2, 4, 6, 8\}$. The bias of the ML estimators is estimated by using the following equation:

$$Bias(\hat{\theta}_{\psi}) = \frac{1}{10,000}\sum_{w=1}^{10,000}\left|\hat{\theta}_{\psi,w} - \theta\right| \tag{2.32}$$

where $\psi =$ SRS, RSS, $GMRSS(R = r)$ and ERSS. Mean square error (MSE) values of the ML estimators are obtained by following equations:

$$MSE(\hat{\theta}_{\psi}) = \frac{1}{10,000}\sum_{w=1}^{10,000}\left(\hat{\theta}_{\psi,w} - \theta\right)^2. \tag{2.33}$$

Relative efficiencies (REs) are obtained by using MSEs of the ML estimators.

$$RE(\hat{\theta}_{\psi'}, \hat{\theta}_{SRS}) = \frac{MSE(\hat{\theta}_{SRS})}{MSE(\hat{\theta}_{\psi'})} \tag{2.34}$$

where ψ' = RSS, $GMRSS(R = r)$ and ERSS.

Table 2.1 presents the bias values of $\hat{\theta}_{SRS}$ and $\hat{\theta}_{RSS}$. In Table 2.1, the ML estimators based on SRS and RSS appear to be biased, and the biases of the ML estimators appear to be decreasing as sample size $n = mk$ increases. Barnett (1985) demonstrated that the ML estimator for SRS has a substantial bias even for large sample sizes (e.g., $n = 100$). On the other hand, as noted by Barnett (1985), bias values decrease while theta values are approaching zero or one. When $\theta = 0$ or 1, $Bias(\hat{\theta}_{RSS})$ is between 0.3 and 0.055 for $n = 6$ and $n = 90$, respectively. On the other hand, $Bias(\hat{\theta}_{RSS})$ takes values in $[0.127, 0.372]$ when $\theta = 0.4$ or 0.6. Also, it is obvious to see that $Bias(\hat{\theta}_{SRS}) \approx Bias(\hat{\theta}_{RSS})$. The REs of the ML estimator based

Table 2.1 The values of $Bias(\hat{\theta}_{SRS})$ and $Bias(\hat{\theta}_{RSS})$

		$\theta = 0$		$\theta = 0.2$		$\theta = 0.4$		$\theta = 0.6$		$\theta = 0.8$		$\theta = 1$	
m	k	SRS	RSS	SRS	RSS	SRS	RSS	SRS	RSS	SRS	RSS	SRS	RSS
2	3	0.296	0.300	0.357	0.361	0.371	0.372	0.353	0.351	0.297	0.295	0.199	0.207
	4	0.253	0.260	0.325	0.323	0.348	0.344	0.333	0.330	0.280	0.279	0.182	0.185
	5	0.230	0.220	0.299	0.303	0.328	0.327	0.314	0.313	0.267	0.262	0.175	0.167
	6	0.206	0.201	0.284	0.283	0.307	0.310	0.298	0.300	0.252	0.249	0.158	0.151
5	3	0.178	0.182	0.257	0.253	0.289	0.286	0.279	0.279	0.234	0.234	0.145	0.147
	4	0.146	0.150	0.230	0.230	0.263	0.259	0.257	0.255	0.214	0.215	0.127	0.128
	5	0.131	0.126	0.209	0.205	0.238	0.236	0.238	0.235	0.200	0.204	0.114	0.114
	6	0.115	0.111	0.191	0.189	0.218	0.220	0.221	0.218	0.189	0.190	0.109	0.106
10	3	0.110	0.115	0.190	0.187	0.217	0.219	0.220	0.221	0.190	0.189	0.106	0.107
	4	0.094	0.094	0.165	0.164	0.189	0.192	0.197	0.197	0.173	0.171	0.097	0.096
	5	0.081	0.080	0.149	0.148	0.173	0.174	0.178	0.180	0.159	0.160	0.086	0.086
	6	0.071	0.069	0.137	0.133	0.156	0.156	0.167	0.167	0.152	0.149	0.077	0.079
15	3	0.089	0.086	0.157	0.158	0.182	0.183	0.188	0.189	0.167	0.165	0.090	0.086
	4	0.069	0.070	0.134	0.136	0.160	0.159	0.169	0.166	0.151	0.149	0.078	0.080
	5	0.062	0.062	0.122	0.121	0.141	0.141	0.149	0.150	0.139	0.137	0.070	0.069
	6	0.054	0.055	0.111	0.111	0.129	0.127	0.138	0.139	0.130	0.131	0.063	0.064

on RSS with respect to the ML estimator based on SRS are presented in Table 2.2. This table demonstrates that the ML estimator based on RSS is as efficient as the ML estimator based on SRS. For researchers who study RSS and its extensions, the outcome might come as a surprise. However, Stokes (1980) also found similar results, and proposed a new RSS technique known as MRSS. In all three cases (1, 2, and 3), it was demonstrated that MRSS offers more efficient ML estimator than SRS and RSS by Stokes (1980).

In addition to RSS, we examine $Bias(\hat{\theta}_{GMRSS(R=r)})$ and $RE(\hat{\theta}_{GMRSS(R=r)}, \hat{\theta}_{SRS})$

Table 2.2 The values of $RE(\hat{\theta}_{RSS}, \hat{\theta}_{SRS})$

m	k	$\theta = 0$	$\theta = 0.2$	$\theta = 0.4$	$\theta = 0.6$	$\theta = 0.8$	$\theta = 1$
2	3	0.984	0.976	0.993	1.018	1.023	0.961
	4	0.968	1.006	1.018	1.012	1.011	0.985
	5	1.070	0.976	1.004	1.011	1.043	1.073
	6	1.043	1.005	0.987	0.993	1.011	1.071
5	3	0.969	1.017	1.016	1.002	1.031	0.989
	4	0.982	0.999	1.020	1.014	0.994	0.987
	5	1.049	1.020	1.011	1.029	0.963	1.006
	6	1.088	1.021	0.981	1.026	0.988	1.040
10	3	0.946	1.017	0.977	0.997	1.015	0.996
	4	1.021	1.017	0.975	1.010	1.021	1.010
	5	1.006	0.996	0.988	0.984	0.988	0.994
	6	1.028	1.058	1.008	1.002	1.026	0.970
15	3	1.034	0.999	0.984	0.990	1.025	1.050
	4	0.985	0.979	1.010	1.036	1.014	0.970
	5	1.014	1.028	0.991	0.993	1.029	1.049
	6	1.004	0.966	1.034	0.995	0.993	0.992

for different rank values r, $r = 1, \cdots, k$. The Figures 2.3-2.10 provides the bias and RE values. The rank values (r) are seen in x-axis. According to the Figures 2.3-2.6, we can say that $\hat{\theta}_{GMRSS(R=r)}$ for $r = 1$ and $r = k$ have slightly lower biases than $\hat{\theta}_{SRS}$ and $\hat{\theta}_{RSS}$. Also, we observe that $GMRSS(R = k)$ is slightly better than $GMRSS(R = 1)$. When $\theta = 0$ or 1, $Bias(\hat{\theta}_{GMRSS(R=k)})$ is between 0.25 and ≤ 0.05 under $n = 6$ and $n = 90$, respectively. When $\theta = 0.4$ or 0.6, $Bias(\hat{\theta}_{GMRSS(R=k)})$ is in the interval $[0.125, 0.35]$. As the number of cycles increase, the bias of $\hat{\theta}_{GMRSS(R=r)}$ reduces. The estimated relative efficiencies of $\hat{\theta}_{GMRSS(R=r)}$ with respect to $\hat{\theta}_{SRS}$ are presented by the Figures 2.7-2.10. In these figures, it is obviously seen that $GMRSS(R = k)$ provides more efficient ML estimator of dependence parameter than $GMRSS(R = 1)$, SRS and RSS even if the set size is 3 and the number of cycles is 2. Also, the REs are not monotone increasing or decreasing as the number of cycles increase.

Since $GMRSS(R = 1)$ (or $GMRSS(R = k)$) ranked pairs has better performance than SRS and RSS, we aim to combine two different procedure using ERSS. We investigate ML estimator based on ERSS in terms of its bias and efficiency. Figures 2.11 and 2.12 give the biases and REs, respectively. In the Figure 2.11, it can be seen that the bias values range between 0.4 and 0.1 as the set size increases (for $m = 2$). On the other hand, the bias values range between 0.25 and 0.05

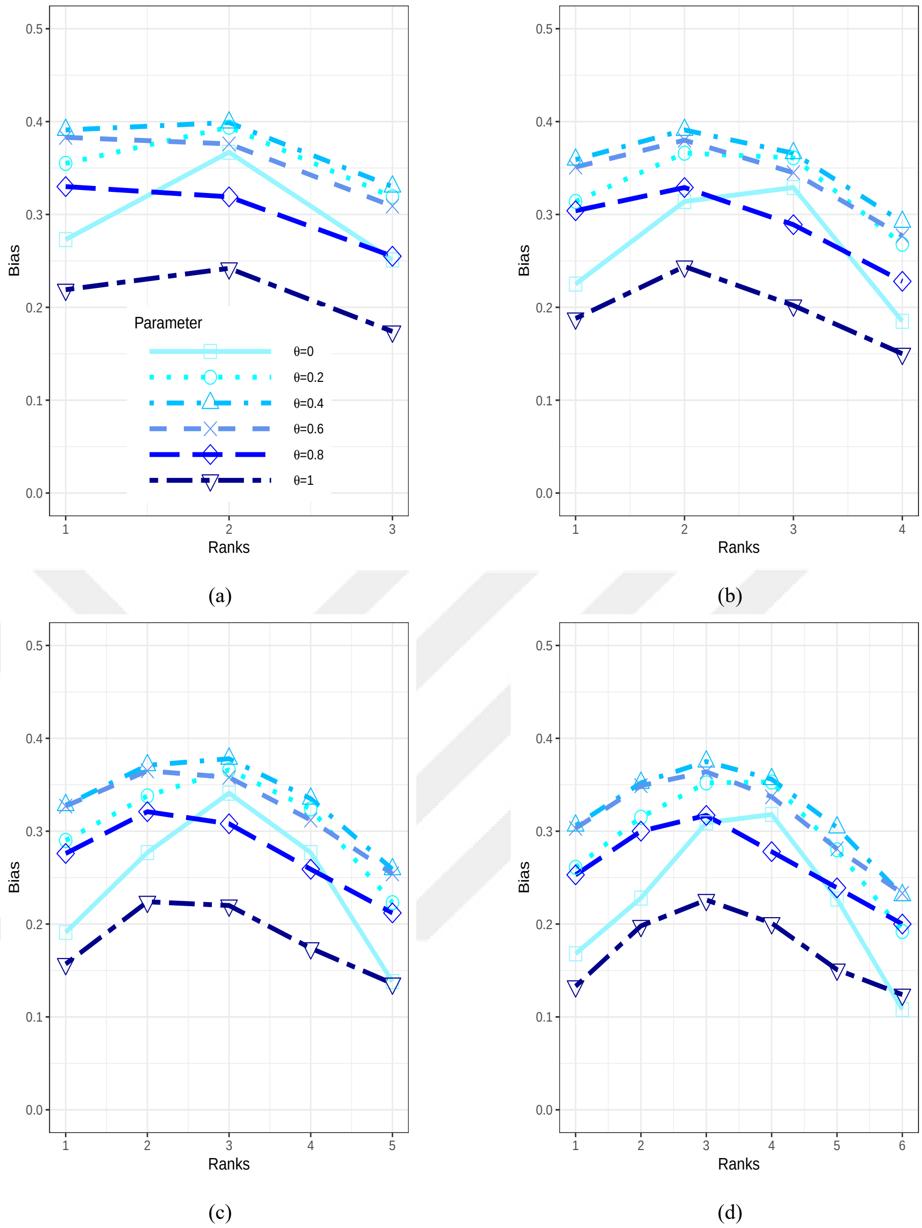
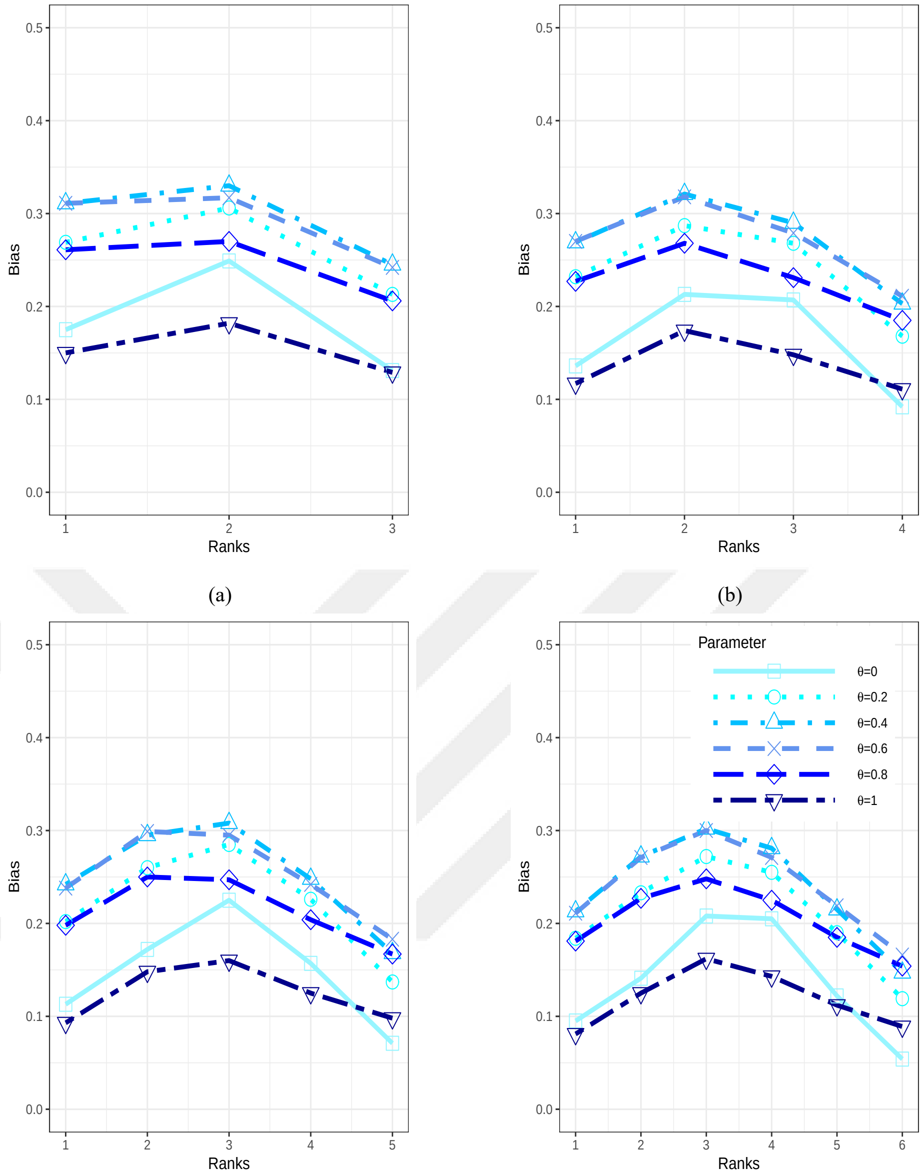


Figure 2.3 For $m = 2$, estimated bias values of $\hat{\theta}_{GMRSS(R=r)}$; (a): $k = 3$, (b): $k = 4$, (c): $k = 5$, and (d): $k = 6$

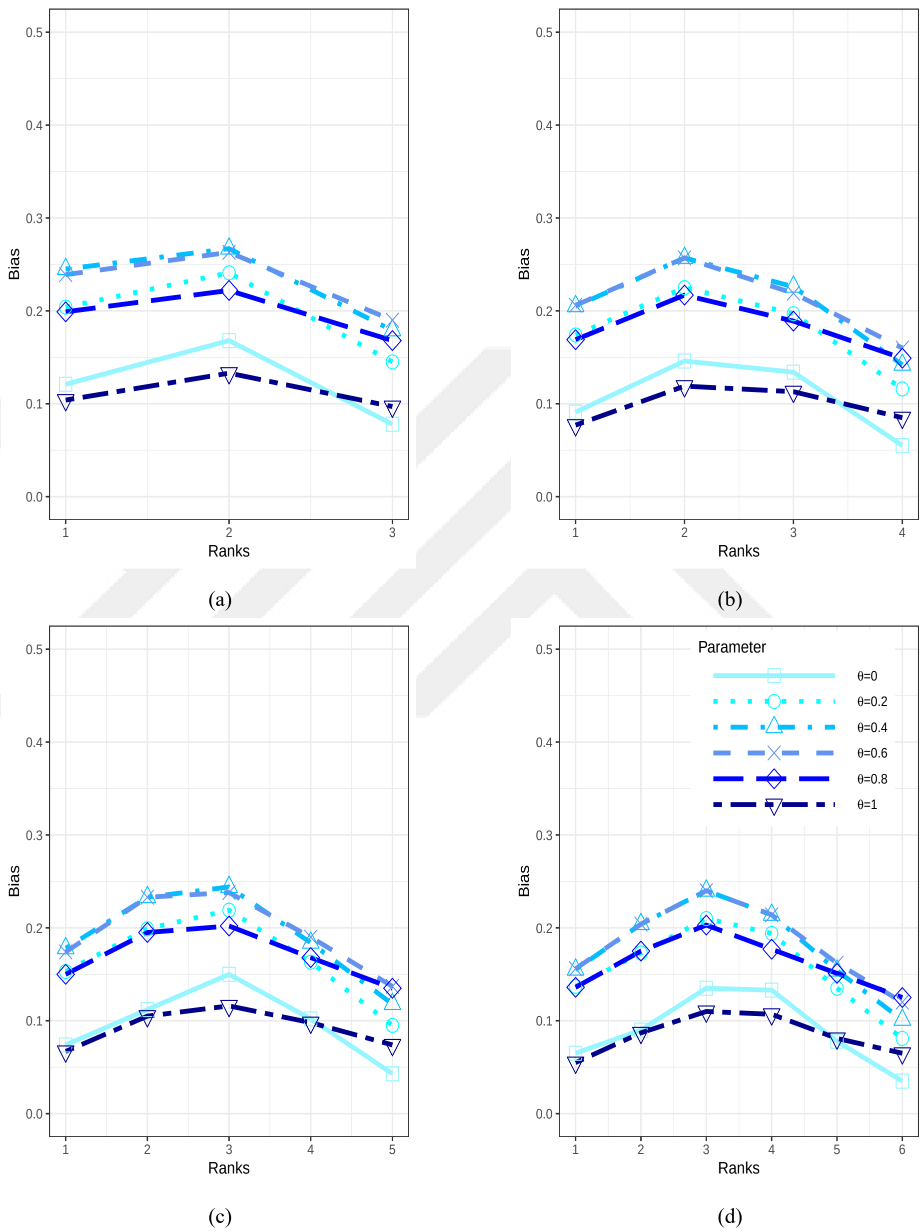
while the set size increases (for $m = 15$). In Figure 2.12, the REs of the ML estimators based on RSS, $GMRSS(R = 1)$, $GMRSS(R = k)$ and ERSS with respect to ML estimator based on SRS are provided. It can be observed that the REs reduce while $\theta \to 1$. Also, there is no monotone increasing or decreasing in REs as the number of cycles increases. On the other hand, it seems that $RE(\hat{\theta}_{GMRSS(R=r)}, \hat{\theta}_{SRS}) > RE(\hat{\theta}_{ERSS}, \hat{\theta}_{SRS}) > RE(\hat{\theta}_{GMRSS(R=r)}, \hat{\theta}_{SRS}) \geq RE(\hat{\theta}_{RSS}, \hat{\theta}_{SRS})$ except for $k = 8$, $m = 10, 15$ and $\theta = 1$. For $k = 8$,



(a) (b) (c) (d)

Figure 2.4 For $m = 5$, estimated bias values of $\hat{\theta}_{GMRSS(R=r)}$; (a): $k = 3$, (b): $k = 4$, (c): $k = 5$, and (d): $k = 6$

$GMRSS(R = 1)$ provides the highest RE when the number of cycles is $m = 10, 15$ and $\theta = 1$. It is shown that the increase in the value of θ less effect on $RE(\hat{\theta}_{GMRSS(R=1)}, \hat{\theta}_{SRS})$ then the other estimators.



(a) (b) (c) (d)

Figure 2.5 For $m = 10$, estimated bias values of $\hat{\theta}_{GMRSS(R=r)}$; (a):$k = 3$, (b):$k = 4$, (c):$k = 5$, and (d):$k = 6$

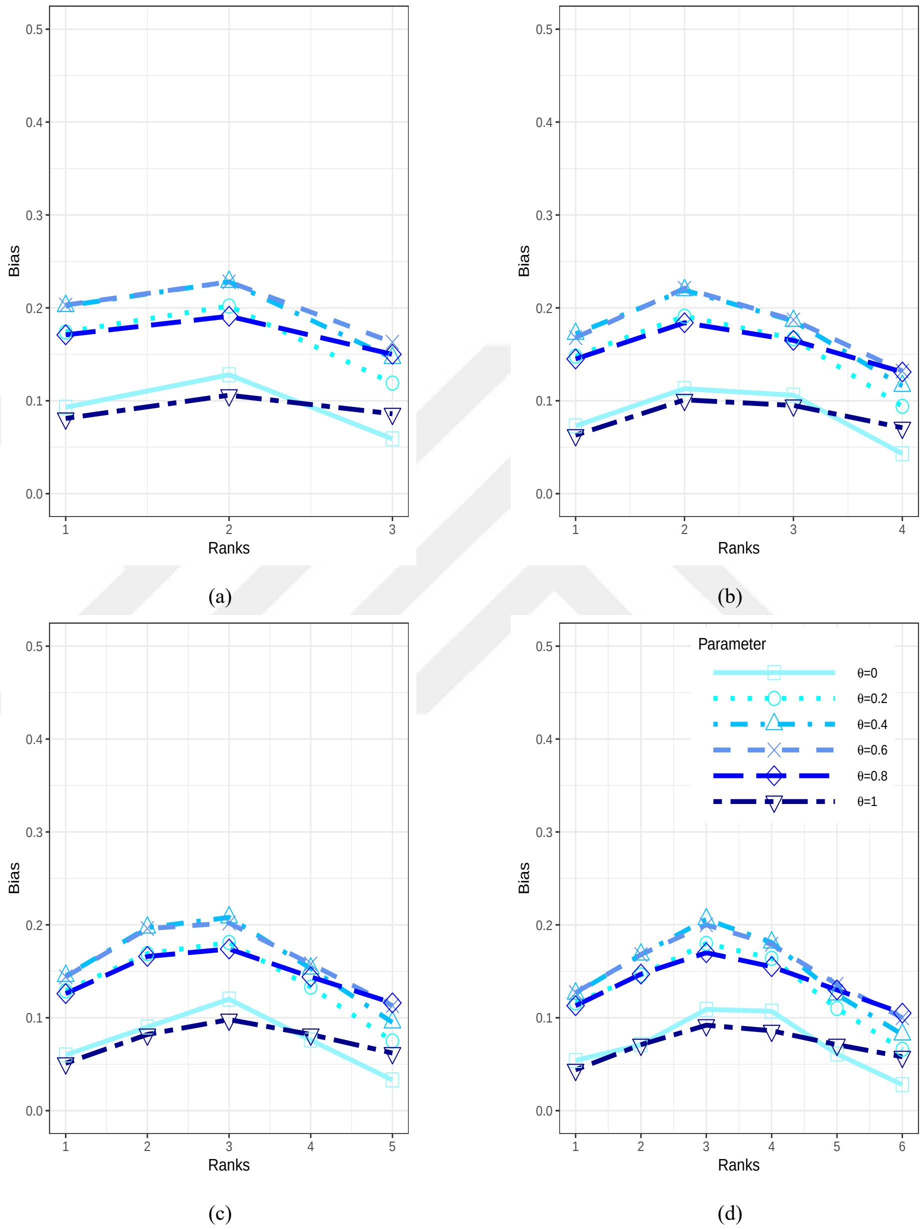


Figure 2.6 For $m = 15$, estimated bias values of $\hat{\theta}_{GMRSS(R=r)}$; (a): $k = 3$, (b): $k = 4$, (c): $k = 5$, and (d): $k = 6$

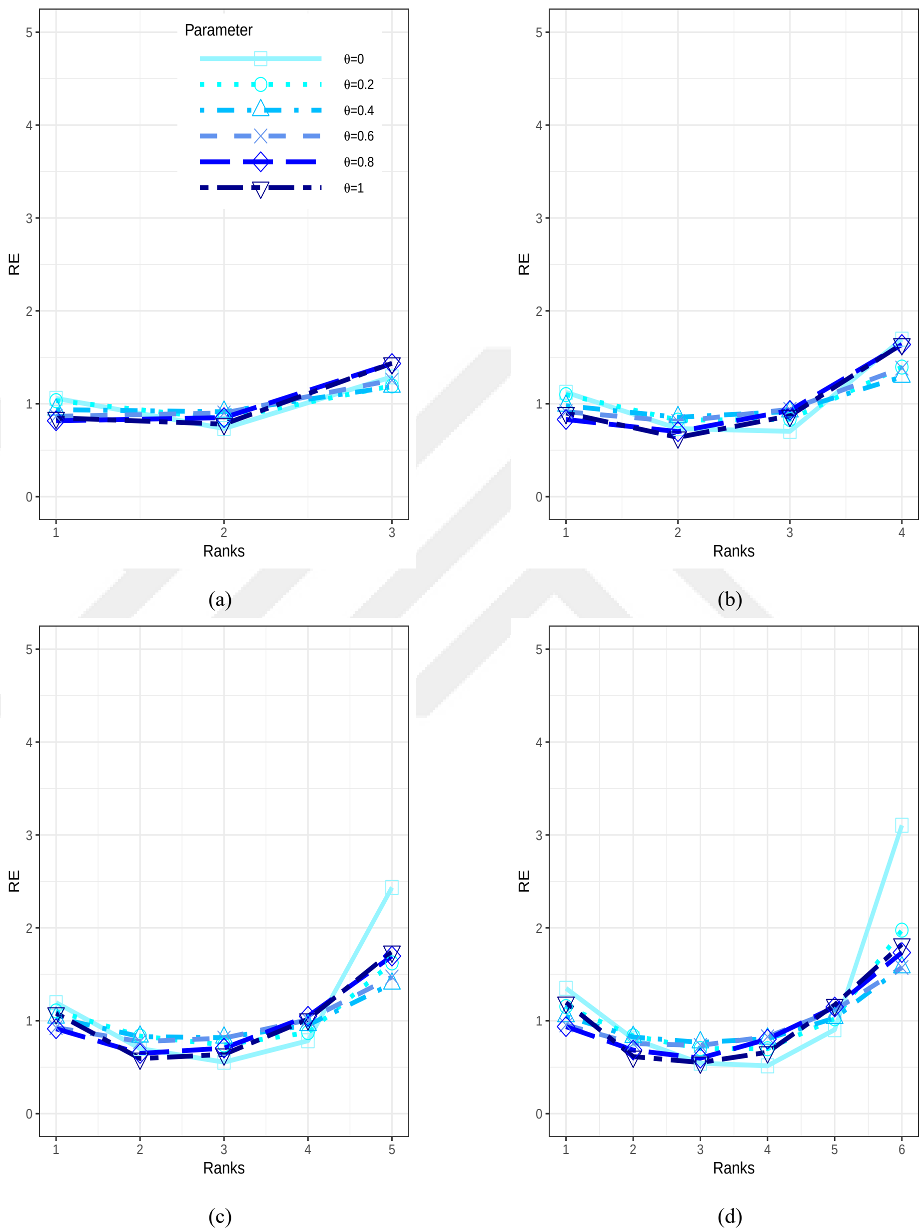


Figure 2.7 For $m = 2$, the estimated values of $RE(\hat{\theta}_{GMRSS(R=r)}, \hat{\theta}_{SRS})$; (a): $k = 3$, (b): $k = 4$, (c): $k = 5$, and (d): $k = 6$

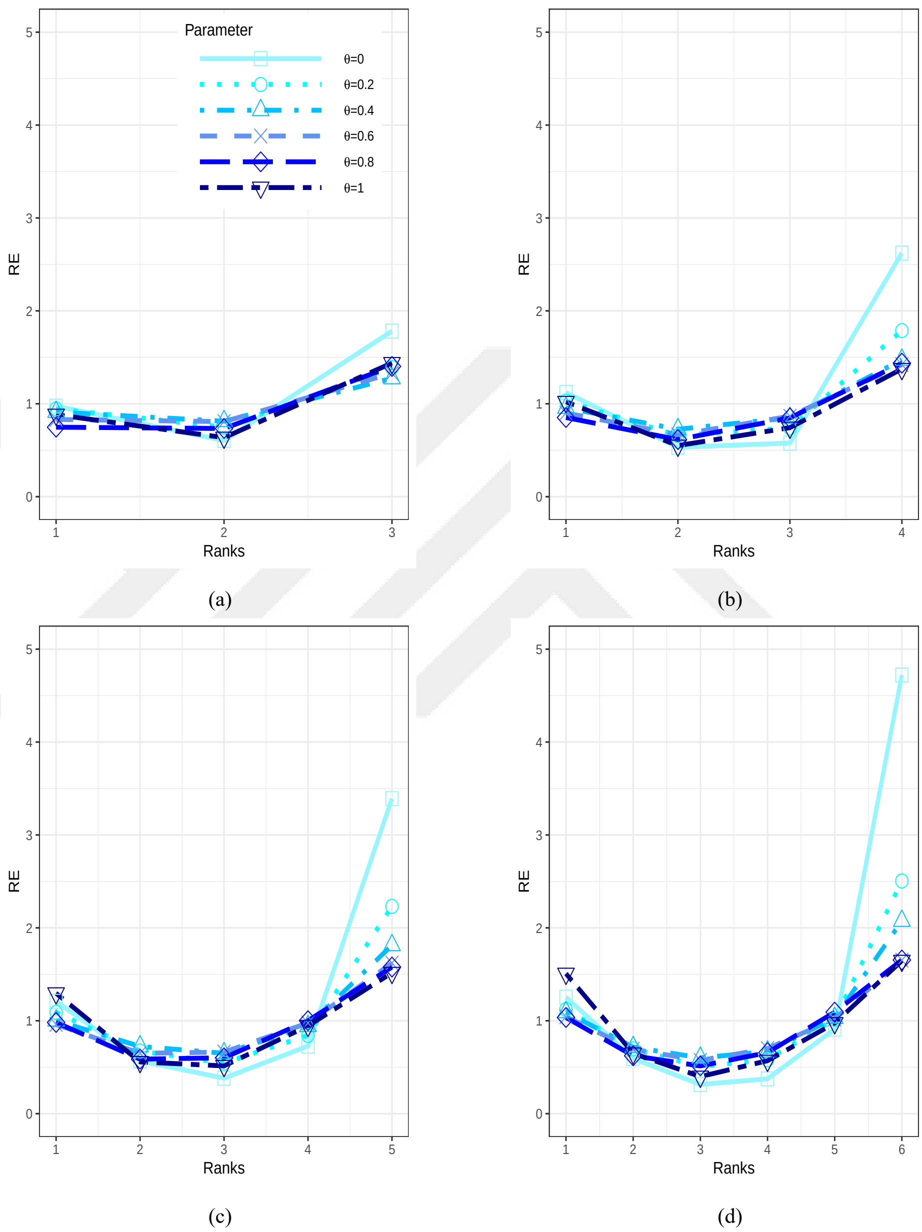


Figure 2.8 For $m = 5$, the estimated values of $RE(\hat{\theta}_{GMRSS(R=r)}, \hat{\theta}_{SRS})$; (a): $k = 3$, (b): $k = 4$, (c): $k = 5$, and (d): $k = 6$

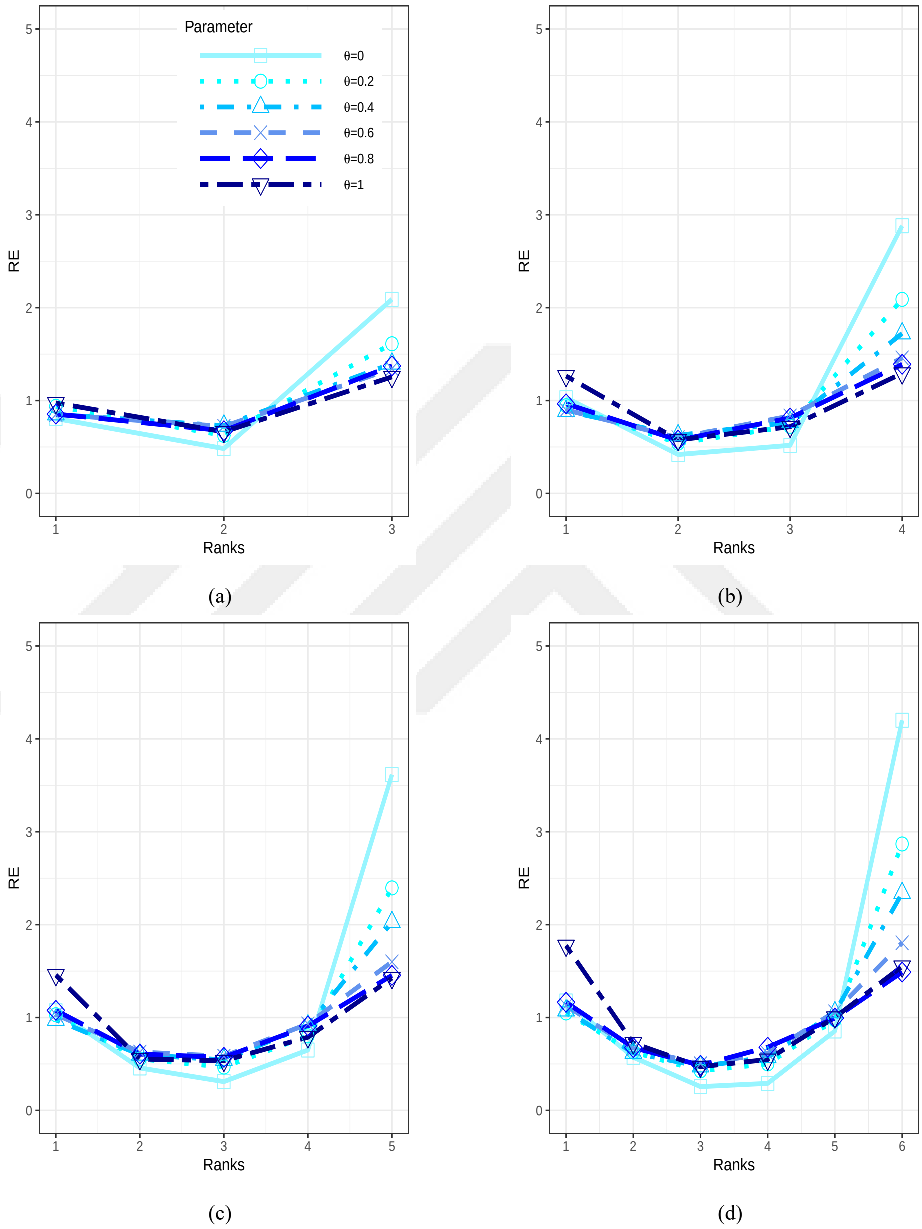


Figure 2.9 For $m = 10$, the estimated values of $RE(\hat{\theta}_{GMRSS(R=r)}, \hat{\theta}_{SRS})$; (a): $k = 3$, (b): $k = 4$, (c): $k = 5$, and (d): $k = 6$

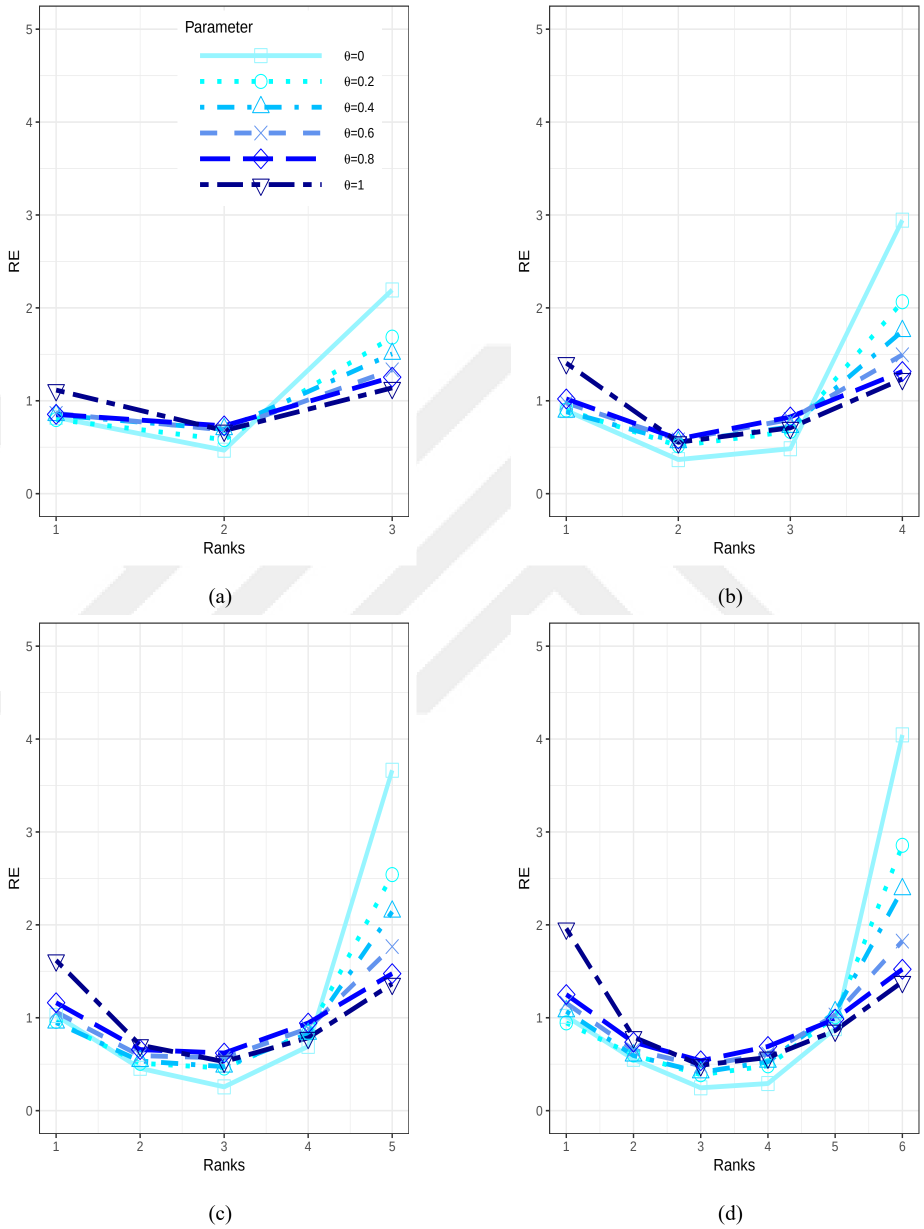


Figure 2.10 For $m = 15$, the estimated values of $RE(\hat{\theta}_{GMRSS(R=r)}, \hat{\theta}_{SRS})$; (a):$k = 3$, (b):$k = 4$, (c):$k = 5$, and (d):$k = 6$

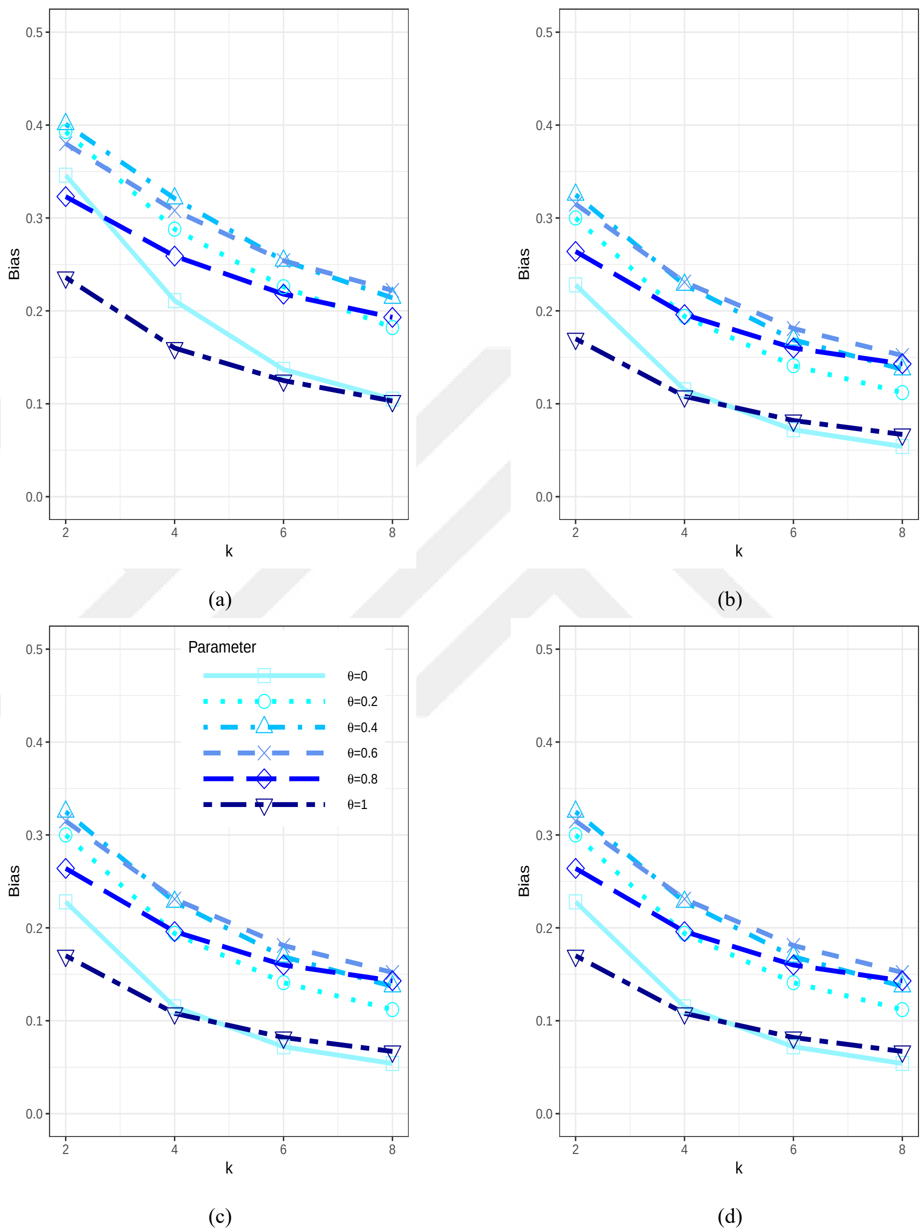
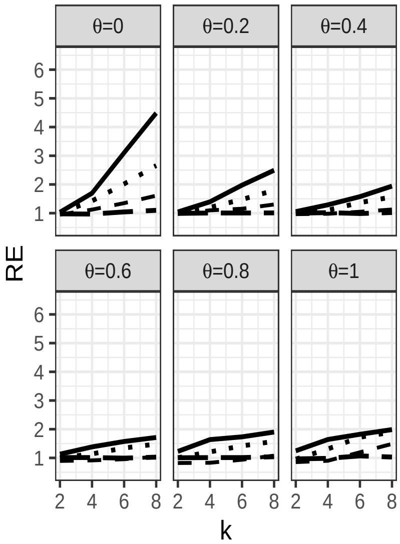
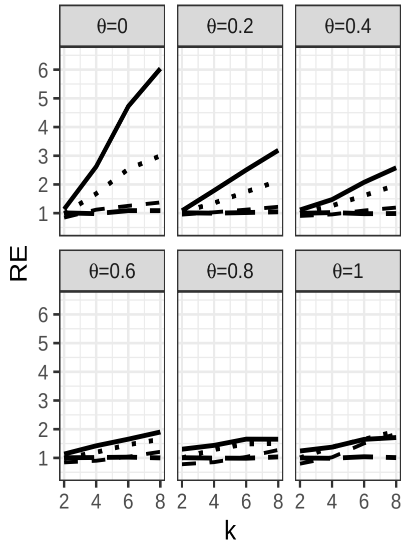


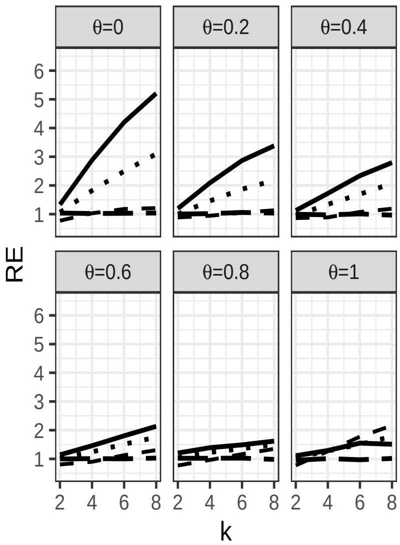
Figure 2.11 The estimated values of $Bias(\hat{\theta}_{ERSS})$; (a): $m = 2$, (b): $m = 5$, (c): $m = 10$, and (d): $m = 15$



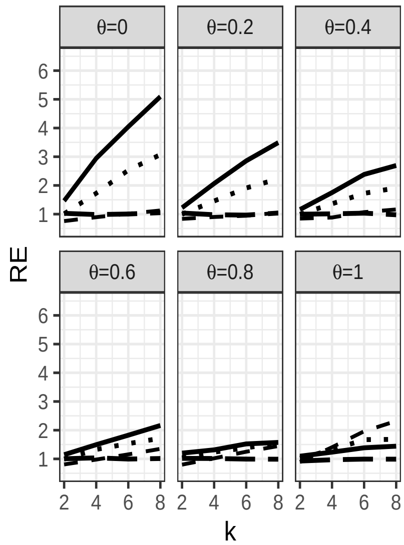
(a)



(b)



(c)



(d) For $m = 15$

Figure 2.12 The estimated values of $RE(\hat{\theta}_{\psi'}, \hat{\theta}_{SRS})$; (a): $m = 2$, (b): $m = 5$, (c): $m = 10$, and (d): $m = 15$ where ψ' = RSS, $GMRSS(R = 1)$, $GMRSS(R = k)$ and ERSS (longdash: RSS, dashed: $GMRSS(R = 1)$, solid: $GMRSS(R = k)$ and dotted: ERSS)

2.1.5 Likelihood Ratio Statistic for Testing the Independence

Let (X, Y) is a sample and $\lambda_\psi(x, y)$ is a LRT statistic for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ where $\psi =$ SRS, RSS, $GMRSS(R = 1)$, $GMRSS(R = k)$ and ERSS.

$$\lambda_\psi (x, y) = \exp \left(L_\psi (\theta_0) - L_\psi \left(\hat{\theta}_\psi \right) \right), \tag{2.35}$$

where $\theta_0 = 0$ and $\hat{\theta}_\psi$ are ML estimators for SRS, RSS, $GMRSS(R = 1)$, $GMRSS(R = k)$ and ERSS. Also, $L_\psi ()$ are log-likelihood functions and are given by Eqs. (2.20), (2.22), (2.24), (2.26.) The rejection region is defined as follows:

$$\{(x, y) : \lambda_\psi (x, y) \leq c_\psi\} \tag{2.36}$$

where c_ψ is any number satisfying $0 \leq c_\psi \leq 1$ and is called cut off point.

Remark 1 *We note that the distribution of $\lambda_\psi (x, y)$ depend on the quality of ranking of X. Hence, in practical terms, it is not feasible to obtain precise critical values as the quality of ranking remains unknown. Against the problem, critical values can be obtained by using a Monte Carlo simulation, see Weer & Basu (1980). Eventually, the critical values can be used to test the null hypothesis.*

The *Algorithm 5* provides cut of points c_ψ for each of the sampling methods, $\psi =$ SRS, RSS, $GMRSS(R = 1)$, $GMRSS(R = k)$ and ERSS. We note that the values of c_ψ

Algorithm 5: For critical value

Step I: Draw a sample of size n by using one of the methods;
Step II: Calculate the LRT statistic by using Eq. (2.22) $\lambda_{\psi,t} (x, y)$;
Step III: Repeat (1)-(2), $t = 1, \cdots, 100,000$;
Step IV: The $(100,000 \times \alpha)$ percentage point of the $\lambda_{\psi,1} (x, y) \leq \cdots \leq$; $\lambda_{\psi,100,000} (x, y)$ is the approximate c_ψ.

---

do not depend on the values θ_0. The values of c_ψ are presented in Appendices section.

Estimated type I errors and the powers of LRTs for SRS, RSS, $GMRSS(R = 1)$,

$GMRSS(R=k)$ and ERSS are obtained by using *Algorithm 6* and are given in Table 2.3. For the *Algorithm 6*, we note that $I\left(\lambda_{\psi,s}(x,y)\leq c_{\psi}\right)$ is an indicator function taking

Algorithm 6: For type I and powers
Step I: Draw a sample of size n by using one of the methods;
Step II: Calculate the LRT statistic by using Eq. (2.22) $\lambda_{\psi,t}\left(x,y\right)$;
Step III: Repeat (1)-(2), $t=1,\cdots,10,000$;
Step IV: Power of $LRT_{\psi}\approx\frac{1}{10{,}000}\sum\limits_{s=1}^{10{,}000}I\left(\lambda_{\psi,s}(x,y)\leq c_{\psi}\right)$.

the value 1 if $\lambda_{\psi,s}(x,y)\leq c_{\psi}$ and 0 otherwise. According to the table, the LRTs based on SRS, RSS, $GMRSS(R=1)$, $GMRSS(R=k)$ and ERSS hold the nominal $\alpha=0.05$ for $\theta=0$. Also, it is noted that power of the $GMRSS(R=k)$ based the test statistic approaches 1 more quickly than those of the other test statistics while $\theta\rightarrow1$.

2.2 Farlie-Gumbel-Morgenstern Type Bivariate Gamma Distribution

The goal of the section is to investigate estimation of dependence parameter of FGM type bivariate gamma distribution. We examine ML estimators based on SRS, RSS, $GMRSS(R=r)$. We make comparison among these estimators in Monte Carlo simulation. Results that support those found in the previous section have occurred.

2.2.1 Preliminaries

Assume that $U=H(x)$ and $V=G(y)$ where the random variables X and Y are continuous. Then, C is called copula which is joint distribution of U and V. According to Sklar's theorem (Sklar, 1959), the bivariate copula is defined,

$$F(x,y)=P\left[H(X)\leq H(x),G(Y)\leq G(y)\right]=C(H(x),G(y))=C\left(u,v\right)\quad(2.37)$$

Table 2.3 Estimated Type I errors and powers of LRT based on SRS, RSS, $GMRSS(R = 1)$, $GMRSS(R = k)$ and ERSS at $\alpha = 0.05$

		LRT_{SRS}				LRT_{RSS}				$LRT_{GMRSS(R=1)}$				$LRT_{GMRSS(R=k)}$				LRT_{ERSS}			
θ	$k \backslash m$	2	5	10	15	2	5	10	15	2	5	10	15	2	5	10	15	2	5	10	15
0	2	0.052	0.049	0.048	0.045	0.051	0.049	0.050	0.054	0.053	0.046	0.049	0.054	0.049	0.049	0.048	0.048	0.051	0.048	0.052	0.049
	3	0.048	0.051	0.048	0.052	0.049	0.048	0.049	0.049	0.053	0.051	0.052	0.052	0.053	0.051	0.048	0.051	$*$	$*$	$*$	$*$
	4	0.048	0.054	0.049	0.047	0.050	0.051	0.052	0.045	0.050	0.057	0.047	0.051	0.051	0.048	0.052	0.053	0.051	0.049	0.052	0.051
	5	0.050	0.050	0.048	0.051	0.051	0.050	0.050	0.051	0.049	0.049	0.050	0.048	0.051	0.050	0.050	0.051	$*$	$*$	$*$	$*$
	6	0.050	0.048	0.052	0.052	0.052	0.050	0.053	0.050	0.051	0.046	0.046	0.053	0.051	0.049	0.047	0.052	0.047	0.050	0.046	0.053
	7	0.049	0.049	0.051	0.048	0.047	0.052	0.055	0.049	0.050	0.055	0.051	0.047	0.049	0.049	0.053	0.052	$*$	$*$	$*$	$*$
	8	0.051	0.047	0.049	0.051	0.049	0.048	0.050	0.052	0.052	0.050	0.049	0.052	0.049	0.049	0.052	0.052	0.049	0.048	0.048	0.051
0.2	2	0.081	0.110	0.147	0.205	0.081	0.118	0.168	0.205	0.086	0.098	0.130	0.168	0.087	0.114	0.188	0.261	0.084	0.117	0.158	0.211
	3	0.096	0.138	0.206	0.273	0.096	0.131	0.196	0.273	0.091	0.127	0.187	0.231	0.106	0.181	0.302	0.425	$*$	$*$	$*$	$*$
	4	0.098	0.163	0.253	0.334	0.108	0.157	0.252	0.342	0.108	0.160	0.225	0.301	0.136	0.259	0.456	0.626	0.126	0.206	0.341	0.464
	5	0.113	0.181	0.294	0.406	0.116	0.179	0.298	0.412	0.122	0.186	0.291	0.370	0.176	0.356	0.615	0.780	$*$	$*$	$*$	$*$
	6	0.126	0.207	0.336	0.466	0.120	0.203	0.339	0.464	0.135	0.220	0.333	0.446	0.217	0.449	0.739	0.893	0.179	0.336	0.548	0.727
	7	0.120	0.232	0.374	0.511	0.133	0.234	0.380	0.515	0.145	0.242	0.377	0.488	0.277	0.563	0.840	0.950	$*$	$*$	$*$	$*$
	8	0.141	0.255	0.420	0.561	0.141	0.251	0.417	0.557	0.167	0.281	0.440	0.551	0.320	0.651	0.908	0.983	0.236	0.479	0.745	0.888
0.4	2	0.127	0.210	0.340	0.462	0.123	0.197	0.328	0.469	0.127	0.181	0.278	0.354	0.136	0.221	0.400	0.572	0.128	0.203	0.342	0.460
	3	0.144	0.269	0.451	0.615	0.153	0.272	0.459	0.623	0.156	0.251	0.395	0.522	0.190	0.384	0.679	0.846	$*$	$*$	$*$	$*$
	4	0.178	0.336	0.567	0.736	0.179	0.337	0.571	0.736	0.185	0.326	0.525	0.667	0.269	0.595	0.876	0.973	0.235	0.449	0.733	0.882
	5	0.200	0.394	0.659	0.827	0.204	0.403	0.663	0.831	0.231	0.410	0.642	0.778	0.375	0.757	0.966	0.998	$*$	$*$	$*$	$*$
	6	0.228	0.468	0.738	0.883	0.226	0.457	0.750	0.883	0.261	0.484	0.729	0.860	0.487	0.874	0.995	1.000	0.368	0.723	0.946	0.992
	7	0.248	0.509	0.796	0.921	0.251	0.512	0.800	0.921	0.312	0.562	0.808	0.922	0.593	0.946	0.999	1.000	$*$	$*$	$*$	$*$
	8	0.282	0.567	0.839	0.948	0.278	0.567	0.847	0.952	0.346	0.639	0.858	0.953	0.691	0.982	1.000	1.000	0.515	0.886	0.993	1.000
0.6	2	0.171	0.309	0.531	0.708	0.168	0.314	0.536	0.711	0.168	0.276	0.455	0.585	0.190	0.349	0.639	0.815	0.172	0.307	0.540	0.699
	3	0.218	0.421	0.703	0.855	0.222	0.428	0.706	0.857	0.232	0.398	0.636	0.777	0.288	0.616	0.901	0.980	$*$	$*$	$*$	$*$
	4	0.263	0.548	0.820	0.938	0.261	0.535	0.827	0.940	0.294	0.546	0.788	0.907	0.422	0.832	0.988	0.999	0.349	0.692	0.936	0.988
	5	0.311	0.632	0.894	0.974	0.318	0.628	0.897	0.976	0.364	0.648	0.887	0.965	0.584	0.947	1.000	1.000	$*$	$*$	$*$	$*$
	6	0.372	0.700	0.934	0.988	0.357	0.697	0.938	0.990	0.429	0.745	0.942	0.987	0.720	0.985	1.000	1.000	0.566	0.930	0.998	1.000
	7	0.406	0.767	0.963	0.996	0.401	0.774	0.967	0.996	0.493	0.823	0.972	0.997	0.834	0.998	1.000	1.000	$*$	$*$	$*$	$*$
	8	0.448	0.814	0.980	0.999	0.447	0.822	0.983	0.998	0.555	0.871	0.989	0.999	0.908	0.999	1.000	1.000	0.766	0.990	1.000	1.000
0.8	2	0.224	0.441	0.722	0.871	0.221	0.437	0.714	0.872	0.210	0.389	0.636	0.784	0.237	0.485	0.797	0.939	0.219	0.448	0.716	0.873
	3	0.293	0.601	0.872	0.963	0.294	0.604	0.867	0.962	0.320	0.559	0.827	0.933	0.390	0.786	0.980	0.998	$*$	$*$	$*$	$*$
	4	0.372	0.717	0.941	0.988	0.369	0.710	0.943	0.990	0.408	0.737	0.932	0.984	0.572	0.944	0.999	1.000	0.492	0.859	0.988	1.000
	5	0.435	0.805	0.979	0.998	0.433	0.804	0.979	0.998	0.515	0.841	0.977	0.997	0.743	0.992	1.000	1.000	$*$	$*$	$*$	$*$
	6	0.518	0.868	0.990	1.000	0.510	0.873	0.992	1.000	0.606	0.914	0.993	1.000	0.873	0.999	1.000	1.000	0.747	0.987	1.000	1.000
	7	0.565	0.915	0.996	1.000	0.571	0.919	0.997	1.000	0.683	0.951	0.998	1.000	0.944	1.000	1.000	1.000	$*$	$*$	$*$	$*$
	8	0.608	0.944	0.999	1.000	0.628	0.948	0.999	1.000	0.760	0.978	1.000	1.000	0.977	1.000	1.000	1.000	0.916	1.000	1.000	1.000
1	2	0.280	0.567	0.851	0.953	0.270	0.568	0.850	0.951	0.255	0.523	0.788	0.910	0.288	0.607	0.902	0.982	0.278	0.570	0.847	0.954
	3	0.376	0.738	0.956	0.992	0.377	0.742	0.955	0.994	0.398	0.738	0.936	0.987	0.480	0.889	0.995	1.000	$*$	$*$	$*$	$*$
	4	0.474	0.843	0.986	0.999	0.476	0.858	0.988	0.999	0.526	0.883	0.987	0.999	0.686	0.984	1.000	1.000	0.616	0.947	0.999	1.000
	5	0.566	0.917	0.998	1.000	0.569	0.924	0.996	1.000	0.665	0.951	0.999	1.000	0.854	0.999	1.000	1.000	$*$	$*$	$*$	$*$
	6	0.650	0.953	0.999	1.000	0.650	0.958	0.999	1.000	0.781	0.983	1.000	1.000	0.947	1.000	1.000	1.000	0.874	0.998	1.000	1.000
	7	0.708	0.978	1.000	1.000	0.714	0.980	1.000	1.000	0.847	0.995	1.000	1.000	0.983	1.000	1.000	1.000	$*$	$*$	$*$	$*$
	8	0.765	0.985	1.000	1.000	0.764	0.988	1.000	1.000	0.900	0.998	1.000	1.000	0.995	1.000	1.000	1.000	0.972	1.000	1.000	1.000

where $F(x,y)$ is joint CDF on $\mathbb{R}^2$ with marginal CDFs $H(x)$ and $G(y)$. If $H(x)$ and $G(y)$ are continuous, C is unique for a pair $(U,V) \in [0,1]^2$. The density function of the bivariate copula can be expressed by taking partial derivatives of C.

$$c\left(u,v\right) = \frac{\partial^2 C\left(u,v\right)}{\partial u \partial v} \tag{2.38}$$

Thus, $f(x,y) = c\left(u,v\right) h(x) g(y)$ where $h(x)$ and $g(y)$ are the PDFs of marginals. The following property gives some characteristics of C.

Property 2.2.1 *Let C be bivariate copula defining on $[0,1]^2$,*

1. *$C(u,1) = u$ and $C(1,v) = v$ for every $u \in [0,1]$ and $v \in [0,1]$.*
2. *$C(u,0) = C(0,v) = 0$ for every $u \in [0,1]$ and $v \in [0,1]$.*
3. *$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$ for every u_1, u_2, v_1 and v_2 in $[0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$.*

Let $c_u(v)$ be conditional distribution function for V given $U = u$,

$$c_u(v) = P\left[V \leq v | U = u\right] = \lim_{\Delta u \to 0} \frac{C\left(u + \Delta u, v\right) - C(u,v)}{\Delta u} = \frac{\partial C(u,v)}{\partial u} \tag{2.39}$$

The following theorem indicates that $c_u(v)$ exists and is nondecreasing almost everywhere $v \in [0,1]$.

Theorem 2.2.2 (Nelsen (2007)) *Let C be a copula. For any $v \in [0,1]$, the partial derivative $\partial C\left(u,v\right)/\partial u$ exists and $0 \leq \frac{\partial C(u,v)}{\partial u} \leq 1$ for almost all u, and for such v and u. Moreover, the function $v \mapsto \partial C\left(u,v\right)/\partial u$ is defined and nondecreasing almost everywhere on $v \in [0,1]$.*

FGM is a quite attractive distribution family due to its straightforward structure and provide a modelling the dependence between two random variables. The distribution family was introduced by Morgenstern (1956), Gumbel (1960) and Farlie (1960). Let

$H(x)$ and $G(y)$ are CDFs of the marginals, the joint CDF of the random variables X and Y on $\mathbb{R}^2$ is given by

$$F(x,y) = H(x)G(y)\left[1 + \lambda\left(1 - H(x)\right)\left(1 - G(y)\right)\right] \tag{2.40}$$

and the corresponding joint PDF is

$$f(x,y) = h(x)g(y)\left[1 + \lambda\left(1 - 2H(x)\right)\left(1 - 2G(y)\right)\right] \tag{2.41}$$

where $h(x)$ and $g(y)$ are PDFs of the marginals. In the Eqs. (2.40) and (2.41), λ denotes the dependence parameter where $-1 \leq \lambda \leq 1$. If $\lambda = 0$, the random variables are said to be independent; however, if $\lambda = 1$ (or $\lambda = -1$), they are said to be dependent. Also, FGM family is known as a "bivariate copula with cubic section in both u and v" where $u = H(x)$ and $v = G(y)$. Nelsen et al. (1997) first coined this phrase and defined the following equation. Suppose that C is a bivariate copula with cubic section,

$$\begin{aligned} C(u,v) = uv + uv\left(1-u\right)\left(1-v\right)[A_1 v\left(1-u\right) + \\ A_2\left(1-v\right)\left(1-u\right) + B_1 uv + B_2 u\left(1-v\right)] \end{aligned} \tag{2.42}$$

where A_1, A_2, B_1 and B_2 are the points in the region which is given by the Figure 2.13. If $A_1 = A_2 = B_1 = B_2 = \lambda$, then the expression of FGM is obtained from the Eq.

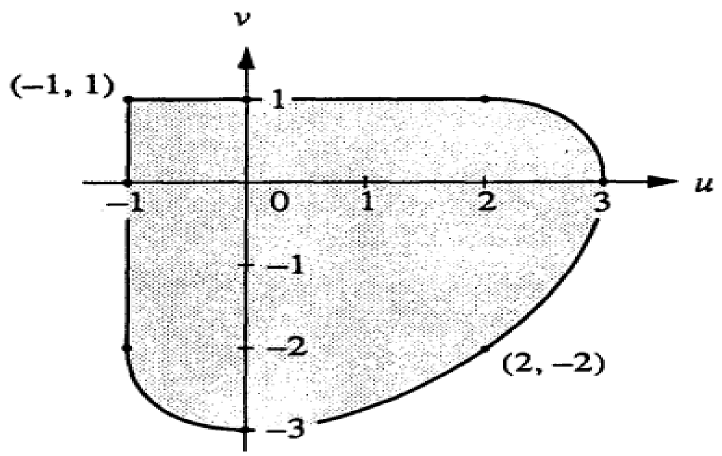


Figure 2.13 The region for copulas cubic section (Nelsen et al., 1997)

(2.42).

$$C(u,v) = uv\left[1 + \lambda(1-u)(1-v)\right] \tag{2.43}$$

where $u, v \in [0,1]$. Since FGM is a popular copula family, several researchers have studied on FGM distribution family, see Bairamov & Kotz (2000), Abo-Eleneen & Nagaraja (2002), Ucer & Yildiz (2012), Ucer & Gurler (2012), Gurler et al. (2015), Yildiz & Ucer (2017).

D'Este (1981) and Gupta & Wong (1984) considered FGM family with gamma marginals. The PDFs of the random variables X and Y on $\mathbb{R}^2$ are

$$h(x) = \frac{1}{\Gamma(a)} e^{-x} x^{a-1}, \quad x > 0, \quad a \geq 0, \tag{2.44}$$

and

$$g(y) = \frac{1}{\Gamma(b)} e^{-y} y^{b-1}, \quad y > 0, \quad b \geq 0. \tag{2.45}$$

where α and β are shape parameters of X and Y, respectively. The CDFs of the random variables X and Y,

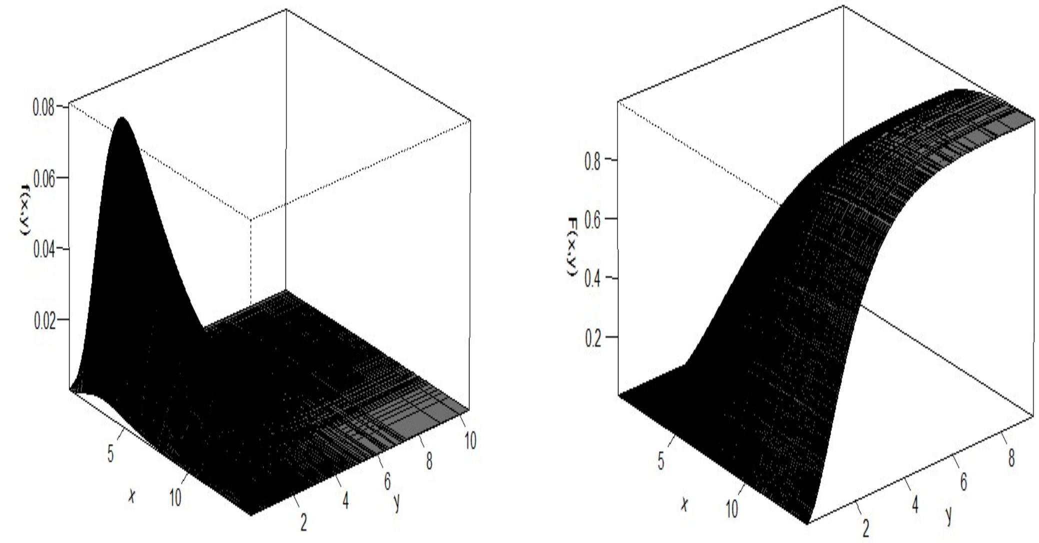
$$H(x) = \frac{\gamma(a,x)}{\Gamma(a)}, \quad \text{and} \quad G(y) = \frac{\gamma(b,y)}{\Gamma(b)}. \tag{2.46}$$

Also, $\Gamma(.)$ and $\gamma(.,.)$ are gamma and incomplete gamma functions, respectively. The following figure illustrates the joint PDF and CDF for FGM type bivariate gamma distribution. Some characteristics of FGM type bivariate gamma distribution was investigated by Gupta & Wong (1984). In their work, the joint moment function was defined as follows:

$$M(s,t) = (1-s)^{-a}(1-t)^{b}\left[1 + \frac{2I\left(a,0;(1-s)^{-1}\right)}{I(a,0;1)} \frac{2I\left(b,0;(1-t)^{-1}\right)}{I(b,0;1)}\right] \tag{2.47}$$

where

$$I(a,w;\vartheta) = \int_0^{\vartheta} \frac{x^{a-1}}{(x+1)^{2a+w}} dx. \tag{2.48}$$



(a) (b)

Figure 2.14 FGM type bivariate gamma with parameters $a = 2$, $b = 5$ and $\lambda = 0.7$; (a): Joint PDF and (b): Joint CDF

The correlation coefficient between X and Y is,

$$Cor(x,y) = \rho = \lambda K(a) K(b). \tag{2.49}$$

where

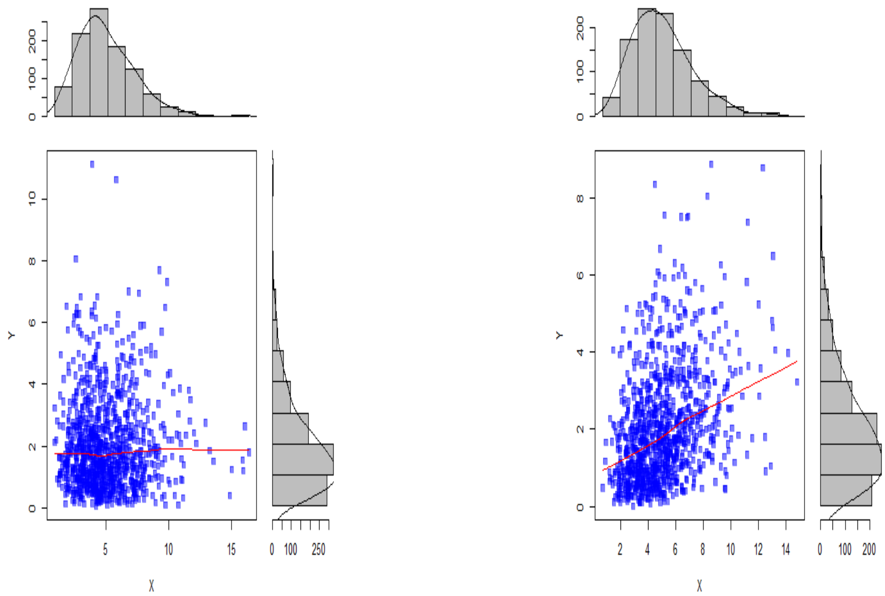
$$K(a) = \frac{\Gamma(2a)}{2^{2a-1}\Gamma^2(a)\sqrt{a}}. \tag{2.50}$$

An illustration for the correlation coefficient between X and Y is given by the following figure. Thus, the estimator for ρ is,

$$\hat{\rho} = \hat{\lambda} K(\hat{a}) K\left(\hat{b}\right). \tag{2.51}$$

The regression function was also provided as following:

$$\begin{aligned} E[Y|X=x] &= E[Y] + \lambda E[Y]\left\{\frac{2I(a,1;1)}{B(a,a+1)} - 1\right\}\{2F(x,y)-1\} \\ &= a + \frac{\lambda a \Gamma(a+1/2)}{\sqrt{\pi}(a+1)}\{2F(x,y)-1\}. \end{aligned} \tag{2.52}$$



(a) (b)

Figure 2.15 Correlation between X and Y; (a): $\lambda = 0$ and (b): $\lambda = 1$

where the duplication formula is

$$\Gamma(a)\,\Gamma(a + 1/2) = 2^{1/2-2a}\Gamma(2a)\sqrt{2\pi}. \tag{2.53}$$

2.2.2 Algorithms to Generate Ranked-based Samples

This section provides some algorithms that use copula tools. First, we have to generate the pairs (X, Y) with a joint distribution function $F(x, y)$. Sklar's theorem states that we need generate standard uniform random variables whose joint CDF is C, which is given by Eq. (2.43), and then transform those uniform random variables. To generate the pairs (U, V) from C, the conditional distribution approach can be used. This approach is provided by **Approach I** in the Section 2.1.2.

Using the approach and Theorem 2.2.2, the *Algorithm 7* was provided by Johnson

(1987) and Nelsen (2007). Recall that c_u is conditional distribution function,

$$
\begin{aligned}
c_u(v) = P\left[V \leq v | U = u\right] = \lim_{\Delta u \to 0} \frac{C\left(u + \Delta u, v\right) - C\left(u, v\right)}{\Delta u} &= \frac{\partial C\left(u, v\right)}{\partial u} \\
= \left[1 + \lambda\left(1 - 2u\right)\right] v - \lambda\left(1 - 2u\right) v^2. &
\end{aligned} \tag{2.54}
$$

A quasi-inverse (c_u^{-1}) of c_u was given by Johnson (1987). The author obtained the c_u^{-1} by solving the equation, $\left[1 + \lambda\left(1 - 2u\right)\right] v - \lambda\left(1 - 2u\right) v^2 = p$, which is quadratic in v. This equation has one root for $0 < p < 1$ and it is $v = c_u^{-1}(p) = \frac{2\zeta}{\varpi + \varkappa}$ where $\varkappa = 1 + \lambda\left(1 - 2u\right)$ and $\varpi = \sqrt{\varkappa^2 - 4\left(\varkappa - 1\right) p}$. In the *Algorithm 7*, H^{-1} and G^{-1}

Algorithm 7: Generating data from FGM type bivariate gamma distribution

Step I: Generate u and p from i.i.d. random variables $U(0, 1)$;
Step II: $\varkappa = 1 + \lambda\left(1 - 2u\right)$ and $\varpi = \sqrt{\varkappa^2 - 4\left(\varkappa - 1\right) p}$;
Step III: Set $v = c_u^{-1}(\zeta) = \frac{2\zeta}{\varpi + \varkappa}$;
Step IV: The desired pair is (u, v);
Step V: $X = H^{-1}(u)$ and $Y = G^{-1}(v)$;
Step VI: RETURN (X, Y).

are inverse functions of H and G, respectively. The desired pair (X, Y) can be obtained by using “qgamma” function in R statistical programming language because there is no closed form of inverse function of gamma distribution function.

Algorithm 2 and *Algorithm 3* are useful for generating RSS and $GMRSS(R = r)$ data from FGM type bivariate gamma distribution, but this time *Algorithm 7* is used in the first step of the algorithms. We construct R functions for generating ranked-based samples from FGM type bivariate gamma distribution and these R functions will be provided upon request.

2.2.3 Maximum Likelihood Estimates

Suppose that $\Theta = \left(a, b, \lambda\right)^T$ is a parameter vector including shape parameters (α and β) and dependence parameter (λ). In this section, ML estimators based on SRS, RSS and $GMRSS(R = r)$ are investigated under the case (i) which assumes that shape parameters are known. Examining the ML estimators for λ beginning with case (i), can be useful and provide some guidance for cases (ii) and (iii).

2.2.3.1 Simple Random Sampling

It is assumed that (X_i, Y_i), $i = 1, \cdots, n$, is a simple random sample. The likelihood function is

$$\mathscr{L}_{SRS} = \prod_{i=1}^{n} f(x_i, y_i) \tag{2.55}$$

and the log-likelihood function is defined

$$L_{SRS}(\lambda) = \sum_{i=1}^{n} \log f(x_i, y_i) \tag{2.56}$$

First derivative of the log-likelihood function provides the following equation.

$$\begin{aligned} \frac{\partial L_{SRS}(\lambda)}{\partial \lambda} &= \sum_{i=1}^{n} \frac{f'(x_i, y_i)}{f(x_i, y_i)} \\ &= \sum_{i=1}^{n} \frac{(1-2u_i)(1-2v_i)}{1+\lambda(1-2u_i)(1-2v_i)} = 0 \end{aligned} \tag{2.57}$$

where $f'(x_i, y_i)$ is the first derivative of the $f(x_i, y_i)$ with respect to λ. In the Eq. (2.57), $u = H^{-1}(x)$ and $v = G^{-1}(y)$. By using a numerical optimization method, the Eqs. (2.56) or (2.57) can be solved and the ML estimator based on SRS $\hat{\lambda}_{SRS}$ is obtained. The variance of $\hat{\lambda}_{SRS}$ is defined by the following equation.

$$\begin{aligned} V\left(\hat{\lambda}_{SRS}\right) &= \left\{-E\left[\frac{\partial^2 L_{SRS}(\lambda)}{\partial \lambda^2}\right]\right\}^{-1} \\ &= \left\{E\left[\sum_{i=1}^{n}\left(\frac{(1-2u_i)(1-2v_i)}{1+\lambda(1-2u_i)(1-2v_i)}\right)^2\right]\right\}^{-1} \end{aligned} \tag{2.58}$$

2.2.3.2 Ranked Set Sampling

Suppose that $RSS(k, m)$ is obtained from FGM type bivariate gamma distribution. In $RSS(k, m)$, the pairs $(X_{(r)j}, Y_{[r]j})$ follows the joint PDF $f_{r:k}(x, y)$ where

$r = 1, \cdots, k$ and $j = 1, \cdots, m$. The likelihood function is given by

$$\mathscr{L}_{RSS} = \prod_{j=1}^{m} \prod_{r=1}^{k} f_{r:k}\left(X_{(r)j}, Y_{[r]j}\right) \tag{2.59}$$

Following that, the log-likelihood function is defined by

$$\begin{aligned} L_{RSS}(\lambda) =& \sum_{j=1}^{m} \sum_{r=1}^{k} \log f_{r:k}\left(X_{(r)j}, Y_{[r]j}\right) \\ =& \sum_{j=1}^{m} \sum_{r=1}^{k} \log\left(\frac{k!}{(r-1)!(k-r)!}\right) \\ &+ \sum_{j=1}^{m} \sum_{r=1}^{k} (r-1) \log\left(H(x_{(r)j})\right) \\ &+ \sum_{j=1}^{m} \sum_{r=1}^{k} (k-r) \log\left(1 - H(x_{(r)j})\right) \\ &+ \sum_{j=1}^{m} \sum_{r=1}^{k} \log\left(f(x_{(r)j}, y_{[r]j})\right) \end{aligned} \tag{2.60}$$

Since λ appears in $f(x, y)$,

$$\begin{aligned} \frac{\partial L_{RSS}(\lambda)}{\partial \lambda} &= \sum_{j=1}^{m} \sum_{r=1}^{k} \frac{f'(x_{(r)j}, y_{[r]j})}{f(x_{(r)j}, y_{[r]j})} \\ &= \sum_{j=1}^{m} \sum_{r=1}^{k} \frac{\left(1 - 2u_{(r)j}\right)\left(1 - 2v_{[r]j}\right)}{1 + \lambda\left(1 - 2u_{(r)j}\right)\left(1 - 2v_{[r]j}\right)} = 0 \end{aligned} \tag{2.61}$$

where $u_{(r)j} = F^{-1}(x_{(r)j})$ and $v_{[r]j} = G^{-1}(y_{[r]j})$. The Eq. (2.60) or (2.61) can be solved by using a computer algorithm and the ML estimator based on RSS ($\hat{\lambda}_{RSS}$) is obtained. Also, the variance of $\hat{\lambda}_{RSS}$ can be found by using the following equation

$$\begin{aligned} V\left(\hat{\lambda}_{RSS}\right) &= \left\{-E\left[\frac{\partial^2 L_{RSS}(\lambda)}{\partial \lambda^2}\right]\right\}^{-1} \\ &= \left\{E\left[\sum_{j=1}^{m} \sum_{r=1}^{k} \left(\frac{\left(1 - 2u_{(r)j}\right)\left(1 - 2v_{[r]j}\right)}{1 + \lambda\left(1 - 2u_{(r)j}\right)\left(1 - 2v_{[r]j}\right)}\right)^2\right]\right\}^{-1} \end{aligned} \tag{2.62}$$

2.2.3.3 Generalized Modified Ranked Set Sampling

Now, we investigate the ML estimator based on $GMRSS(R=r)$ ($\hat{\lambda}_{GMRSS(R=r)}$). Let $GMRSS(R=r)$ is generated (or is selected from a population) from FGM type bivariate gamma distribution. The following equation gives the likelihood function.

$$\mathscr{L}_{GMRSS(R=r)} = \prod_{j=1}^{m}\prod_{\tau=1}^{k} f_{r:k}\left(X_{\tau(r)j}, Y_{\tau[r]j}\right), \tag{2.63}$$

The log-likelihood function is

$$\begin{aligned} L_{GMRSS(R=r)}(\lambda) =& \sum_{j=1}^{m}\sum_{\tau=1}^{k} \log f_{r:k}\left(X_{\tau(r)j}, Y_{\tau[r]j}\right) \\ =& mk \log\left(\frac{k!}{(r-1)!(k-r)!}\right) \\ &+ (r-1)\sum_{j=1}^{m}\sum_{r=1}^{k} \log\left(H(x_{\tau(r)j})\right) \\ &+ (k-r)\sum_{j=1}^{m}\sum_{r=1}^{k} \log\left(1-H(x_{\tau(r)j})\right) \\ &+ \sum_{j=1}^{m}\sum_{r=1}^{k} \log\left(f(x_{\tau(r)j}, y_{\tau[r]j})\right) \end{aligned} \tag{2.64}$$

First derivative of $L_{GMRSS(R=r)}(\lambda)$ with respect to λ gives

$$\begin{aligned} \frac{\partial L_{GMRSS(R=r)}(\lambda)}{\partial \lambda} &= \sum_{j=1}^{m}\sum_{\tau=1}^{k} \frac{f'(x_{\tau(r)j}, y_{\tau[r]j})}{f(x_{\tau(r)j}, y_{\tau[r]j})} \\ &= \sum_{j=1}^{m}\sum_{r=1}^{k} \frac{\left(1-2u_{\tau(r)j}\right)\left(1-2v_{\tau[r]j}\right)}{1+\lambda\left(1-2u_{\tau(r)j}\right)\left(1-2v_{\tau[r]j}\right)} = 0 \end{aligned} \tag{2.65}$$

where $u_{\tau(r)j} = F^{-1}(x_{\tau(r)j})$ and $v_{\tau[r]j} = G^{-1}(y_{\tau[r]j})$. By solving the Eq. (2.64) or (2.65), $\hat{\lambda}_{GMRSS(R=r)}$ can be found. The variance of $\hat{\lambda}_{GMRSS(R=r)}$ is defined as follows:

$$\begin{aligned} V\left(\hat{\lambda}_{GMRSS(R=r)}\right) &= \left\{-E\left[\frac{\partial^2 L_{GMRSS(R=r)}(\lambda)}{\partial\lambda^2}\right]\right\}^{-1} \\ &= \left\{E\left[\sum_{j=1}^{m}\sum_{\tau=1}^{k}\left(\frac{\left(1-2u_{\tau(r)j}\right)\left(1-2v_{\tau[r]j}\right)}{1+\lambda\left(1-2u_{\tau(r)j}\right)\left(1-2v_{\tau[r]j}\right)}\right)^2\right]\right\}^{-1} \end{aligned} \tag{2.66}$$

2.2.4 Simulation Results

In this section, a Monte Carlo simulation is presented. Using *Algorithm 2*, *Algorithm 3* and *Algorithm 7*, $10,000$ samples are generated. The values of dependence parameter, the sample sizes and the set sizes are taken to be $\lambda = \{-0.5, 0, 0.5\}$, $n = \{30, 60, 90\}$ and $k = \{3, 5\}$, respectively. Tables 2.4-2.6 provides values of $\hat{\lambda}_{\Im}$, RE and relative information (RI) where $\Im$ = SRS, RSS, $GMRSS(R = r)$ where $r \in [1, k]$. REs are calculated as follows:

$$RE = \frac{MSE\left(\hat{\lambda}_{SRS}\right)}{MSE\left(\hat{\lambda}_{RSS}\right)}, \quad RE_r = \frac{MSE\left(\hat{\lambda}_{SRS}\right)}{MSE\left(\hat{\lambda}_{GMRSS(R=r)}\right)}, \tag{2.67}$$

where

$$MSE(\hat{\lambda}_{\Im}) = \frac{1}{10,000}\sum_{w=1}^{10,000}\left(\hat{\lambda}_{\Im,w} - \lambda\right)^2, \tag{2.68}$$

for $\Im$=SRS, RSS and $GMRSS(R = r)$. Also, the RIs are expressed as follows:

$$RI = \frac{FI\left(\hat{\lambda}_{SRS}\right)}{FI\left(\hat{\lambda}_{RSS}\right)}, \quad RI_r = \frac{FI\left(\hat{\lambda}_{SRS}\right)}{FI\left(\hat{\lambda}_{GMRSS(R=r)}\right)}, \tag{2.69}$$

where FI is fisher information,

$$\begin{aligned} FI\left(\hat{\lambda}_{SRS}\right) &= -E\left[\frac{\partial^2 L_{SRS}(\lambda)}{\partial\lambda^2}\right], \quad FI\left(\hat{\lambda}_{RSS}\right) = -E\left[\frac{\partial^2 L_{RSS}(\lambda)}{\partial\lambda^2}\right] \\ \text{and} \quad FI\left(\hat{\lambda}_{GMRSS(R=r)}\right) &= -E\left[\frac{\partial^2 L_{GMRSS(R=r)}(\lambda)}{\partial\lambda^2}\right] \end{aligned} \tag{2.70}$$

We obtain the first and second derivatives of the log-likelihoods with respect to λ by using "optim" function in R programming language. In the R function, *method* is set to "L-BFGS-B". According to Table 2.4, it is seen that $\hat{\lambda}_{SRS} \approx \hat{\lambda}_{RSS}$. On the other hand,

Table 2.4 Estimated values ($\hat{\lambda}_{SRS}$, $\hat{\lambda}_{RSS}$) and relative efficiencies of $\hat{\lambda}_{RSS}$ with respect to $\hat{\lambda}_{SRS}$.

λ	n	k	$\hat{\lambda}_{SRS}$	$\hat{\lambda}_{RSS}$	RE	RI
-0.5	30	3	-0.458	-0.455	0.984	0.998
		5	-0.458	-0.449	0.981	1.007
	60	3	-0.485	-0.490	1.032	0.993
		5	-0.485	-0.489	1.016	1.007
	90	3	-0.498	-0.498	0.994	1.002
		5	-0.498	-0.497	1.037	0.999
0	30	3	0.002	-0.001	1.002	1.014
		5	0.002	0.000	1.001	0.999
	60	3	-0.004	-0.001	1.035	0.999
		5	-0.004	0.000	1.017	0.995
	90	3	-0.006	0.001	1.004	1.003
		5	-0.006	0.003	1.016	1.002
0.5	30	3	0.459	0.458	0.975	1.007
		5	0.459	0.456	0.976	0.991
	60	3	0.492	0.482	0.982	1.000
		5	0.492	0.486	0.987	0.992
	90	3	0.498	0.493	0.990	0.999
		5	0.498	0.496	0.997	1.000

it is observed that $\hat{\lambda}_{RSS}$ is as efficient as $\hat{\lambda}_{SRS}$. Despite the fact that these results come as surprise to researchers who study on RSS and its extensions, similar results can be seen in Stokes (1980) and Sevil & Yildiz (2022b). In Table 2.5, $GMRSS(R = 1)$

Table 2.5 Estimated values ($\hat{\lambda}_{R=r}$) and relative efficiencies of $\hat{\lambda}_{R=r}$ with respect to $\hat{\lambda}_{SRS}$ ($k = 3$).

λ	n	$\hat{\lambda}_{R=1}$	$\hat{\lambda}_{R=2}$	$\hat{\lambda}_{R=3}$	$RE_{R=1}$	$RI_{R=1}$	$RE_{R=2}$	$RI_{R=2}$	$RE_{R=3}$	$RI_{R=3}$
-0.5	30	-0.461	-0.417	-0.471	1.139	1.222	0.695	0.569	1.130	1.229
	60	-0.496	-0.463	-0.492	1.202	1.218	0.656	0.572	1.144	1.213
	90	-0.499	-0.483	-0.501	1.138	1.213	0.658	0.570	1.179	1.210
0	30	0.005	-0.001	-0.004	1.175	1.216	0.707	0.583	1.154	1.207
	60	0.002	-0.006	0.001	1.147	1.202	0.630	0.597	1.230	1.194
	90	-0.005	-0.004	-0.003	1.228	1.196	0.618	0.604	1.185	1.191
0.5	30	0.465	0.400	0.459	1.165	1.230	0.650	0.570	1.126	1.218
	60	0.492	0.467	0.488	1.177	1.216	0.660	0.574	1.193	1.223
	90	0.505	0.487	0.500	1.208	1.207	0.658	0.573	1.221	1.211

$\hat{\lambda}_{R=r}$ is $\hat{\lambda}_{GMRSS(R=r)}$ for $r \in [1, k]$

and $GMRSS(R = 3)$ provides smaller biases than the $GMRSS(R = 2)$. Also, the ML estimator based on $GMRSS(R = 1)$ and $GMRSS(R = 3)$ are more efficient than the ML estimators based on SRS, RSS and $GMRSS(R = 2)$. Moreover, there is no evidence that REs and RIs are not monotone increasing or decreasing function of n. Table 2.6 demonstrates that $GMRSS(R = 1)$ and $GMRSS(R = 5)$ provides unbiased and more efficient ML estimators for dependence parameter than SRS, RSS, $GMRSS(R = r)$ where $1 < r < k$. The REs and RIs decrease as the r is increasing

Table 2.6 Estimated values ($\hat{\lambda}_{R=r}$) and relative efficiencies of $\hat{\lambda}_{R=r}$ with respect to $\hat{\lambda}_{SRS}$ ($k = 5$).

λ	n	$\hat{\lambda}_{R=1}$	$\hat{\lambda}_{R=2}$	$\hat{\lambda}_{R=3}$	$\hat{\lambda}_{R=4}$	$\hat{\lambda}_{R=5}$	$RE_{R=1}$	$RI_{R=1}$	$RE_{R=2}$	$RI_{R=2}$	$RE_{R=3}$	$RI_{R=3}$	$RE_{R=4}$	$RI_{R=4}$	$RE_{R=5}$	$RI_{R=5}$
−0.5	30	−0.489	−0.432	−0.379	−0.426	−0.475	1.456	1.667	0.792	0.689	0.543	0.399	0.753	0.686	1.460	1.665
	60	−0.497	−0.481	−0.439	−0.477	−0.503	1.530	1.623	0.787	0.683	0.506	0.396	0.744	0.689	1.457	1.624
	90	−0.501	−0.488	−0.457	−0.490	−0.497	1.549	1.614	0.743	0.691	0.470	0.397	0.713	0.692	1.605	1.600
0	30	−0.005	0.002	−0.002	0.002	−0.005	1.424	1.571	0.792	0.695	0.602	0.406	0.801	0.703	1.451	1.581
	60	0.000	0.000	0.004	−0.004	−0.002	1.542	1.549	0.741	0.713	0.498	0.425	0.734	0.713	1.536	1.563
	90	−0.001	0.006	0.000	−0.007	−0.005	1.535	1.557	0.709	0.713	0.456	0.429	0.718	0.713	1.576	1.551
0.5	30	0.480	0.430	0.371	0.423	0.483	1.485	1.655	0.792	0.687	0.545	0.399	0.775	0.697	1.455	1.675
	60	0.496	0.476	0.439	0.475	0.496	1.497	1.630	0.775	0.684	0.523	0.398	0.724	0.690	1.503	1.610
	90	0.497	0.486	0.455	0.489	0.502	1.520	1.605	0.761	0.686	0.481	0.399	0.741	0.689	1.504	1.597

$\hat{\lambda}_{R=r}$ is $\hat{\lambda}_{GMRSS(R=r)}$ for $r \in [1, k]$

for $1 \leq r \leq (k+1)/2$. Also, it is seen that REs and RIs are increasing in r for $(k+1)/2 \leq r \leq k$.

CHAPTER THREE

STATISTICAL INFERENCE ON DEPENDENCE PARAMETER UNDER IMPERFECT RANKING

In RSS process, there are two potential ranking errors: one is brought on by ordering of X and the other by ordering of Y when ρ is low. The chapter's objective is to look into dependence parameter estimators under the case when the random variable X is ranked imperfectly. This is an important issue since the assumption of perfect ranking could not be realistic especially for environmental researches.

To provide our rationale, let us consider a scenario where (X, Y) represents an absolutely continuous random vector characterized by the joint CDF, $F(x, y)$, and PDF, $f(x, y)$. In the given literature on the estimation of dependence parameter, it is assumed that $X_{r(1)j} \leq X_{r(2)j} \leq \cdots \leq X_{r(k)j}$ from the rth set and the rth smallest X, say $X_{(r)j}$, is selected from the set where $r = 1, \cdots, k$ and $j = 1, \cdots, m$. On the other hand, the concomitant variable is denoted by $Y_{[r]j}$ since Ys are ranked by using Xs. Because the ordering of the Ys could be imperfect depending on the correlation coefficient between X and Y. Nevertheless, we now assume that Xs are ranked imperfectly. Under the imperfect ranking assumption, the pairs are $\left(X_{[r]j}, Y_{[r]j}\right)$ are selected from the rth set for the jth cycle. In this case, the ML estimator for dependence parameter performs poorly because the joint PDF of $\left(X_{[r]j}, Y_{[r]j}\right)$ deviates from $f_{r:k}(x, y)$. This issue has motivated us to develop a robust estimator.

In this chapter, we investigate MPL estimate of dependence parameter of bivariate normal distribution. An adaptation of ML methodology which is based on ranks was used to estimate dependence parameters by Genest & Favre (2007). This kind of methodology was first described by Genest et al. (1995) and then it was studied by Shih & Louis (1995). For estimating the dependence parameter of FGM copula, Ucer & Yildiz (2012) preferred the MPL methodology.

Let (X_i, Y_i), $i = 1, \cdots, n$, be random sample from a continuous bivariate distribution with CDF, $F(x, y)$, and PDF, $f(x, y)$. Using the transformations

$U = H(x)$ and $V = G(y)$, the copula form C is obtained. This form is given by Eq. (2.26). Suppose that $\hat{H}(x)$ and $\hat{G}(y)$ are EDFs,

$$\hat{H}(x) = \frac{1}{n}\sum_{i=1}^{n} I\left(X_i \leq x\right), \quad \text{and} \quad \hat{G}(y) = \frac{1}{n}\sum_{i=1}^{n} I\left(Y_i \leq y\right) \tag{3.1}$$

where $I\left(X_i \leq x\right)$ is an indicator function taking the value 1 if $X_i \leq x$ and 0 otherwise. The empirical log-likelihood function is defined as follows:

$$L\left(\rho\right) = \sum_{i=1}^{n} \log\left[c\left(\tilde{H}\left(X_i\right), \tilde{G}(Y_i)\right)\right] \tag{3.2}$$

where

$$\tilde{H}(x) = \frac{n}{n+1}\hat{H}(x), \quad \text{and} \quad \tilde{G}(y) = \frac{n}{n+1}\hat{G}(y). \tag{3.3}$$

Also, c is joint PDF of the random variables U and V. It is assumed that $\left(\Re_1, \mathcal{S}_1\right), \cdots, \left(\Re_n, \mathcal{S}_n\right)$ are the pairs of rank of (X, Y). The log-likelihood based on ranks can be expressed as following:

$$L\left(\rho\right) = \sum_{i=1}^{n} \log\left[c\left(\frac{\Re_i}{n+1}, \frac{\mathcal{S}_i}{n+1}\right)\right]. \tag{3.4}$$

In this form, $\tilde{H}(x) = \Re_i/(n+1)$ and $\tilde{G}(y) = \mathcal{S}_i/(n+1)$ for all $i \in 1, \cdots, n$.

It is proved that the following equation has a unique root $\hat{\rho}$ by Genest et al. (1995).

$$\frac{\partial L\left(\rho\right)}{\partial \rho} = \sum_{i=1}^{n} \frac{c'\left(\frac{\Re_i}{n+1}, \frac{\mathcal{S}_i}{n+1}\right)}{c\left(\frac{\Re_i}{n+1}, \frac{\mathcal{S}_i}{n+1}\right)} = 0 \tag{3.5}$$

where $c' = \partial c(u, v)/\partial \rho$. Also,

$$\hat{\rho} \approx N\left(\rho, \frac{v^2}{n}\right) \tag{3.6}$$

where $\hat{v}^2 = \hat{\delta}^2/\varphi^2$. Here, $\hat{\delta}^2$ and φ^2 are sample variances,

$$\hat{\delta}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(M_i - \bar{M}\right)^2 \tag{3.7}$$

and

$$\varphi^2 = \frac{1}{n}\sum_{i=1}^{n}\left(N_i - \bar{N}\right)^2 \tag{3.8}$$

where M and N are pseudo-observations with means $\hat{M}$ and $\hat{N}$, respectively. To obtain pseudo-observations, the procedure is followed.

1. We redeclare the initial data $(X_1, Y_1), \cdots, (X_n, Y_n)$ in a manner that $X_1 < \cdots < X_n$.

2. ℓ_ρ, ℓ_u, and ℓ_v are computed by using $\ell(\rho, u, v) = \log(c(u, v))$. Then N_i and M_i are obtained,

$$N_i = \ell\left(\hat{\rho}, \frac{i}{n+1}, \frac{\mathcal{S}_i}{n+1}\right) \tag{3.9}$$

and

$$\begin{aligned} M_i = N_i &- \frac{1}{n}\sum_{j=1}^{n}\ell_\rho\left(\hat{\rho}, \frac{j}{n+1}, \frac{\mathcal{S}_j}{n+1}\right)\ell_u\left(\hat{\rho}, \frac{j}{n+1}, \frac{\mathcal{S}_j}{n+1}\right) \\ &- \frac{1}{n}\sum_{\mathcal{S}_j \geq \mathcal{S}_i}^{n}\ell_\rho\left(\hat{\rho}, \frac{j}{n+1}, \frac{\mathcal{S}_j}{n+1}\right)\ell_v\left(\hat{\rho}, \frac{j}{n+1}, \frac{\mathcal{S}_j}{n+1}\right) \end{aligned} \tag{3.10}$$

3.1 Existing Correlation Coefficient Estimators based on RSS

Suppose that the pairs (X, Y) follows bivariate normal distribution with the parameters μ_x, μ_y, σ_x, σ_y and ρ. Joint PDF of (X, Y) is,

$$f(x, y; \mu_x, \mu_y, \sigma_x, \sigma_y, \rho) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left\{-d/2\right\}, \tag{3.11}$$

where

$$d = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right].$$

In the literature, some robust estimators based on RSS were investigated for dependence parameter (ρ). $RSS(k, m)$ is selected from a population having bivariate normal distribution. In this section, we introduce correlation coefficient estimators for dependence parameter of bivariate normal distribution.

First, Stokes (1980) investigated ML estimator based on RSS and MRSS for estimating the dependence parameter of bivariate normal distribution. Let $(X_{(r)j}, Y_{[r]j})$ be ranked set sample of size mk where $r = 1, \cdots, k$ and $j = 1, \cdots, m$. The joint PDF of rth ranked pairs is

$$\begin{aligned} f_{r:k}(x, y) = {} & \frac{k!}{(r-1)!\,(k-r)!} \Phi^{r-1}\left(\frac{x-\mu_x}{\sigma_x}\right) \left[1 - \Phi\left(\frac{x-\mu_x}{\sigma_x}\right)\right]^{k-r} \times \\ & f(x, y; \mu_x, \mu_y, \sigma_x, \sigma_y, \rho) \end{aligned} \tag{3.12}$$

where $\Phi(x)$ is the CDF of $X \sim N(0,1)$. The log-likelihood function was given as follows:

$$\begin{aligned} L_{RSS_{ML}}(\rho) = {} & \sum_{j=1}^{m}\sum_{r=1}^{k} \log\left(\frac{k!}{(r-1)!\,(k-r)!}\right) + \\ & (r-1)\log\left(\Phi\left(\frac{x_{(r)j}-\mu_x}{\sigma_x}\right)\right) + \\ & (k-r)\log\left[1-\Phi\left(\frac{x_{(r)j}-\mu_x}{\sigma_x}\right)\right] + \\ & \log\left(f\left(x_{(r)j}, y_{[r]j}; \mu_x, \mu_y, \sigma_x, \sigma_y, \rho\right)\right) \end{aligned} \tag{3.13}$$

By taking the first derivative of the Eq. (3.13), the following cubic equation is obtained.

$$\begin{aligned} \frac{\partial L_{RSS_{ML}}(\rho)}{\partial \rho} = {} & mk\sigma_x^2\sigma_y^2\rho\left(1-\rho^2\right) + \\ & \sigma_x\sigma_y\left(1+\rho^2\right)\sum_{j=1}^{m}\sum_{r=1}^{k}\left(x_{(r)j}-\mu_x\right)\left(y_{[r]j}-\mu_y\right) - \\ & \rho\left[\sigma_y^2\sum_{j=1}^{m}\sum_{r=1}^{k}\left(x_{(r)j}-\mu_x\right)^2 + \sigma_x^2\sum_{j=1}^{m}\sum_{r=1}^{k}\left(y_{[r]j}-\mu_y\right)^2\right] = 0 \end{aligned} \tag{3.14}$$

The solution of the cubic equation provides the value of the ML estimator based on RSS for dependence parameter. Under case (i), Stokes (1980) showed that the ML estimator based on RSS is as efficient as its counterpart in SRS. Therefore, Stokes (1980) defined MRSS procedure that requires only minimum or maximum ranked pairs. Then the author showed that the ML estimator based on MRSS is more efficient than the ML estimators based on SRS and RSS. After that, case (ii) was investigated. First, μ_y and σ_y^2 need to be estimated in this case. Stokes (1980) used the RSS counterparts of the estimators suggested by Barnett et al. (1976) for that. The author proposed following estimators

$$\hat{\mu}_y = \bar{Y} - \mathscr{R} S_y \left(\bar{X} - \hat{\mu}_x\right)/S_x, \quad \text{and} \quad \hat{\sigma}_y^2 = S_y^2 \left(1 - \mathscr{R}^2 + \mathscr{R}^2 \sigma_x^2 / S_x^2\right)^{-1/2}, \tag{3.15}$$

and

$$\hat{\rho}_{RSS} = \left(\mathscr{R} \sigma_x / S_x\right) \left(1 - \mathscr{R}^2 + \mathscr{R}^2 \sigma_x^2 / S_x^2\right)^{-1/2} \tag{3.16}$$

where

$$S_x^2 = \Sigma_{j=1}^m \Sigma_{r=1}^k \left(x_{(r)j} - \bar{X}\right)^2 / mk, \quad S_y^2 = \Sigma_{j=1}^m \Sigma_{r=1}^k \left(y_{[r]j} - \bar{Y}\right)^2 / mk,$$

and

$$\mathscr{R} S_x S_y = \Sigma_{j=1}^m \Sigma_{r=1}^k \left(x_{(r)j} - \bar{X}\right) \left(y_{[r]j} - \bar{Y}\right) / mk.$$

Furthermore, $\bar{X} = \Sigma_{j=1}^m \Sigma_{r=1}^k x_{(r)j}/mk$ and $\bar{Y} = \Sigma_{j=1}^m \Sigma_{r=1}^k y_{[r]j}/mk$. Stokes (1980) proved that $\hat{\rho}_{RSS}$ is more efficient than its counterpart in SRS under case (ii). Also, Stokes (1980) studied the MLE of ρ under case (iii) and the author noted that the closed form of the ML estimator of ρ based on RSS is not present even when the ranking is perfect.

Zheng & Modarres (2006) suggested the following sample correlation coefficient.

$$\hat{\rho}_{ZM} = \frac{\sum_{j=1}^{m} \sum_{r=1}^{k} \left(x_{(r)j} - \bar{X}\right) \left(y_{[r]j} - \bar{Y}\right)}{\left[\sum_{j=1}^{m} \sum_{r=1}^{k} \left(x_{(r)j} - \bar{X}\right)^2 \sum_{j=1}^{m} \sum_{r=1}^{k} \left(y_{[r]j} - \bar{Y}\right)^2\right]^{1/2}} \tag{3.17}$$

The authors stated that the estimator is obtained by using partial likelihood equation,

$$\sum_{j=1}^{m} \sum_{r=1}^{k} \frac{\partial \log f\left(x_{(r)j}, y_{[r]j}; \mu_x, \mu_y, \sigma_x, \sigma_y, \rho\right)}{\partial \rho} = 0 \tag{3.18}$$

Thus, $\hat{\rho}_{ZM}$ was called as a modified maximum likelihood (MML) estimate in this study.

It should be noted that MML was proposed by Mehrotra & Nanda (1974). In their work, $\hat{\rho}_{ZM}$ was investigated when all parameters are unknown (case (iii)). They made comparison between $\hat{\rho}_{ZM}$ and its counterpart $\hat{\rho}_{SRS}$ where

$$\hat{\rho}_{SRS} = \frac{\sum_{i=1}^{n} \left(x_i - \bar{X}\right)\left(y_i - \bar{Y}\right)}{\left[\sum_{i=1}^{n} \left(x_i - \bar{X}\right)^2 \sum_{i=1}^{n} \left(y_i - \bar{Y}\right)^2\right]^{1/2}} \tag{3.19}$$

The authors obtained that $\hat{\rho}_{ZM}$ has smaller variance than $\hat{\rho}_{SRS}$ under either perfect or imperfect ranking. On the other hand, the REs of $\hat{\rho}_{ZM}$ with respect to $\hat{\rho}_{SRS}$ showed that $\hat{\rho}_{ZM}$ should be used when the correlation between X and Y is strong ($\rho \geq 0.75$) and $k \geq 6$. Also, the authors proved that the Eq. (3.17) is a robust estimator against imperfect ranking.

Hui et al. (2009) developed pseudo ML (PML) estimates for dependence parameter of bivariate normal distribution under the case when μ_x and σ_x^2 are known (case (ii)). Without loss of generality, it is assumed that $\mu_x = 0$ and $\sigma_x^2 = 1$. The authors used two consistent estimators,

$$\bar{Y} = \frac{1}{mk} \sum_{j=1}^{m} \sum_{r=1}^{k} y_{[r]j}, \tag{3.20}$$

and

$$\hat{\sigma}_y^2 = \frac{1}{mk} \sum_{j=1}^{m} \sum_{r=1}^{k} \left(y_{[r]j} - \bar{Y}\right)^2 \tag{3.21}$$

to estimate μ_y and σ_y^2. Then, the pseudo log-likelihood function is

$$\begin{aligned} L_{RSS_{PML}}(\rho) &= \sum_{j=1}^{m} \sum_{r=1}^{k} \log(f_{r:k}\left(x_{(r)j}, y_{[r]j}\right)) \\ &= \sum_{j=1}^{m} \sum_{r=1}^{k} (r-1) \log\left(\Phi\left(x_{(r)j}\right)\right) + \\ &\quad \sum_{j=1}^{m} \sum_{r=1}^{k} (k-r) \log\left(1 - \Phi\left(x_{(r)j}\right)\right) + \\ &\quad \sum_{j=1}^{m} \sum_{r=1}^{k} \log\left(f\left(x_{(r)j}, y_{[r]j}; \bar{Y}, \hat{\sigma}_y^2, \rho\right)\right) \end{aligned} \tag{3.22}$$

where $\Phi(x)$ is the CDF of $X \sim N(0,1)$. Thus,

$$\begin{aligned}\frac{\partial L_{RSS_{PML}}(\rho)}{\partial\rho} = \rho^3 - \frac{\rho^2}{mk}\sum_{j=1}^{m}\sum_{r=1}^{k}\left(x_{(r)j}\frac{y_{[r]j}-\bar{Y}}{\hat{\sigma}_y}\right) - \\ \rho\left\{1-\frac{1}{mk}\sum_{j=1}^{m}\sum_{r=1}^{k}\left(x_{(r)j}^2+\frac{\left(y_{[r]j}-\bar{Y}\right)^2}{\hat{\sigma}_y^2}\right)\right\} - \\ \frac{1}{mk}\sum_{j=1}^{m}\sum_{r=1}^{k}\left(x_{(r)j}\frac{y_{[r]j}-\bar{Y}}{\hat{\sigma}_y}\right) = 0\end{aligned} \tag{3.23}$$

Using a numerical algorithm, (3.22) or (3.23) can be solved to obtain value of PML estimator ($\hat{\rho}_{HMZ}$). Then the authors gave numerical comparisons among the estimators $\hat{\rho}_{RSS}$, $\hat{\rho}_{ZM}$, $\hat{\rho}_{SRS}$, $\hat{\rho}_{HMZ}$. In the simulation study, samples are selected from bivariate normal and contaminated bivariate normal distributions. To obtain contaminated distribution, samples are generated from $f(x, y; 0, 0, 1, 1, \rho)$ with 90% probability and $f(x, y; 0, 0, 9, 9, \rho)$ with %10 probability. According to their results, $\hat{\rho}_{HMZ}$ is more efficient than $\hat{\rho}_{SRS}$ and $\hat{\rho}_{ZM}$ for larger ρ, while being slightly less efficient when $\rho \to 0$. Moreover, $\hat{\rho}_{HMZ}$ is less efficient than $\hat{\rho}_{RSS}$ when ρ is large while being relatively more efficient when $\rho \to 0$. For contaminated bivariate normal distribution, $\hat{\rho}_{HMZ}$ is more efficient than the other three estimators for small ρ but less efficient for large ρ.

3.2 Maximum Pseudo Likelihood Estimates

3.2.1 MPL Estimator from Simple Random Sample

Let (X_i, Y_i), $i \cdots, n$, be a simple random sample that follows bivariate normal distribution function and the copula form of the joint density in the Eq. (3.11) is,

$$\begin{aligned}c(u,v) = \frac{1}{\sqrt{(1-\rho^2)}}\exp\left[-\frac{\rho^2}{2(1-\rho^2)}\left\{\left(\Phi^{-1}(u)\right)^2+\left(\Phi^{-1}(v)\right)^2\right\}+\right. \\ \left.\frac{\rho}{1-\rho^2}\Phi^{-1}(u)\,\Phi^{-1}(v)\right]\end{aligned} \tag{3.24}$$

where $\rho \in (0,1)$, $u = H(x)$, $v = G(y)$ and Φ^{-1} is inverse CDF, see Balakrishnan & Lai (2009). The empirical likelihood function is

$$\mathscr{L}_{SRS_{MPL}} = \left[\frac{1}{\sqrt{(1-\rho^2)}}\right]^n \exp\left[-\frac{\rho^2}{2(1-\rho^2)}\sum_{i=1}^{n}\left\{\left(\Phi^{-1}(u_i)\right)^2 + \left(\Phi^{-1}(v_i)\right)^2\right\} + \frac{\rho^2}{1-\rho^2}\sum_{i=1}^{n}\Phi^{-1}(u_i)\,\Phi^{-1}(v_i)\right] \tag{3.25}$$

where $u_i = \tilde{H}(x_i)$ and $v_i = \tilde{G}(x_i)$ that are calculated by using Eq. (3.3) for $i = 1, \cdots, n$. The empirical log-likelihood function is defined as following:

$$L_{SRS_{MPL}}(\rho) = \frac{-n}{2}\log\left(1-\rho^2\right) - \frac{\rho^2}{2(1-\rho^2)}\sum_{i=1}^{n}\left\{\left(\Phi^{-1}(u_i)\right)^2 + \left(\Phi^{-1}(v_i)\right)^2\right\} + \frac{\rho}{1-\rho^2}\sum_{i=1}^{n}\Phi^{-1}(u_i)\,\Phi^{-1}(v_i) \tag{3.26}$$

By taking the first derivative of $L_{SRS_{MPL}}(\rho)$ with respect to ρ,

$$\frac{\partial L_{SRS_{MPL}}(\rho)}{\partial \rho} = \frac{n\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2}\sum_{i=1}^{n}\left\{\left(\Phi^{-1}(u_i)\right)^2 + \left(\Phi^{-1}(v_i)\right)^2\right\} + \frac{1+\rho^2}{1-\rho^2}\sum_{i=1}^{n}\Phi^{-1}(u_i)\,\Phi^{-1}(v_i) \tag{3.27}$$

3.2.2 MPL Estimator from Ranked Set Sample

We suppose that $(X_{(r)j}, Y_{[r]j})$ are the pairs in $RSS(k,m)$ for $r = 1, \cdots, k$ and $j = 1, \cdots, m$. The joint density of $(X_{(r)j}, Y_{[r]j})$ is given as

$$f_{r:k}(u,v) = \frac{k!}{(r-1)!\,(k-r)!}u^{r-1}\left[1-u\right]^{k-r}c(u,v) \tag{3.28}$$

where $u = H(x)$, $v = G(y)$ and $c(u, v)$ is given by Eq. (3.24). The empirical log-likelihood function for RSS is

$$
\begin{aligned}
L_{RSS_{MPL}}(\rho) = & \sum_{j=1}^{m}\sum_{r=1}^{k} \log\left(\frac{k!}{(r-1)!\,(k-r)!}\right) + \\
& \sum_{j=1}^{m}\sum_{r=1}^{k} (r-1)\log\left(u_{(r)j}\right) + \\
& \sum_{j=1}^{m}\sum_{r=1}^{k} (k-r)\log\left(1-u_{(r)j}\right) - \frac{mk}{2}\log\left(1-\rho^2\right) - \\
& \frac{\rho^2}{2\left(1-\rho^2\right)}\sum_{j=1}^{m}\sum_{r=1}^{k}\left\{\left(\Phi^{-1}\left(u_{(r)j}\right)\right)^2 + \left(\Phi^{-1}\left(v_{[r]j}\right)\right)^2\right\} + \\
& \frac{\rho}{\left(1-\rho^2\right)}\sum_{j=1}^{m}\sum_{r=1}^{k}\Phi^{-1}\left(u_{(r)j}\right)\Phi^{-1}\left(v_{[r]j}\right)
\end{aligned}
\tag{3.29}
$$

where

$$
\begin{aligned}
u_{(r)j} &= \tilde{H}^*\left(x_{(r)j}\right) = \frac{mk}{mk+1}\hat{H}^*\left(x_{(r)j}\right) \\
v_{[r]j} &= \tilde{G}^*\left(y_{[r]j}\right) = \frac{mk}{mk+1}\hat{G}^*\left(y_{[r]j}\right).
\end{aligned}
\tag{3.30}
$$

For estimating the distribution function, Stokes & Sager (1988) proposed EDF based on RSS that is presented as following:

$$
\hat{H}^*(x) = \frac{1}{mk}\sum_{j=1}^{m}\sum_{r=1}^{k} I\left(X_{(r)j} \le x\right), \quad \text{and} \quad \hat{G}^*(y) = \frac{1}{mk}\sum_{j=1}^{m}\sum_{r=1}^{k} I\left(Y_{[r]j} \le y\right).
\tag{3.31}
$$

Both $\hat{H}^*(x)$ and $\hat{G}^*(y)$ are unbiased for $H(x)$ and $G(y)$. Also, Stokes & Sager (1988) showed that variance of the EDF based on RSS is smaller than or equal to variance of the EDF based on SRS. Thus, the following equation is obtained by taking the first derivative of $L_{RSS_{MPL}}(\rho)$ with respect to ρ,

$$\frac{\partial L_{RSS_{MPL}}(\rho)}{\partial\rho} = \frac{mk\rho}{1-\rho^2} - \frac{\rho}{\left(1-\rho^2\right)^2}\sum_{j=1}^{m}\sum_{r=1}^{k}\left\{\left(\Phi^{-1}\left(u_{(r)j}\right)\right)^2 + \left(\Phi^{-1}\left(v_{[r]j}\right)\right)^2\right\} + \frac{1+\rho^2}{1-\rho^2}\sum_{j=1}^{m}\sum_{r=1}^{k}\Phi^{-1}\left(u_{(r)j}\right)\Phi^{-1}\left(v_{[r]j}\right) = 0 \tag{3.32}$$

The MPL estimator for RSS is denoted by $\hat{\rho}_{YT}$. Note that Y and T are the initials of the authors. Estimated value of $\hat{\rho}_{YT}$ is obtained by using a numerical algorithm for solving the Eq. (3.29) in the Monte Carlo simulation.

3.3 Simulation Results

In this section, the developed estimator $\hat{\rho}_{YT}$ compare with three other estimators of the correlation coefficient that are $\hat{\rho}_{SRS}$, $\hat{\rho}_{ZM}$, and $\hat{\rho}_{HMZ}$. We consider case (iii) which are μ_x, μ_y, σ_x, and σ_y are unknown. However, $\hat{\rho}_{HMZ}$ was investigated under case (ii) by Hui et al. (2009). Therefore, the following equation is used instead of Eq. (3.22) to obtain $\hat{\rho}_{HMZ}$ under the case (iii).

$$\begin{aligned} L_{RSS_{PML}}(\rho) &= \sum_{j=1}^{m}\sum_{r=1}^{k}\log(f_{r:k}\left(x_{(r)j}, y_{[r]j}\right)) \\ &= \sum_{j=1}^{m}\sum_{r=1}^{k}(r-1)\log\left(\Phi\left(x;\bar{X},\hat{\sigma}_x^2\right)\right) + \\ &\quad \sum_{j=1}^{m}\sum_{r=1}^{k}(k-r)\log\left(1-\Phi\left(x;\bar{X},\hat{\sigma}_x^2\right)\right) + \\ &\quad \sum_{j=1}^{m}\sum_{r=1}^{k}\log\left(f\left(x_{(r)j}, y_{[r]j};\bar{X},\bar{Y},\hat{\sigma}_x^2,\hat{\sigma}_y^2,\rho\right)\right) \end{aligned} \tag{3.33}$$

where the consistent estimators are used

$$\bar{X} = \sum_{j=1}^{m}\sum_{r=1}^{k}X_{(r)j}/mk, \quad \text{and} \quad \hat{\sigma}_x^2 = \sum_{j=1}^{m}\sum_{r=1}^{k}\left(X_{(r)j}-\bar{X}\right)^2/mk \tag{3.34}$$

for estimating the μ_y and σ_x.

In the simulation, we consider bivariate normal distribution with parameters $\mu_x = 0$, $\mu_y = 0$, $\sigma_x = 1$, $\sigma_y = 1$, $\rho = 0.1, 0.5, 0.9$, $k = 3, 4, 5, 6, 7, 8$, and $m = 5, 10, 15$. The results are obtained based on $10,000$ trials. For the imperfect ranking case, we used the following imperfect ranking model,

$$F_i = X_i + \Delta_i \tag{3.35}$$

where $\Delta_i \sim N(0, \sigma_\Delta^2)$ for $i = 1, \cdots, k$. The imperfect ranking model is known as visual ranking model that is suggested by Dell & Clutter (1972). If $\sigma_\Delta^2 = 0$, ranking is performed perfectly, otherwise, it is performed imperfetcly. The following equation was defined by Nahhas et al. (2002).

$$Corr\left(F, X\right) = \frac{\sigma_x^2}{\sqrt{\sigma_x^2 + \sigma_\Delta^2}} \tag{3.36}$$

In the simulation, we suppose that $\sigma_\Delta^2 = 0.778$, so $Corr\left(F, X\right) = 0.75$.

Algorithm 2 can be used to generate ranked set samples, but in this case, (X_i, Y_i) are constructed using bivariate normal distribution, $i = 1, \cdots, k$. For imperfect ranking, Xs are first ranked using F-values, and then Ys are ordered using ranks of Xs.

In this simulation, estimated values and REs of $\hat{\rho}_{YT}$ with respect to $\hat{\rho}_{SRS}$, $\hat{\rho}_{ZM}$, and $\hat{\rho}_{HMZ}$ are obtained. For the estimated values and MSEs, the following equations are used.

$$\hat{\rho}_h = \frac{1}{10,000} \sum_{i=1}^{10,000} \hat{\rho}_{h,i}, \quad \text{and} \quad MSE\left(\hat{\rho}_h\right) = \frac{1}{10,000} \sum_{i=1}^{10,000} \left(\hat{\rho}_{h,i} - \rho\right)^2, \tag{3.37}$$

where $h = SRS, ZM, HMZ$, and YT. The REs are computed by using the Eq. (3.32).

$$RE\left(\hat{\rho}_{SRS}, \hat{\rho}_{h'}\right) = \frac{MSE\left(\hat{\rho}_{SRS}\right)}{MSE\left(\hat{\rho}_{h'}\right)}, \tag{3.38}$$

where $h' = ZM, HMZ$, and YT. The results are presented in the Table 3.1. In this table, we observe that $\hat{\rho}_{YT}$ has larger biases than the other estimators especially when ρ is small. However, these biases of $\hat{\rho}_{YT}$ reduce as $n = mk$ increases. On the other hand, it is seen that the bias values do not vary depending on the ranking quality. Also, we can say that $\hat{\rho}_{SRS}$, $\hat{\rho}_{ZM}$, and $\hat{\rho}_{HMZ}$ are unbiased estimators for ρ under perfect and imperfect ranking. According to the REs in Table 3.1, $\hat{\rho}_{YT}$ is more efficient than $\hat{\rho}_{SRS}$ and $\hat{\rho}_{ZM}$. Also, $\hat{\rho}_{YT}$ is less efficient than $\hat{\rho}_{HMZ}$ when $\rho \to 0$ but more efficient for larger ρ. Additionally, it is noted that REs do not increase or decrease monotonically while $n = mk$ increases. However, there is an evidence that REs increase monotonically as $\rho \to 1$. When $\rho \geq 0.5$, it should be noted that the REs under perfect ranking are mostly higher than those under imperfect ranking. In other words, if the Xs are not perfectly ranked, $\hat{\rho}_{ZM}$ and $\hat{\rho}_{HMZ}$ loss efficiency even if Y is highly correlated with X. Thus, it appears that the proposed estimator is robust to the imperfect ranking case. In particular, $RE\left(\hat{\rho}_{ZM}, \hat{\rho}_{YT}\right)$, and $RE\left(\hat{\rho}_{HMZ}, \hat{\rho}_{YT}\right)$ are similar when $\rho \geq 0.5$ and $n = mk \geq 25$.

Table 3.1 The estimated values ($\hat{\rho}_h$) for $h = SRS, ZM, HMZ$, and YT and relative efficiencies of $\hat{\rho}_{YT}$ with respect to $\hat{\rho}_{SRS}$, $\hat{\rho}_{ZM}$, and $\hat{\rho}_{HMZ}$

				Perfect ranking						Perfect ranking					
				Yes			No			Yes			No		
ρ	m	k	$\hat{\rho}_{SRS}$	$\hat{\rho}_{ZM}$	$\hat{\rho}_{HMZ}$	$\hat{\rho}_{YT}$	$\hat{\rho}_{ZM}$	$\hat{\rho}_{HMZ}$	$\hat{\rho}_{YT}$	$RE1$	$RE2$	$RE3$	$RE1$	$RE2$	$RE3$
0.1	5	3	0.099	0.098	0.098	0.232	0.095	0.095	0.228	1.008	**1.916**	1.027	1.004	**2.027**	1.078
		4	0.098	0.099	0.099	0.202	0.099	0.099	0.205	1.009	**1.918**	1.077	1.039	**2.007**	1.117
		5	0.096	0.103	0.103	0.186	0.100	0.100	0.184	1.029	**1.906**	1.132	1.036	**1.912**	1.136
		6	0.099	0.097	0.097	0.170	0.099	0.099	0.171	0.980	**1.745**	1.126	1.003	**1.766**	1.101
		7	0.098	0.099	0.099	0.159	0.100	0.100	0.162	0.973	**1.730**	1.128	1.000	**1.773**	1.166
		8	0.098	0.099	0.099	0.152	0.102	0.102	0.155	1.017	**1.768**	1.194	1.009	**1.702**	1.153
	10	3	0.101	0.100	0.100	0.171	0.098	0.098	0.170	0.970	**1.760**	1.102	1.018	**1.809**	1.128
		4	0.101	0.101	0.101	0.154	0.100	0.100	0.156	1.021	**1.782**	1.212	0.999	**1.748**	1.180
		5	0.100	0.101	0.101	0.143	0.097	0.097	0.141	1.005	**1.699**	1.218	1.011	**1.686**	1.207
		6	0.100	0.098	0.098	0.134	0.099	0.099	0.135	1.009	**1.620**	1.203	1.014	**1.623**	1.212
		7	0.100	0.099	0.099	0.129	0.100	0.100	0.130	0.996	**1.606**	1.261	0.992	**1.573**	1.212
		8	0.099	0.100	0.100	0.126	0.102	0.102	0.129	1.013	**1.556**	1.238	1.012	**1.537**	1.219
	15	3	0.101	0.097	0.097	0.145	0.101	0.101	0.150	0.984	**1.715**	1.178	1.043	**1.737**	1.195
		4	0.097	0.102	0.102	0.137	0.099	0.099	0.134	1.052	**1.673**	1.246	0.997	**1.614**	1.201
		5	0.100	0.098	0.098	0.126	0.096	0.096	0.125	0.999	**1.548**	1.196	1.013	**1.569**	1.219
		6	0.101	0.100	0.100	0.123	0.099	0.099	0.123	1.016	**1.513**	1.222	1.014	**1.491**	1.218
		7	0.099	0.099	0.099	0.119	0.101	0.101	0.120	0.986	**1.431**	1.196	1.006	**1.473**	1.210
		8	0.098	0.099	0.099	0.115	0.100	0.100	0.117	0.999	**1.408**	1.203	1.002	**1.419**	1.200
0.5	5	3	0.486	0.496	0.499	0.576	0.491	0.494	0.572	1.037	1.125	**1.151**	1.006	1.096	**1.103**
		4	0.492	0.497	0.498	0.561	0.495	0.496	0.561	1.030	1.072	**1.120**	1.020	1.075	**1.131**
		5	0.495	0.497	0.498	0.552	0.495	0.496	0.552	1.042	1.056	**1.118**	1.067	1.084	**1.176**
		6	0.493	0.501	0.501	0.547	0.500	0.500	0.548	1.057	1.061	**1.126**	0.993	1.003	**1.062**
		7	0.492	0.501	0.501	0.541	0.494	0.494	0.539	1.034	1.036	**1.113**	0.996	1.000	**1.110**
		8	0.495	0.500	0.500	0.537	0.498	0.498	0.537	0.986	0.990	**1.073**	1.043	1.045	**1.154**
	10	3	0.493	0.496	0.497	0.546	0.496	0.496	0.546	1.057	1.063	**1.150**	0.964	0.974	**1.063**
		4	0.493	0.498	0.498	0.536	0.498	0.498	0.538	1.042	1.043	**1.135**	1.037	1.040	**1.126**
		5	0.496	0.500	0.500	0.533	0.496	0.496	0.530	1.074	1.074	**1.163**	0.992	0.992	**1.115**
		6	0.497	0.499	0.499	0.526	0.497	0.497	0.527	1.039	1.039	**1.136**	1.029	1.029	**1.130**
		7	0.497	0.500	0.500	0.524	0.497	0.497	0.524	1.066	1.066	**1.166**	0.997	0.997	**1.126**
		8	0.497	0.501	0.501	0.523	0.499	0.499	0.523	1.051	1.051	**1.153**	0.989	0.989	**1.115**
	15	3	0.494	0.497	0.497	0.534	0.496	0.496	0.533	1.013	1.014	**1.137**	1.008	1.010	**1.129**
		4	0.496	0.499	0.499	0.527	0.499	0.499	0.528	1.075	1.075	**1.185**	1.024	1.024	**1.159**
		5	0.496	0.500	0.500	0.524	0.498	0.498	0.522	1.046	1.046	**1.156**	0.987	0.987	**1.116**
		6	0.497	0.499	0.499	0.520	0.499	0.499	0.520	1.046	1.046	**1.143**	1.008	1.008	**1.141**
		7	0.498	0.500	0.500	0.518	0.500	0.500	0.518	1.054	1.054	**1.172**	0.986	0.986	**1.142**
		8	0.499	0.501	0.501	0.517	0.499	0.499	0.516	1.054	1.054	**1.175**	1.033	1.033	**1.186**
0.9	5	3	0.893	0.898	0.898	0.905	0.895	0.895	0.904	1.075	1.075	**1.964**	1.067	1.067	**2.026**
		4	0.894	0.899	0.899	0.903	0.897	0.897	0.903	1.194	1.194	**1.924**	1.048	1.049	**1.847**
		5	0.896	0.900	0.900	0.903	0.898	0.898	0.901	1.264	1.264	**1.870**	1.050	1.050	**1.722**
		6	0.897	0.901	0.901	0.903	0.899	0.899	0.901	1.309	1.310	**1.793**	1.062	1.062	**1.688**
		7	0.898	0.901	0.901	0.903	0.899	0.899	0.901	1.242	1.243	**1.707**	1.093	1.093	**1.726**
		8	0.897	0.901	0.901	0.903	0.899	0.899	0.902	1.298	1.298	**1.752**	1.070	1.070	**1.644**
	10	3	0.896	0.899	0.899	0.902	0.898	0.898	0.902	1.129	1.129	**1.744**	0.984	0.984	**1.715**
		4	0.898	0.899	0.899	0.901	0.898	0.898	0.901	1.144	1.145	**1.716**	1.023	1.023	**1.672**
		5	0.899	0.900	0.900	0.901	0.899	0.899	0.901	1.169	1.169	**1.710**	1.047	1.047	**1.674**
		6	0.899	0.900	0.900	0.902	0.900	0.900	0.901	1.152	1.152	**1.719**	1.034	1.034	**1.672**
		7	0.899	0.900	0.900	0.901	0.900	0.900	0.901	1.198	1.198	**1.724**	1.048	1.048	**1.698**
		8	0.899	0.900	0.900	0.901	0.900	0.900	0.901	1.252	1.253	**1.729**	1.051	1.051	**1.654**
	15	3	0.898	0.899	0.899	0.901	0.898	0.898	0.901	1.048	1.048	**1.607**	1.048	1.048	**1.716**
		4	0.898	0.900	0.900	0.901	0.899	0.899	0.901	1.106	1.106	**1.680**	1.018	1.018	**1.603**
		5	0.899	0.900	0.900	0.901	0.900	0.900	0.901	1.145	1.146	**1.741**	1.059	1.059	**1.694**
		6	0.899	0.900	0.900	0.901	0.900	0.900	0.901	1.190	1.190	**1.757**	1.069	1.069	**1.730**
		7	0.899	0.900	0.900	0.901	0.900	0.900	0.901	1.126	1.126	**1.660**	1.050	1.050	**1.713**
		8	0.899	0.900	0.900	0.901	0.900	0.900	0.901	1.184	1.184	**1.700**	1.029	1.029	**1.664**

$RE1 : RE(\hat{\rho}_{SRS}, \hat{\rho}_{ZM})$, $RE2 : RE(\hat{\rho}_{SRS}, \hat{\rho}_{HMZ})$, $RE3 : RE(\hat{\rho}_{SRS}, \hat{\rho}_{YT})$. Bold faced values shows the largest REs in the rows for perfect and imperfect ranking.

CHAPTER FOUR

CONCOMITANT BASED BOOTSTRAP TECHNIQUES

Modarres & Zheng (2004) conducted a study on constructing confidence intervals for the dependence parameter of the bivariate normal distribution using parametric bootstrap. Surprisingly, they found that all coverage probabilities of the bootstrap confidence intervals were 1, suggesting that further research was needed on the bootstrap method for the dependence parameter. In light of that, this section aims to address this issue and proposes non-parametric bootstrap methods based on concomitants. In this section, different nonparametric bootstrap techniques are introduced.

4.1 Bootstrap Techniques for Univariate RSS

Suppose that $RSS_1(k,m)$ is an original sample which is selected from a population. Let the following table be transpose of the matrix $RSS_1(k,m)$. By using the techniques, a bootstrap sample

$$\left\{Y^*_{(1)1},\cdots,Y^*_{(1)m}\right\},\cdots,\left\{Y^*_{(k)1},\cdots,Y^*_{(k)m}\right\} \tag{4.1}$$

are obtained. Chen et al. (2004) defined the following procedure that is called bootstrap

Table 4.1 Ranked set sample of size mk

	Cycles			
Judgement ranks	1	2	$\cdots$	m
1	$Y_{[1]1}$	$Y_{[1]2}$	$\cdots$	$Y_{[1]m}$
2	$Y_{[2]1}$	$Y_{[2]2}$	$\cdots$	$Y_{[2]m}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
k	$Y_{[k]1}$	$Y_{[k]2}$	$\cdots$	$Y_{[k]m}$

RSS by row (BRSSR).

I: A sample of size k is drawn with replacement from rth row of the Table 4.1.

II: The Step I is repeated for each $r = 1, \cdots, k$.

This procedure provides m units from each stratum. The other two methods were suggested by Modarres et al. (2006). These methods were named bootstrap RSS (BRSS) and mixed row bootstrap RSS (MRBRSS). The BRSS process is described below.

I: A sample of size k^2 is drawn with replacement from the whole of Table 4.1.

II: These units are split into k sets of k units each.

III: The units in each set are ranked in ascending order.

IV: From the rth set, the rth ranked unit is selected and is represented by $Y^*_{(r)}$ for $r = 1 \cdots, k$.

V: The first four steps are repeated m times and the bootstrap sample in Eq (4.1) is obtained.

MRBRSS combined the BRSSR and BRSS.

I: A unit is randomly selected from each row of Table 4.1.

II: The k units are ranked from the smallest to the largest. The smallest unit, let's say $Y^*_{(1)}$, is then selected.

III: Steps I and II are repeated k times. At the rth repetition of the process, $Y^*_{(r)}$ is selected.

IV: The first three steps are repeated m times to obtain the bootstrap sample in Eq (4.1).

For these bootstrap techniques, a literature review is provided by Chatterjee & Ghosh (2022).

4.2 Concomitant based Non-parametric Bootstrap Techniques

In this section, we provide some non-parametric bootstrap techniques based on BRSSR, BRSS, and MRBRSS. Suppose that $RSS(k,m)$ is selected from a population and the transpose of the matrix $RSS(k,m)$ is given by the following table. Let $\hat{\varrho}$ is the sample correlation coefficient that is obtained by using pairs $\left(X_{(r)j}, Y_{[r]j}\right)$

Table 4.2 Ranked set sample of size mk

	Cycles			
Judgement ranks	1	2	$\cdots$	m
1	$\left(X_{(1)1}, Y_{[1]1}\right)$	$\left(X_{(1)2}, Y_{[1]2}\right)$	$\cdots$	$\left(X_{(1)m}, Y_{[1]m}\right)$
2	$\left(X_{(2)1}, Y_{[2]1}\right)$	$\left(X_{(2)2}, Y_{[2]2}\right)$	$\cdots$	$\left(X_{(2)m}, Y_{[2]m}\right)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
k	$\left(X_{(k)1}, Y_{[k]1}\right)$	$\left(X_{(k)2}, Y_{[k]2}\right)$	$\cdots$	$\left(X_{(k)m}, Y_{[k]m}\right)$

in the Table 4.2 where $r = 1, \cdots, k$ and $j = 1, \cdots, m$. For the illustration purpose, we assume that $\hat{\varrho}$ is calculated by using the Eq. (3.12). Now, we aim to obtain $(1-\alpha)\%$ confidence interval at significance level α. There are several ways to confidence interval of dependence parameter ρ. For example, Modarres & Zheng (2004) provided parametric bootstrap and asymptotic confidence intervals of ϱ of bivariate normal distribution under case (i) and case (ii). The non-parametric approach is another option that has not been studied in the literature yet. We define concomitant based BRSSR (CBRSSR), BRSS (CBRSS) and MRBRSS (CBRSSR) as following.

Method 1: Concomitant based BRSSR (CBRSSR)

I: The pairs $\left(X^*_{(r)1}, Y^*_{[r]1}\right), \cdots, \left(X^*_{(r)m}, Y^*_{[r]m}\right)$ are selected from the rth row of the Table 4.2.

II: Repeat the Item I k times, $r = 1, 2, \cdots, k$, to obtain a bootstrap ranked set sample $\left(X^*_{(r)j}, Y^*_{[r]j}\right)$, $j = 1, 2, \cdots, m$.

III: By using the Eq. (3.17), $\hat{\varrho}^{M1}$ is obtained.

IV: Repeat the Items I-III B times to obtain $\left\{\hat{\varrho}_1^{M1}, \cdots, \hat{\varrho}_B^{M1}\right\}$.

V: Construct a $100(1-\alpha)\%$ percentile confidence interval,

$$\left(\hat{\varrho}_{(\alpha/2)}^{M1}, \hat{\varrho}_{(1-\alpha/2)}^{M1}\right),$$

where $\hat{\varrho}_{(\alpha/2)}^{M1}$ and $\hat{\varrho}_{(1-\alpha/2)}^{M1}$ are the points corresponding to the $\alpha/2$ and $1-\alpha/2$ percentiles of the distribution of $\hat{\varrho}^{M1}$, respectively.

Method 2: Concomitant based BRSS (CBRSS)

I: Select k^2 pairs from the whole of the Table 4.2.

II: Split them into k sets at random. Let rth set is

$$\left\{\left(X_{r1}^*, Y_{r1}^*\right), \cdots, \left(X_{rk}^*, Y_{rk}^*\right)\right\}$$

for $r = 1, \cdots, k$.

III: Rank the X-values in ascending order,

$$X_{r(1)}^* \leq \cdots \leq X_{r(k)}^*$$

IV: Select $\left(X_{(r)1}^*, Y_{[r]1}^*\right)$ from the rth set for each $r = 1, \cdots, k$.

V: Repeat the Steps I-IV m times to obtain $\left(X_{(r)j}^*, Y_{[r]j}^*\right)$ where $j = 1, \cdots, m$.

VI: By using the Eq. (3.17), $\hat{\varrho}^{M2}$ is obtained.

VII: Repeat the Items I-VI B times to obtain $\left\{\hat{\varrho}_1^{M2}, \cdots, \hat{\varrho}_B^{M2}\right\}$.

VIII: Construct a $100(1-\alpha)\%$ percentile confidence interval,

$$\left(\hat{\varrho}_{(\alpha/2)}^{M2}, \hat{\varrho}_{(1-\alpha/2)}^{M2}\right),$$

where $\hat{\varrho}_{(\alpha/2)}^{M2}$ and $\hat{\varrho}_{(1-\alpha/2)}^{M2}$ are the points corresponding to the $\alpha/2$ and $1-\alpha/2$ percentiles of the distribution of $\hat{\varrho}^{M2}$, respectively.

Method 3: Concomitant based MRBRSS (CBRSS)

I: Select one pairs at random from each row of the Table 4.2.

II: Rank the X-values in ascending order and select $\left(X^*_{(1)}, Y^*_{[1]}\right)$.

III: Repeat the Steps I and II k times. At the rth repetition of the procedure, select $\left(X^*_{(r)1}, Y^*_{[r]1}\right)$.

IV: Repeat the Steps I-III m cycles to obtain $\left(X^*_{(r)j}, Y^*_{[r]j}\right)$ where $j = 1, \cdots, m$.

V: By using the Eq. (3.17), $\hat{\varrho}^{M3}$ is obtained.

VI: Repeat the Items I-V B times to obtain $\left\{\hat{\varrho}^{M3}_1, \cdots, \hat{\varrho}^{M3}_B\right\}$.

VII: Construct a $100(1-\alpha)$% percentile confidence interval,

$$\left(\hat{\varrho}^{M3}_{(\alpha/2)}, \hat{\varrho}^{M3}_{(1-\alpha/2)}\right),$$

where $\hat{\varrho}^{M3}_{(\alpha/2)}$ and $\hat{\varrho}^{M3}_{(1-\alpha/2)}$ are the points corresponding to the $\alpha/2$ and $1-\alpha/2$ percentiles of the distribution of $\hat{\varrho}^{M3}$, respectively.

This issue will be evaluated as a future work. Making comparisons between different sample correlation coefficients could be meaningful. Also, other bootsrap confidence intervals can be considered in addition to the percentile bootstrap.

CHAPTER FIVE
CONCLUSION

RSS is quite attractive sampling scheme in the environmental studies. It has been observed that estimators based on RSS have outperformed their counterparts in many problems. Thus, it can be seen that the sampling process is vital. Also, the sampling process becomes even more crucial when taking into account the challenges encountered in the environmental studies during the process. In this dissertation, SRS, RSS, $GMRSS(R = r)$ and ERSS sampling procedures are studied. The sampling procedures can be easily applied to environmental problems. If some information about the data, such as location, is recorded in the computer, an algorithm can be defined for the sampling step. Upon request, we will provide the R functions used for the sampling schemes. Note that the rank-based samples from a given bivariate probability distribution can be generated using these algorithms. They can, however, be used to draw a sample from a population with a few mirror adjustments in codes.

Particularly in case where the random variables have a conditionally nonlinear relationship, there is a few studies on estimating the dependence parameter. Therefore, it can be said that there is a considerable gap in the literature. This dissertation addresses this gap. In this dissertation, the dependence parameters of Gumbel (type I) bivariate exponential, FGM type bivariate gamma and bivariate normal distributions are investigated.

Some concluding remarks and suggestions are listed below in light of the findings of the dissertation.

- The first part of the Chapter 2 provides ML estimators RSS, $GMRSS(R = r)$, and ERSS for the dependence parameter of Gumbel (type I) bivariate exponential distribution. The developed estimators are compared with Barnett's ML estimator based on SRS (Barnett, 1985).

  I. We can say that the biases range from 0 and 0.4 for all studied ML estimators. However, it is observed that $GMRSS(R = k)$ provides ML estimator with the lowest degree of bias.

II. RSS offers an ML estimator as efficient as its counterpart in SRS.

III. It is obvious that the ML estimator based on $GMRSS(R = 1)$ (or $r = k$) and ERSS have lower MSEs than the ML estimator based on SRS and RSS.

IV. Furthermore, the power of test statistics based on $GMRSS(R = k)$ is seen to reach 1 more quickly.

V. In comparison to RSS, and ERSS, the $GMRSS(R = r)$ procedure makes it easier to obtain sample of size $n = mk$ because the measurements are only obtained from a specific rank. As a result, the authors suggest using $GMRSS(R = k)$ to estimate the dependence parameter of (type I) bivariate exponential distribution.

- The second part of the Chapter 2 gives ML estimators for the dependence parameter of FGM type bivariate gamma distribution.

I. It is observed that there is no significant difference between the ML estimator based on SRS and RSS in terms of their estimated values, REs and RIs.

II. It is showed that ML estimator based on $GMRSS(R = 1)$ (or $r = k$) has smallest bias value when the set size is k. Also, $GMRSS(R = 1)$ and $GMRSS(R = k)$ provides more efficient ML estimators than other sampling designs.

III. Furthermore, it can be seen that the findings from this part of Chapter 2 confirm those from the first part of Chapter 2.

- In the Chapter 3, we investigate MPL estimator ($\hat{\rho}_{YT}$) based on RSS for the dependence parameter of bivariate normal distribution. The developed estimator is compared with its counterparts in SRS and RSS under perfect and imperfect ranking cases.

I. It is seen that $\hat{\rho}_{YT}$ is a biased estimator for the dependence parameter. The bias values decrease while the sample size increases. Also, we can say that other estimators are unbiased.

II. According to the REs, it is obvious that $\hat{\rho}_{YT}$ is more efficient than $\hat{\rho}_{RSS_{SRS}}$ and $\hat{\rho}_{RSS_{ZM}}$.

III. On the other hand, $\hat{\rho}_{YT}$ appears to be less efficient than $\hat{\rho}_{RSS_{HMZ}}$ as $\rho \to 0$ but more efficient for larger ρ.

IV. The REs under imperfect ranking case are relatively higher than those under perfect ranking case when $\rho \geq 0.5$. Thus, the $\hat{\rho}_{YT}$ seems to be robust to the imperfect ranking case.

V. Most importantly, this chapter offers different perspective by discussing an alternative approach to MML. This approach does not require estimating the parameters of X and Y. We only need to know the copula form of the bivariate probability distribution and EDFs.

REFERENCES

Abo-Eleneen, Z., & Nagaraja, H. (2002). Fisher information in an order statistic and its concomitant. *Annals of the Institute of Statistical Mathematics*, *54*(3), 667–680.

Al-Nasser, A. D., & Radaideh, A. (2008). Estimation of simple linear regression model using l ranked set sampling. *International Journal of Open Problems in Computer Science and Mathematics*, *1*(1), 18–33.

Al-Rawwash, M., Alodat, M., & Nawajah, I. (2009). Analysis of simple linear regression via median ranked set sampling. *Metron-International Journal of Statistics*, *67*(1), 57–74.

Al-Saleh, M. F., & Al-Kadiri, M. A. (2000). Double-ranked set sampling. *Statistics & Probability Letters*, *48*(2), 205–212.

Al-Saleh, M. F., & Samawi, H. M. (2005). Estimation of the correlation coefficient using bivariate ranked set sampling with application to the bivariate normal distribution. *Communications in Statistics-Theory and Methods*, *34*(4), 875–889.

Al-Saleh, M. F., & Zheng, G. (2002). Estimation of bivariate characteristics using ranked set sampling. *Australian & New Zealand Journal of Statistics*, *44*(2), 221–232.

Alirezaei Dizicheh, M., Iranpanah, N., & Zamanzade, E. (2021). Bootstrap methods for judgment post stratification. *Statistical Papers*, *62*, 2453–2471.

Bairamov, I., & Bekci, M. (2012). Concomitant of order statistics in fgm type bivariate uniform distributions. *Istatistik, Journal of the Turkish Statistical Association*, *2*(2), 135–144.

Bairamov, I., & Kotz, S. (2000). On local dependence function for multivariate distributions. *New Trends in Probability and Statistics*, *5*, 27–44.

Balakrishnan, N., & Lai, C. D. (2009). *Continuous bivariate distributions*. Springer.

Balasubramanian, K., & Beg, M. (1998). Concomitant of order statistics in gumbel's bivariate exponential distribution. *Sankhyā: The Indian Journal of Statistics, Series B*, *60*(3), 399–406.

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, *42*(11), 1–12.

Barnett, V. (1985). The bivariate exponential distribution; a review and some new results. *Statistica Neerlandica*, *39*(4), 343–356.

Barnett, V. (1999). Ranked set sample design for environmental investigations. *Environmental and Ecological Statistics*, *6*, 59–74.

Barnett, V., Green, P., & Robinson, A. (1976). Concomitants and correlation estimates. *Biometrika*, *63*(2), 323–329.

Bhave, P. P., & Sadhwani, K. (2022). Sampling in environmental matrices: a critical review. *Environmental Forensics*, *23*(1-2), 75–92.

Bilgin, O. C., Öztürk, O., & Wolfe, D. A. (2004). Estimation of population variance: A ranke set sampling approach in a finite population setting. *Atatürk Üniversitesi Ziraat Fakültesi Dergisi*, *35*(1-2).

Bortey-Sam, N., Nakayama, S. M., Ikenaka, Y., Akoto, O., Baidoo, E. et al. (2016). Heavy metals and metalloid accumulation in livers and kidneys of wild rats around gold-mining communities in tarkwa, ghana.

Brent, R. P. (2013). *Algorithms for minimization without derivatives*. Courier Corporation.

Ceruti, R., Ghisleni, G., Ferretti, E., Cammarata, S., Sonzogni, O., & Scanziani, E. (2002). Wild rats as monitors of environmental lead contamination in the urban area of milan, italy. *Environmental Pollution*, *117*(2), 255–259.

Chatterjee, A., & Ghosh, S. (2022). A review of bootstrap methods in ranked set sampling. In *Ranked Set Sampling Models and Methods* (171–189). IGI Global.

Chen, H., Stasny, E. A., & Wolfe, D. A. (2007). Improved procedures for estimation of disease prevalence using ranked set sampling. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *49*(4), 530–538.

Chen, M., & Lim, J. (2011). Estimating variances of strata in ranked set sampling. *Journal of Statistical Planning and Inference*, *141*(8), 2513–2518.

Chen, Z., Bai, Z., & Sinha, B. K. (2004). *Ranked set sampling: theory and applications*, volume 176. Springer.

Chen, Z., & Wang, Y.-G. (2004). Efficient regression analysis with ranked-set sampling. *Biometrics*, *60*(4), 997–1004.

Cobby, J., Ridout, M., Bassett, P., & Large, R. (1985). An investigation into the use of ranked set sampling on grass and grass-clover swards. *Grass and Forage Science*, *40*(3), 257–263.

David, H., & Galambos, J. (1974). The asymptotic theory of concomitants of order statistics. *Journal of Applied Probability*, *11*(4), 762–770.

David, H., & Levine, D. (1972). Ranked set sampling in the presence of judgment error. *Biometrics*, *28*, 553–555.

David, H. A. (1973). Concomitants of order statistics. *Bulletin of International Statistical Institute*, *45*, 295–300.

David, H. A., & Nagaraja, H. N. (2004). *Order statistics*. John Wiley & Sons.

Dell, T., & Clutter, J. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, *28*(2), 545–555.

Deshpande, J. V., Frey, J., & Ozturk, O. (2006). Nonparametric ranked-set sampling confidence intervals for quantiles of a finite population. *Environmental and Ecological Statistics*, *13*, 25–40.

D'Este, G. (1981). A morgenstern-type bivariate gamma distribution. *Biometrika*, *68*(1), 339–340.

Devroye, L. (1986). *Non-uniform random variate generation*. Springer, New York.

Farlie, D. J. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, *47*(3/4), 307–323.

Genest, C., & Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, *12*(4), 347–368.

Genest, C., Ghoudi, K., & Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, *82*(3), 543–552.

Göçoğlu, A., & Demirel, N. (2019). Estimating the population proportion in modified ranked set sampling methods. *Journal of Statistical Computation and Simulation*, *89*(14), 2694–2710.

Gulay, B. K., & Demirel, N. (2019). Two-layer median ranked set sampling. *Hacettepe Journal of Mathematics and Statistics*, *48*(5), 1560–1569.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, *55*(292), 698–707.

Gupta, A. K., & Wong, C. (1984). On a morgenstern-type bivariate gamma distribution. *Metrika*, *31*, 327–332.

Gurler, S., Ucer, B. H., & Bairamov, I. (2015). On the mean remaining strength at the system level for some bivariate survival models based on exponential distribution. *Journal of computational and applied mathematics*, *290*, 535–542.

Halls, L. K., & Dell, T. R. (1966). Trial of ranked-set sampling for forage yields. *Forest Science*, *12*(1), 22–26.

Hanandeh, A. A., & Al-Saleh, M. F. (2013). Inference on downton's bivariate exponential distribution based on moving extreme ranked set sampling. *Austrian Journal of Statistics*, *42*(3), 161–179.

Haq, A., Brown, J., Moltchanova, E., & Al-Omari, A. I. (2014). Mixed ranked set sampling design. *Journal of Applied Statistics*, *41*(10), 2141–2156.

Hatefi, A., Jozani, M. J., & Ozturk, O. (2015). Mixture model analysis of partially rank-ordered set samples: age groups of fish from length-frequency data. *Scandinavian Journal of Statistics*, *42*(3), 848–871.

Hatefi, A., Jozani, M. J., & Ziou, D. (2014). Estimation and classification for finite mixture models under ranked set sampling. *Statistica Sinica*, *24*(2), 675–698.

Hatefi, A., Reid, N., Jozani, M. J., & Ozturk, O. (2020). Finite mixture modeling, classification and statistical learning with order statistics. *Statistica Sinica*, *30*(4), 1881–1903.

Hazratian, L., Naderi, M., & Mollashahi, M. (2017). Norway rat, rattus norvegicus in metropolitans, a bio-indicator for heavy metal pollution (case study: Tehran, iran). *Caspian Journal of Environmental Sciences*, *15*(1), 85–92.

He, Q. (2007). *Inference on correlation from incomplete bivariate samples*. PhD thesis, The Ohio State University.

He, Q., & Nagaraja, H. (2009). Distribution of concomitants of order statistics and their order statistics. *Journal of statistical planning and inference*, *139*(8), 2643–2655.

He, Q., Nagaraja, H., & Wu, C. (2013). Efficient simulation of complete and censored samples from common bivariate exponential distributions. *Computational Statistics*, *28*(6), 2479–2494.

Hui, T. P., Modarres, R., & Zheng, G. (2005). Bootstrap confidence interval estimation of mean via ranked set sampling linear regression. *Journal of Statistical Computation and Simulation*, *75*(7), 543–553.

Hui, T. P., Modarres, R., & Zheng, G. (2009). Pseudo maximum likelihood estimates using ranked set sampling with applications to estimating correlation. *Test*, *18*, 365–380.

Husby, C. E., Stasny, E. A., & Wolfe, D. A. (2006). An application of ranked set sampling for mean and median estimation using usda crop production data. *Journal of Agricultural, Biological, and Environmental Statistics*, *10*(3), 354–373.

Jasso-Pineda, Y., Espinosa-Reyes, G., González-Mille, D., Razo-Soto, I., Carrizales, L., Torres-Dosal, A., Mejía-Saavedra, J., Monroy, M., Ize, A. I., Yarto, M. et al. (2007). An integrated health risk assessment approach to the study of mining sites contaminated with arsenic and lead. *Integrated Environmental Assessment and Management: An International Journal*, *3*(3), 344–350.

Johnson, M. B. (1987). *Multivariate Statistical Simulation*. Wiley, New York.

Kvam, P. H. (2003). Ranked set sampling based on binary water quality data with covariates. *Journal of Agricultural, Biological, and Environmental Statistics*, *8*(3), 271–279.

Lacayo, H., Neerchal, N. K., & Sinha, B. K. (2002). Ranked set sampling from a dichotomous population. *Journal of Applied Statistical Science*, *11*(1), 83–90.

Long, D., & Krzysztofowicz, R. (1992). Farlie-gumbel-morgenstern bivariate densities: Are they applicable in hydrology? *Stochastic Hydrology and Hydraulics*, *6*, 47–54.

Mahdizadeh, M., & Zamanzade, E. (2021). New estimator for the variances of strata in ranked set sampling. *Soft Computing*, *25*(13), 8007–8013.

Mahdizadeh, M., & Zamanzade, E. (2022). Using a rank-based design in estimating prevalence of breast cancer. *Soft Computing*, *26*(7), 3161–3170.

Martin, W. L., Sharik, T. L., Oderwald, R. G., & Smith, D. W. (1980). Evaluation of ranked set sampling for estimating shrub phytomass in appalachian oak forests.

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

McIntyre, G. (2005). A method for unbiased selective sampling, using ranked sets. *The American Statistician*, *59*(3), 230–232.

Mehrotra, K., & Nanda, P. (1974). Unbiased estimation of parameters by order statistics in the case of censored samples. *Biometrika*, *61*(3), 601–606.

Minamia, T., Yamazakia, H., Ohamab, N., & City, H. (2009). Accumulation of heavy metals in the organs of wild rodents. *Annual reports by Research Institute for Science and Technology*, *1*(21), 11–7.

Modarres, R., Hui, T. P., & Zheng, G. (2006). Resampling methods for ranked set samples. *Computational statistics & data analysis*, *51*(2), 1039–1050.

Modarres, R., & Zheng, G. (2004). Maximum likelihood estimation of dependence parameter using ranked set sampling. *Statistics & Probability Letters*, *68*(3), 315–323.

Moore, R., & Clarke, R. (1981). A distribution function approach to rainfall runoff modeling. *Water Resources Research*, *17*(5), 1367–1382.

Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitteilungsblatt fur Mathematishe Statistik*, *8*, 234–235.

Murray, R., Ridout, M., & Cross, J. (2000). The use of ranked set sampling in spray deposit assessment. *Aspects of Applied Biology*, *57*, 141–146.

Muttlak, H. (1997). Median ranked set sampling. *Journal of Applied Statistical Science*, *6*, 245–255.

Muttlak, H. A. (2003). Modified ranked set sampling methods. *Pakistan Journal of Statistics-All Series*, *19*(3), 315–324.

Nagaraja, H. (2003). Functions of concomitants of order statistics. *Journal of the Indian Society for Probability and Statistics*, *7*, 16–32.

Nahhas, R. W., Wolfe, D. A., & Chen, H. (2002). Ranked set sampling: cost and optimal set size. *Biometrics*, *58*(4), 964–971.

Nazari, S., Jafari Jozani, M., & Kharrati-Kopaei, M. (2014). Nonparametric density estimation using partially rank-ordered set samples with application in estimating the distribution of wheat yield. *Electronic Journal of Statistics*, *8*(1).

Nelsen, R., Quesada-Molina, J., & Rodriguez-Lallena, a. J. (1997). Bivariate copulas with cubic sections. *Journal of Nonparametric Statistics*, *7*(3), 205–220.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer, New York.

Oakes, D. (1994). Multivariate survival distributions. *Nonparametric Statistics*, *3*(3-4), 343–354.

Ozturk, O. (2014). Statistical inference for population quantiles and variance in judgment post-stratified samples. *Computational Statistics & Data Analysis*, *77*, 188–205.

Ozturk, O., Bilgin, O. C., & Wolfe, D. A. (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), 905–919.

Ozturk, O., & Demirel, N. (2016). Estimation of population variance from multi-ranker ranked set sampling designs. *Communications in Statistics-Simulation and Computation*, *45*(10), 3568–3583.

Patil, G., Sinha, A., & Taille, C. (1993). Relative precision of ranked set sampling: a comparison with the regression estimator. *Environmetrics*, *4*(4), 399–412.

Philip, A., & Thomas, P. Y. (2017). On concomitants of order statistics and its application in defining ranked set sampling from farlie-gumbel-morgenstern bivariate lomax distribution. *JIRSS*, *16*(2), 67–95.

Reynolds, K. D., Schwarz, M. S., McFarland, C. A., McBride, T., Adair, B., Strauss, R. E., Cobb, G. P., Hooper, M. J., & McMurry, S. T. (2006). Northern pocket gophers (thomomys talpoides) as biomonitors of environmental metal contamination. *Environmental Toxicology and Chemistry: An International Journal*, *25*(2), 458–469.

Robertson, B., Ozturk, O., Kravchuk, O., & Brown, J. (2022). Spatially balanced sampling with local ranking. *Journal of Agricultural, Biological and Environmental Statistics*, *27*(4), 622–639.

Robertson, B., Reale, M., Price, C., & Brown, J. (2021). Quasi-random ranked set sampling. *Statistics & Probability Letters*, *171*, 1–7.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The annals of mathematical statistics*, *23*(3), 470–472.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method*. John Wiley & Sons.

Samawi, H., & Ababneh, F. (2001). On regression analysis using ranked set sample. *Journal of statistical Research*, *35*(2), 93–105.

Samawi, H. M., Ahmed, M. S., & Abu-Dayyeh, W. (1996). Estimating the population mean using extreme ranked set sampling. *Biometrical Journal*, *38*(5), 577–586.

Samawi, H. M., & Chen, D.-G. (2021). Ranked simulated resampling: a more efficient and accurate resampling approximations for bootstrap inference. *Journal of Statistical Computation and Simulation*, *91*(18), 3709–3720.

Sevil, Y. C. (2017). Goodness-of-fit tests in ranked set sampling. Master's thesis, Dokuz Eylul University.

Sevil, Y. C., & Yildiz, T. O. (2017). Power comparison of the kolmogorov–smirnov test under ranked set sampling and simple random sampling. *Journal of Statistical Computation and Simulation*, *87*(11), 2175–2185.

Sevil, Y. C., & Yildiz, T. O. (2020). Performances of the distribution function estimators based on ranked set sampling using body fat data. *Türkiye Klinikleri Biyoistatistik*, *12*(2), 218–228.

Sevil, Y. C., & Yildiz, T. O. (2021). Tests of normality based on edf statistics using partially rank ordered set sampling designs. *Istatistik Journal of The Turkish Statistical Association*, *13*(2), 52–73.

Sevil, Y. C., & Yildiz, T. O. (2022a). Design-based estimators of the distribution function in ranked set sampling with an application. *Statistics*, *56*(4), 891–918.

Sevil, Y. C., & Yildiz, T. O. (2022b). Gumbel's bivariate exponential distribution: estimation of the association parameter using ranked set sampling. *Computational Statistics*, *37*, 1695–1726.

Sevil, Y. C., & Yildiz, T. O. (2023a). Estimating the dependence parameter of farlie-gumbel-morgenstern type bivariate gamma distribution using ranked set sampling. *Computer Sciences and Mathematics Forum*, 1–8.

Sevil, Y. C., & Yildiz, T. O. (2023b). Estimation of distribution function using percentile ranked set sampling. *REVSTAT-Statistical Journal*, *21*(1), 39–62.

Shih, J. H., & Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, *51*(4), 1384–1399.

Shtiza, A., & Tashko, A. (2009). Appropriate sampling strategy and analytical methodology to address contamination by industry: Part 1. conceptual model of a sampling design and sampling types. *Central European Journal of Geosciences*, *1*(2), 193–206.

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

Sklar, M. (1959). Fonctions de répartition àn dimensions et leurs marges. *Publications de I'Institut de statistique de I'Université de Paris*, *8*, 229–231.

Stokes, S. L. (1977). Ranked set sampling with concomitant variables. *Communications in Statistics-Theory and Methods*, *6*(12), 1207–1211.

Stokes, S. L. (1980). Inferences on the correlation coefficient in bivariate normal populations from ranked set samples. *Journal of the American Statistical Association*, *75*(372), 989–995.

Stokes, S. L., & Sager, T. W. (1988). Characterization of a ranked-set sample with application to estimating distribution functions. *Journal of the American Statistical Association*, *83*(402), 374–381.

Taconeli, C. A., & de Lara, I. A. R. (2022). Improved confidence intervals based on ranked set sampling designs within a parametric bootstrap approach. *Computational Statistics*, *37*(5), 2267–2293.

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

Terpstra, J. (2004). On estimating a population proportion via ranked set sampling. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *46*(2), 264–272.

Terpstra, J. T., & Liudahl, L. A. (2004). Concomitant-based rank set sampling proportion estimates. *Statistics in medicine*, *23*(13), 2061–2070.

Ucer, B. H., & Gurler, S. (2012). On the mean residual lifetime at system level in two-component parallel systems for the fgm distribution. *Hacettepe Journal of Mathematics and Statistics*, *41*(1), 139–145.

Ucer, B. H., & Yildiz, T. O. (2012). Estimation and goodness-of-fit procedures for farlie-gumbel-morgenstern bivariate copula of order statistics. *Journal of Statistical Computation and Simulation*, *82*(1), 137–147.

Wang, K. (2008). *On concomitants of order statistics*. PhD thesis, The Ohio State University.

Weer, D., & Basu, A. (1980). Testing for indepedence in multivariate exponential distributions[2]. *Australian Journal of Statistics*, *22*(3), 276–288.

Wolfe, D. A. (2012). Ranked set sampling: its relevance and impact on statistical inference. *International scholarly research notices*, *2012*, 1–32.

Yamaguchi, H., & Murakami, H. (2021). Symmetric smoothed bootstrap methods for ranked set samples. *Journal of Nonparametric Statistics*, *33*(3-4), 435–463.

Yildiz, T., & Ucer, B. H. (2017). Fisher information of dependence in progressive type ii censored order statistics and their concomitants. *International Journal of Applied Mathematics & Statistics*, *56*(6), 1–10.

Yildiz, T. O., & Sevil, Y. C. (2018). Performances of some goodness-of-fit tests for sampling designs in ranked set sampling. *Journal of Statistical Computation and Simulation*, *88*(9), 1702–1716.

Yildiz, T. O., & Sevil, Y. C. (2019). Empirical distribution function estimators based on sampling designs in a finite population using single auxiliary variable. *Journal of Applied Statistics*, *46*(16), 2962–2974.

Younes, M. (2020). Integration of mathematical median ranked set sample and decision making ahp tools to enhance decentralized wastewater treatment system. *Journal of Water Chemistry and Technology*, *42*(6), 472–479.

Yue, S., Ouarda, T. B., & Bobée, B. (2001). A review of bivariate gamma distributions for hydrological application. *Journal of Hydrology*, *246*, 1–18.

Zamanzade, E. (2019). Edf-based tests of exponentiality in pair ranked set sampling. *Statistical Papers*, *60*(6), 2141–2159.

Zamanzade, E., & Mahdizadeh, M. (2020). Using ranked set sampling with extreme ranks in estimating the population proportion. *Statistical methods in medical research*, *29*(1), 165–177.

Zheng, G., & Modarres, R. (2006). A robust estimate of the correlation coefficient for bivariate normal distribution using ranked set sampling. *Journal of statistical planning and inference*, *136*(1), 298–309.

APPENDICES

Appendix 1: Critical values for LRT statistics

Table A.1 Values of c_ψ, ψ = SRS, RSS, GMRSS($R=1$), GMRSS($R=k$) and ERSS

m	k	LRT_{SRS}	LRT_{RSS}	$LRT_{GMRSS(R=1)}$	$LRT_{GMRSS(R=k)}$	LRT_{ERSS}
2	2	0.292	0.288	0.302	0.277	0.289
	3	0.242	0.242	0.250	0.208	–
	4	0.218	0.221	0.225	0.189	0.210
	5	0.211	0.212	0.232	0.196	–
	6	0.213	0.208	0.233	0.204	0.216
	7	0.206	0.204	0.241	0.208	–
	8	0.211	0.209	0.249	0.208	0.221
5	2	0.210	0.211	0.227	0.194	0.212
	3	0.212	0.212	0.229	0.201	–
	4	0.214	0.215	0.246	0.210	0.220
	5	0.216	0.213	0.252	0.216	–
	6	0.221	0.219	0.258	0.217	0.227
	7	0.223	0.225	0.262	0.226	–
	8	0.225	0.224	0.265	0.226	0.229
10	2	0.214	0.216	0.232	0.204	0.214
	3	0.218	0.216	0.244	0.213	–
	4	0.224	0.224	0.254	0.225	0.228
	5	0.227	0.229	0.264	0.228	–
	6	0.232	0.231	0.259	0.232	0.232
	7	0.232	0.233	0.260	0.233	–
	8	0.235	0.235	0.267	0.231	0.240
15	2	0.220	0.225	0.234	0.216	0.219
	3	0.226	0.224	0.252	0.220	–
	4	0.230	0.229	0.259	0.231	0.236
	5	0.235	0.241	0.261	0.234	–
	6	0.234	0.237	0.265	0.237	0.237
	7	0.232	0.236	0.259	0.236	–
	8	0.236	0.237	0.263	0.242	0.240