Maximum Product of Spacings Estimators of Block and Basu Bivariate Weibull Distribution

Debasis Kundu *

Abstract

Maximum product of spacings estimators for the univariate distributions have received a considerable amount of attention in the statistical literature. They can be used as an alternative of the maximum likelihood estimators. It is well known that for the regular families both the estimators are asymptotically equivalent. Moreover, in certain cases, even though the maximum likelihood estimators do not exist, the maximum product of spacings estimators exist. Since, the product of spacings estimators are based on ordered samples, they do not have a natural extension to the higher dimension. Due to this reason, not much attention has been paid in developing maximum product of spacings estimators for the higher dimensional model. In this paper, we have defined the bivariate maximum product of spacings estimators based on the marginal and the conditional distribution. It seems it is a very natural generalization. It can be extended to more than two dimensions also. We have illustrated this procedure by using the absolutely continuous Block and Basu bivariate Weibull model. Extensive simulations have been performed to show the effectiveness of the proposed method. One data set has been analyzed, and the results are quite satisfactory.

Key Words and Phrases: maximum product of spacings estimators; robust estimators; maximum likelihood estimators; bivariate Weibull distribution; Block and Basu bivariate exponential distribution; lifetime data.

AMS Subject Classifications: 62F10, 62F03, 62H12.

*Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. E-mail: kundu@iitk.ac.in, Phone no. 91-512-2597141, Fax no. 91-512-2597500.

1 Introduction

Cheng and Amin [3] and Ranneby [14] independently developed the maximum product of spacings (MPS) estimators for independently and identically distributed (i.i.d.) univariate distributions. The main idea behind this method is to approximate the Kullback-Leibler (KL) information in such a way that each contribution is bounded from above, unlike likelihood contribution, which may not be bounded from above always. It may be recalled that if $f(x)$ and $g(x)$ are two probability density functions (PDFs), then Kullback and Leibler [10] proposed

$$I(f,g) = \int f(x) \ln \frac{f(x)}{g(x)} dx$$

as the measure of the mean information between f and g, when the observation comes from f. The measure $I(f,g)$ is known as KL-divergence or relative entropy. The estimation method obtained from this approximation is called the maximum spacing method. Interestingly this method works also in situations when the maximum likelihood (ML)-method breaks down.

The basic idea of the MPS method is quite simple and intuitive. The concept underlying the method is based on probability integral transformation, in that a set of independent random samples derived from any random variable should on average be uniformly distributed with respect to the cumulative distribution function (CDF) of the random variable. The MPS method chooses the parameter values that make the observed data as uniform as possible, according to a specific quantitative measure of uniformity. The method can be described as follows. Suppose $\{x_1, \ldots, x_n\}$ is an i.i.d. sample of size n from a univariate distribution with continuous CDF $F(x;\theta^0)$, where $\theta^0 \in \Theta$ is an unknown parameter to be estimated. Let $\{x_{(1)}, \ldots, x_{(n)}\}$ be the corresponding ordered sample. Let us denote the spacings as the gaps between the values of the distribution function at adjacent ordered points, i.e.

$$D_i(\theta) = F(x_{(i)};\theta) - F(x_{(i-1)};\theta), \quad i = 1, \ldots, n+1, \tag{1}$$

where $F(x_{(0)};\theta) = 0$ and $F(x_{(n+1)};\theta) = 1$. Then the MPS estimator of θ^0 is defined as

$$\widehat{\theta} = \arg\max_{\theta\in\Theta} \prod_{i=1}^{n+1} D_i(\theta) = \arg\max_{\theta\in\Theta} \prod_{i=1}^{n+1} F(x_{(i)};\theta) - F(x_{(i-1)};\theta). \tag{2}$$

If $f(x;\theta)$ denotes the PDF corresponds to CDF $F(x;\theta)$, then it has been shown that under certain regularity conditions, maximizing the product of spacings as in (1) is equivalent to maximizing $\sum_{i=1}^{n} \ln f(x_{(i)};\theta)$. Hence, under certain regularity conditions, the maximum likelihood estimators (MLEs) and the MPS estimators are asymptotically equivalent. The main advantage of the MPS estimators is that it involves maximizing a bounded objective function (1), hence, it avoids dealing with the unbounded likelihood function for example three parameter Weibull, three-parameter log-normal, or three-parameter gamma distribution. Although, in case of three-parameter Weibull, three-parameter log-normal, or three-parameter gamma distribution, the MLEs do not exist, the MPS estimators exist in all these cases. Ferguson [6] gave several examples, where the MLEs exist but they are not consistent, it has been observed, see for example Ranneby [14], that in all these cases the MPS estimators exist and they are consistent. In many cases, it has been observed that the MLEs exist but they are not unique and it has been shown that in those cases MPS estimators exist and they are unique. Based on extensive simulations Ekstrom [5] and Khan et al. [8] claimed that the MPS estimators are more robust than the MLEs for small sample sizes.

One major issue about the MPS estimators is that it is based on ordered samples, hence, it does not have a natural extension to higher dimension. Although, the MLEs can be extended to the higher dimension in a unique manner, the same cannot be done in case of MPS estimators. Due to this reason, although it has several advantages for univariate case, and an extensive amount of work has been done related to MPS estimators for univariate case, not much attention has been paid to higher-dimensional model. Ranneby et al. [15], see also Kuljus and Ranneby [9], defined the multivariate MPS estimators based on nearest neighborhood and established their consistency properties. The implementation of the multivariate MPS estimators as proposed by Ranneby et al. [15] is not very simple. Due to this reason, it has not received much attention in the literature.

The main aim of this paper is to define the bivariate MPS estimators for absolutely continuous bivariate distribution function. Although, we will define the MPS estimators for the bivariate distribution, it can be easily extended for any multivariate distribution also. The main idea is simple and intuitive and the implementation is also quite straight forward. The idea is exactly the same as the univariate case. Here we have decomposed the joint PDF as the product of the marginal and the conditional, and we have used probability integral transformation for both cases separately. It has been illustrated using the bivariate absolutely continuous Block and Basu bivariate Weibull (BBBW) model as introduced by Kundu and Gupta [12]. The BBBW is an extension of the Block and Basu bivariate exponential (BBBE) model as introduced by Block and Basu [2]. It may be mentioned that the BBBW distribution is more flexible than the BBBE distribution. The PDFs and hazard functions of the BBBW distribution can take variety of shapes, hence, it can be used quite effectively for data analysis purposes. Extensive simulations have been performed to see the effectiveness of the proposed method and one data set has been analyzed. The performances are quite satisfactory.

The main contributions of this paper can be summarized as follows:

• We have introduced the maximum producet of spacings estimators for a bivariate distribution. It is a very natural extension from a univariate to a bivariate distribution.

• The implementaion of the proposed method is very simple.

• We have derived the asymptotic properties of the proposed estimators under a failrly general class of sufficinet conditions.

• We have implements the proposed method for the Block and Basu bivariate Weibull distribution.

The rest of the paper is organized as follows. We have described the BBBE distribution in Section 2 and discuss some of its properties. The bivariate MPS (BMPS) estimators have been defined in Section 3. In Section 4 we have shown how the proposed method can be implemented

for BBBW model. Extensive numerical results and the analysis of a data set have been presented in Section 5 and Section 6, respectively. Finally we conclude the paper in Section 7.

2 Block and Basu Bivariate Weibull Distribution

Marshall and Olkin [13] in their classical paper introduced the bivariate exponential distribution via a reduction technique. It has been defined as follows. Let us use the following notation. A random variable X is said to have an exponential distribution with scale parameter $\lambda > 0$, denoted by EX(λ), if it has the following PDF;

$$f_{EX}(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x \leq 0. \end{cases} \tag{3}$$

It has the following distribution function, survival function and hazard function for $x > 0$;

$$F_{EX}(x;\lambda) = 1 - e^{-\lambda x}, \quad S_{EX}(x;\lambda) = e^{-\lambda x}, \quad h_{EX}(x;\lambda) = \lambda,$$

respectively. Suppose U_1 follows ($\sim$) EX(λ_1), $U_2 \sim$ EX(λ_2) and $U_3 \sim$ EX(λ_3), and they are independently distributed, then Marshall and Olkin [13] define

$$X = \min\{U_1, U_3\} \quad \text{and} \quad Y = \min\{U_2, U_3\}.$$

as a bivariate exponential distribution. This is known as the Marshall-Olkin bivariate exponential distribution (MOBE). From now on we will define it as MOBE($\lambda_1, \lambda_2, \lambda_3$). It can be easily seen that although, U_1, U_2 and U_3 are independently distributed, X and Y are dependent due to presence of U_3. Further, observe that $X \sim$ EX($\lambda_1 + \lambda_3$), $Y \sim$ EX($\lambda_2 + \lambda_3$), and

$$P(X = Y) = P(U_3 < \min\{U_1, U_2\}) = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}.$$

Hence, although X and Y are absolutely continuous univariate random variables, (X, Y) is not an absolutely continuous bivariate random variable. It does not have two-dimensional PDF. It has a positive support on the line $x = y$

The MOBE$(\lambda_1, \lambda_2, \lambda_3)$ has the following PDF

$$f_{MOBE}(x,y) = \begin{cases} f_{EX}(x;\lambda_1) f_{EX}(y;\lambda_2+\lambda_3) & \text{if} \quad x<y \\ f_{EX}(x;\lambda_1+\lambda_3) f_{EX}(y;\lambda_2) & \text{if} \quad x>y \\ \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3} f_{EX}(z;\lambda_1+\lambda_2+\lambda_3) & \text{if} \quad x=y=z \end{cases} \tag{4}$$

under the two dimensional Lebesgue measure on $x \neq y$ and under one dimensional Lebesgue measure on $x = y$, see for example Kundu [11] for details. The MOBE distribution has several interesting properties and a physical interpretation.

Block and Basu [2] introduced the absolutely continuous bivariate exponential distribution from the MOBE distribution by removing the singular component and after proper normalization as follows:

$$f_{BBBE}(x,y) = \frac{\lambda_1+\lambda_2+\lambda_3}{\lambda_1+\lambda_2} \times \begin{cases} f_{EX}(x;\lambda_1) f_{EX}(y;\lambda_2+\lambda_3) & \text{if} \quad x<y \\ f_{EX}(x;\lambda_1+\lambda_3) f_{EX}(y;\lambda_2) & \text{if} \quad x>y. \end{cases} \tag{5}$$

From now on it will be denoted by BBBE$(\lambda_1, \lambda_2, \lambda_3)$. Block and Basu [2] developed several interesting properties of the BBBE model. Although, it has several interesting properties, the PDF and the hazard function (HF) of the BBBE distribution are always a decreasing function. Hence, it cannot be used very effectively for data analysis purposes. Due to this reason Kundu and Gupta [12] introduced the Block and Basu bivariate Weibull (BBBW) distribution by replacing the exponential distribution with the two-parameter Weibull distribution. It makes the model very flexible and it can be used quite effectively for data analysis purposes. It can be defined as follows. Let us introduce the following notation. A random variable X is said to have a Weibull distribution with the shape parameter $\alpha > 0$ and the scale parameter $\lambda > 0$, denoted by WE(α, λ), if it has the following PDF;

$$f_{WE}(x;\alpha,\lambda) = \begin{cases} \alpha\lambda x^{\alpha-1} e^{-\lambda x^\alpha} & \text{if} \quad x>0 \\ 0 & \text{if} \quad x \leq 0. \end{cases} \tag{6}$$

It has the following distribution function, survival function and hazard function for $x > 0$;

$$F_{WE}(x;\alpha,\lambda) = 1 - e^{-\lambda x^\alpha}, \quad S_{WE}(x;\alpha,\lambda) = e^{-\lambda x^\alpha}, \quad h_{WE}(x;\alpha,\lambda) = \alpha\lambda x^{\alpha-1},$$

respectively.

Now instead of exponential distributions, if we take $U_1 \sim \text{WE}(\alpha, \lambda_1)$, $U_2 \sim \text{WE}(\alpha, \lambda_2)$, $U_3 \sim \text{WE}(\alpha, \lambda_3)$ and proceed the same way as before, we can obtain the BBBW distribution. The joint PDF of the BBBW distribution becomes,

$$f_{BBBW}(x,y) = \frac{\lambda_1+\lambda_2+\lambda_3}{\lambda_1+\lambda_2} \times \begin{cases} f_{WE}(x;\alpha,\lambda_1) f_{WE}(y;\alpha,\lambda_2+\lambda_3) & \text{if} \quad x<y \\ f_{WE}(x;\alpha,\lambda_1+\lambda_3) f_{WE}(y;\alpha,\lambda_2) & \text{if} \quad x>y. \end{cases} \tag{7}$$

From now on it will be denoted by BBBW$(\alpha, \lambda_1, \lambda_2, \lambda_3)$. The shape of the joint PDF of the BBBW can be unimodal or a decreasing shape depending on the value of the shape parameter. The BBBE becomes a special case of the BBBW distribution.

3 Bivariate MPS Estimators

3.1 Methodology

Suppose the bivariate random variable (X, Y) has the joint CDF $F_{X,Y}(x, y; \theta^0)$. Here θ^0 denotes the true parameter value and $\theta^0 \in \boldsymbol{\Theta} \subset \mathbb{R}^p$. Let $f_{X,Y}(x, y; \theta)$ denote the joint PDF of $F(x, y; \theta)$. Further, $F_X(x; \gamma_1^0)$, $f_X(x; \gamma_1^0)$, $F_{Y|X=x}(y; \theta^0)$, where $\theta^0 = (\gamma_1^0, \gamma_2^0)$, denote the CDF of X, PDF of X, the conditional distribution of Y given $X = x$, respectively. It is assumed $\gamma_1^0 \in \boldsymbol{\Gamma} \subset \mathbb{R}^q$, where $q \leq p$. If $p = q$, then $\boldsymbol{\Theta} = \boldsymbol{\Gamma}$. Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. random variables with the joint CDF $F_{X,Y}(x, y, \theta^0)$. We denote $X_{(1)} < \ldots < X_{(n)}$ and $Y_{(1)} < \ldots < Y_{(n)}$ as the ordered values of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively.

The bivariate MPS estimator based on $(X_1, Y_1), \ldots, (X_n, Y_n)$ is defined as follows. Suppose for $i = 1, \ldots, n$,

$$U_i(\gamma_1) = F_X(X_i; \gamma_1), \qquad V_i(\theta) = F_{Y|X=x_i}(Y_i; \theta)$$

and $U_{(1)}(\gamma_1) < \ldots < U_{(n)}(\gamma_1)$, $V_{(1)}(\theta) < \ldots < V_{(n)}(\theta)$ denote the ordered values of $U_1(\gamma_1), \ldots, U_n(\gamma_1)$ and $V_1(\theta), \ldots, V_n(\theta)$, respectively. Further, $U_{(0)}(\gamma_1) = V_{(0)}(\theta) = 0$ and $U_{(n+1)}(\gamma_1) = V_{(n+1)}(\theta) = 1$.

We make the following assumptions and need the following results for further development.

(A1) The parameter space $\Theta \subset \mathbb{R}^p$ is compact and the true parameter θ^0 is an interior point of Θ.

(A2) The model is identifiable. More precisely, $F_X(x;\gamma_1) = F_X(x;\gamma_2)$ for almost every x implies $\gamma_1 = \gamma_2$ and similarly, $F_{Y|X=x}(y;\theta_1) = F_{Y|X=x}(y;\theta_2)$, for almost every (x,y) implies $\theta_1 = \theta_2$.

(A3) For every (x,y), the marginal distribution $F_X(x;\theta)$ and the conditional distribution $F_{Y|X=x}(y;\theta)$ are absolutely continuous in θ. Furthermore, their first and second order partial derivatives with respect to the parameter exist and are continuous.

(A4) The corresponding marginal and conditional densities, $f_X(x;\theta)$ and $f_{Y|X=x}(y;\theta)$, respectively, are positive on the interior of their supports. Moreover, for every compact subset $\mathcal{K}$ of the support, $\inf_{(x,y)\in\mathcal{K},\,\theta\in\Theta} f_X(x;\theta) > 0$ and $\inf_{(x,y)\in\mathcal{K},\,\theta\in\Theta} f_{Y|X=x}(y;\theta) > 0$.

(A5) The inverse distribution functions $F_X^{-1}(u;\theta)$ and $F_{Y|X=x}^{-1}(u;\theta)$ exist for every $0 < u < 1$ and are jointly continuous in (u,θ).

Lemma 1: Suppose Assumptions (A1)–(A5) hold. $(X_1,Y_1),\ldots,(X_n,Y_n)$ and $U_1(\gamma_1^0),\ldots,U_n(\gamma_1^0)$ are same as defined before. Then $U_1(\gamma_1^0),\ldots,U_n(\gamma_1^0)$ are i.i.d. U$(0,1)$ random variables.

Proof: It just follows from the probability integral transformation. ∎

Lemma 2: Suppose Assumptions (A1)–(A5) hold. $(X_1,Y_1),\ldots,(X_n,Y_n)$ and $V_1(\theta^0),\ldots,V_n(\theta^0)$ are same as defined before. Then $V_1(\theta^0),\ldots,V_n(\theta^0)$ are i.i.d. U$(0,1)$ random variables.

Proof: Using Lemma 1, for each $i = 1,\ldots,n$, it follows that $V_i \,|\, X_i = x \sim \text{U}(0,1)$, for every possible value of x. Now remove the conditioning, using the law of total probability,

$$P(V_i \le v) = E\left[P(V_i \le v \,|\, X_i)\right] = E(v) = v.$$

Thus, $V_i \sim \mathrm{U}(0,1)$. Further, since (X_i, Y_i)'s are independently distributed, V_i is a measurable function of (X_i, Y_i), it follows that V_i's are also independently distributed. ∎

Lemma 3: Suppose Assumptions (A1)–(A5) hold. $(X_1, Y_1), \ldots, (X_n, Y_n)$, $U_1(\gamma_1^0), \ldots, U_n(\gamma_1^0)$ and $V_1(\theta^0), \ldots, V_n(\theta^0)$ are same as defined before. Then each $U_i(\gamma_1^0)$ and $V_i(\theta^0)$ are independent random variables. Consequently, $(U_1(\gamma_1^0), V_1(\theta^0)), \ldots, (U_n(\gamma_1^0), V_n(\theta^0))$ are independent and identically distributed random vectors having the joint distribution $(U_i(\gamma_1^0), V_i(\theta^0)) \sim \mathrm{U}(0,1) \times \mathrm{U}(0,1)$

Proof: By Lemma 1 and Lemma 2, $U_i(\gamma_1^0) \sim \mathrm{U}(0,1)$ and $V_i(\theta^0) \sim \mathrm{U}(0,1)$. It therefore remains to prove their independence. For any $0 \leq u, v \leq 1$,

$$\begin{aligned} P(U_i(\gamma_1^0) \leq u, V_i(\theta^0) \leq v) &= E\left[P(U_i(\gamma_1^0) \leq u, V_i(\theta^0) \leq v \,|\, X_i)\right] \\ &= E\left[I\big\{F_X(X_i;\theta^0) \leq u\big\}\, P(V_i(\theta^0) \leq v \,|\, X_i)\right], \end{aligned}$$

where $I(\cdot)$ denotes the indicator function. By the conditional probability integral transform, $P(V_i \leq v \,|\, X_i) = v$. Hence,

$$P(U_i(\gamma_1^0) \leq u, V_i(\theta^0) \leq v) = v\, E\left[I\big\{F_X(X_i;\theta^0) \leq u\big\}\right] = v\, P(U_i(\gamma_1^0) \leq u) = uv.$$

Therefore,

$$P(U_i(\gamma_1^0) \leq u, V_i(\theta^0) \leq v) = P(U_i(\gamma_1^0) \leq u)\, P(V_i(\theta^0) \leq v),$$

which proves that $U_i(\gamma_1^0)$ and $V_i(\theta^0)$ are independent. Since, $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independently distributed, it follows that $(U_1(\gamma_1^0), V_1(\theta^0)), \ldots, (U_n(\gamma_1^0), V_n(\theta^0))$ are also independent. ∎

Let us define

$$\widetilde{H}_1(\gamma_1) = \frac{1}{n+1} \sum_{i=1}^{n+1} \ln D_i^1(\gamma_1) \quad \text{and} \quad \widetilde{H}_2(\theta) = \frac{1}{n+1} \sum_{i=1}^{n+1} \ln D_i^2(\theta) \tag{8}$$

where $D_i^1(\gamma_1) = U_{(i)}(\gamma_1) - U_{(i-1)}(\gamma_1)$ and $D_i^2(\theta) = V_{(i)}(\theta) - V_{(i-1)}(\theta)$ for $i = 1, \ldots, n$. Suppose $D_1^1, \ldots, D_{n+1}^1$ can be freely chosen subject to the condition that $D_i^1 \geq 0$, for $i = 1, \ldots, n+1$ and

$\sum_{i=1}^{n+1} D_i^1 = 1$, with $\widetilde{H}_1 = \widetilde{H}_1(D_1^1, \ldots, D_{n+1}^1)$ as a function of D_i^1's, then we have the upper bound

$$\widetilde{H}_1(D_1^1, \ldots, D_{n+1}^1) \leq \max_{D_i^1} \frac{1}{n+1} \sum_{i=1}^{n+1} \ln D_i^1 = -\ln(n+1). \tag{9}$$

Further, the upper bound is obtained when all the D_i^1's are equal, with $D_i^1 = \dfrac{1}{n+1}$, see Darling [4]. Following similar notations for $\widetilde{H}_2$, it can be said that

$$\widetilde{H}_2(D_1^2, \ldots, D_{n+1}^2) \leq \max_{D_i^2} \frac{1}{n+1} \sum_{i=1}^{n+1} \ln D_i^2 = -\ln(n+1). \tag{10}$$

and the upper bound is obtained when $D_i^2 = \dfrac{1}{n+1}$, for $i = 1, \ldots, n+1$. Observe that if $\theta = \theta^0$, hence $\gamma_1 = \gamma_1^0$, then $\{D_1^1(\gamma_1), \ldots, D_{n+1}^1(\gamma_1)\}$ are identically distributed and the same is true for $\{D_1^2(\theta), \ldots, D_{n+1}^2(\theta)\}$. Further, $\{D_1^1(\gamma_1^0), \ldots, D_{n+1}^1(\gamma_1^0)\}$ and $\{D_1^2(\theta^0), \ldots, D_{n+1}^2(\theta^0)\}$ are independently distributed. It can be viewed as the statistical analogue of setting all the D_i^1's and D_i^2's equal in the deterministic situations where D_i^1's and D_i^2's can be chosen freely. A simple way of trying to get as close to this deterministic case as possible is to choose θ, so that

$$\widetilde{H}(\theta) = \widetilde{H}(\gamma_1, \gamma_2) = \widetilde{H}_1(\gamma_1) + \widetilde{H}_2(\theta) = \widetilde{H}_1(\gamma_1) + \widetilde{H}_2(\gamma_1, \gamma_2), \tag{11}$$

for $\theta \in \Theta$ is as large as possible. If $\widehat{\theta}$ maximizes $H(\theta)$, then $\widehat{\theta}$ is called maximum product of spacings (MPS) estimators of θ^0.

3.2 Asymptotic Properties

In this section we provide the asymptotic properties of the BMPS estimators under suitable regularity conditions. Throughout this section, let $\theta \in \mathbf{\Theta} \subset \mathbb{R}^p$ denote the unknown parameter vector and let θ^0 denote the true parameter value. For notational simplicity, in this section, it is assumed that $p = q$, $\mathbf{\Theta} = \mathbf{\Gamma}$. We use the following notations:

$$H_n(\theta) = H_{1n}(\theta) + H_{2n}(\theta),$$

where

$$H_{1n}(\theta) = \frac{1}{n+1}\sum_{i=1}^{n+1} \ln[(n+1)D_i^{(1)}(\theta)], \quad \text{and} \quad H_{2n}(\theta) = \frac{1}{n+1}\sum_{i=1}^{n+1} \ln[(n+1)D_i^{(2)}(\theta)].$$

Observe that

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \widetilde{H}(\theta) = \arg\max_{\theta\in\Theta} H_n(\theta).$$

Other than the Assumptions (A1)-(A5), the following regularity assumptions will be used throughout this section.

(A6) The support of the joint distribution does not depend on the unknown parameter.

(A7) There exist measurable function $M_1(x)$ such that $\sup_{\theta\in\Theta} |\log f_X(x;\theta)| \leq M_1(x)$ where $E_{\theta^0}[M_1(X)] < \infty$ and similarly there exists a measurable function $M_2(x,y)$ such that $\sup_{\theta\in\Theta} \left|\log f_{Y|X=x}(y;\theta)\right| \leq M_2(x,y)$, where $E_{\theta^0}[M_2(X,Y)] < \infty$.

Comments: Assumptions (A1)–(A7) are standard regularity conditions in the asymptotic theory of maximum product of spacings estimators. They guarantee the existence of the probability integral transform, continuity of the BMPS objective function, and the validity of the limiting arguments used in proving the consistency of the proposed estimator.

Theorem 1: Suppose Assumptions (A1)–(A7) hold. Let us define

$$H_1(\theta) = E_{\theta^0}\left[\ln \frac{f_X(X;\theta)}{f_X(X;\theta^0)}\right] \quad \text{and} \quad H_2(\theta) = E_{\theta^0}\left[\ln \frac{f_{Y|X}(Y|X;\theta)}{f_{Y|X}(Y|X;\theta^0)}\right].$$

Then

$$H_n(\theta) = H_{1n}(\theta) + H_{2n}(\theta) \xrightarrow{P} H(\theta) = H_1(\theta) + H_2(\theta),$$

where $H(\theta)$ has a unique maximizer at θ^0. Consequently,

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} H_n(\theta) \xrightarrow{P} \theta^0$$

Proof: Since $H_{1n}(\theta)$ is the ordinary univariate maximum product of spacings objective function based on the marginal distribution, Theorem 2 of Shao [16] implies

$$\sup_{\theta\in\Theta} |H_{1n}(\theta) - H_1(\theta)| \xrightarrow{P} 0.$$

Similarly, based on the Assumptions A1-A7, the conditional spacing criterion $H_{2n}(\theta)$ also satisfies the regularity conditions of Shao [16]. Therefore,

$$\sup_{\theta\in\Theta} |H_{2n}(\theta) - H_2(\theta)| \xrightarrow{P} 0.$$

Hence,

$$\sup_{\theta\in\Theta} |H_n(\theta) - H(\theta)| = \sup_{\theta\in\Theta} |(H_{1n} - H_1) + (H_{2n} - H_2)| \leq \sup_{\gamma_1\in\Gamma} |H_{1n} - H_1| + \sup_{\theta\in\Theta} |H_{2n} - H_2|.$$

Since each term on the right-hand side converges to zero in probability,

$$\sup_{\theta\in\Theta} |H_n(\theta) - H(\theta)| \xrightarrow{P} 0.$$

Note that

$$H(\theta) = E\left[\ln \frac{f_X(X;\theta)}{f_X(X;\theta^0)}\right] + E\left[\ln \frac{f_{Y|X}(Y;\theta)}{f_{Y|X}(Y;\theta^0)}\right] \leq 0,$$

as the Kullback-Leibler divergence measure is always non-negative. Note that each term in the above expression of $H(\theta)$ is the negative of the Kullback-Leibler divergence measure. Moreover $H(\theta) = 0$, with equality only when $\theta = \theta^0$, by the identifiability assumption, $H(\theta)$ has a unique maximizer at θ^0. Therefore,

$$\widehat{\theta}_n \xrightarrow{P} \theta^0.$$

This completes the proof. ∎

Now we would like to provide the asymptotic normality of the BMPS estimators. We need the following assumptions for further development.

(A8) The first and second order derivatives (partial) of the marginal (conditional) distribution function $F_X(x;\gamma_1)$ ($F_{Y|X=x}(y;\theta)$) with respect to the parameter vector γ_1 (θ) exist in an open neighborhood of $\gamma_1^0(\theta^0)$, and are continuous jointly in (x,γ_1) (x,y,θ).

(A9) There exist measurable dominating functions $M_{1j}(x)$ and $M_{2jk}(x)$, such that for every θ in a neighborhood of θ^0,

$$\sup_{\theta \in N(\theta^0)} \left| \frac{\partial}{\partial \theta_j} \ln f_X(x;\theta) \right| \leq M_{1j}(x); \quad j = 1, \ldots, p,$$

with $E_{\theta^0}[M_{1j}(X)] < \infty$. Similarly, assume for every j, k,

$$\sup_{\theta \in N(\theta^0)} \left| \frac{\partial^2}{\partial \theta_j \partial \theta_k} \ln f_X(x;\theta) \right| \leq M_{2jk}(x); \quad j, k = 1, \ldots, p,$$

with $E_{\theta^0}[M_{2jk}(X)] < \infty$. We have analogous assumptions for the conditional density also, i.e. there exists measurable dominating function $M_{3jk}(x, y)$, such that for every θ in a neighborhood of θ^0,

$$\sup_{\theta \in N(\theta^0)} \left| \frac{\partial^2}{\partial \theta_j \partial \theta_k} \ln f_{Y|X}(y|x;\theta) \right| \leq M_{3jk}(x, y); \quad j, k = 1, \ldots, p,$$

with $E_{\theta^0}[M_{3jk}(X, Y)] < \infty$.

(A10) The Fisher information matrix

$$I(\theta^0) = E_{\theta^0}\left[-\frac{\partial}{\partial \theta \partial \theta^\top} \ln f(X, Y; \theta^0) \right]$$

exists, is finite and positive definite.

(A11) For every $j = 1, \ldots, p$,

$$E_{\theta^0}\left[\frac{\partial}{\partial \theta} \ln f_X(X;\theta^0) \right] = 0, \quad \text{and} \quad E_{\theta^0}\left[\frac{\partial}{\partial \theta} \ln f_{Y|X}(Y|X;\theta^0) \right] = 0,$$

for every $j, k = 1, \ldots, p$,

$$E_{\theta^0}\left[\left(\frac{\partial}{\partial \theta_j} \ln f_X(X;\theta^0) \right)^2 \right] < \infty, \quad \text{and} \quad E_{\theta^0}\left[\left(\frac{\partial}{\partial \theta_j} \ln f_{Y|X}(Y|X;\theta^0) \right)^2 \right] < \infty,$$

and

$$E_{\theta^0}\left| \frac{\partial^2}{\partial \theta_j \partial \theta_k} \ln f_X(X;\theta^0) \right| < \infty, \quad \text{and} \quad E_{\theta^0}\left| \frac{\partial^2}{\partial \theta_j \partial \theta_k} \ln f_{Y|X}(Y|X;\theta^0) \right| < \infty$$

Lemma 4: Suppose Assumptions (A1)–(A7) hold. Suppose $U_i(\theta)$'s, $U_{(i)}$'s V_i's, $V_{(i)}$'s $D_i^{(1)}(\theta)$'s and $D_i^{(2)}(\theta)$' are same as defined before. Then, uniformly over every compact subset of the parameter space,

$$(n+1)D_i^{(1)}(\theta) = \frac{f_X(x_i^*;\theta)}{f_X(x_i^*;\theta^0)} + o_p(1), \tag{12}$$

where x_i^* lies between the consecutive order statistics $X_{(i-1)}$ and $X_{(i)}$. Consequently,

$$\log D_i^{(1)}(\theta) = -\log(n+1) + \log \frac{f_X(x_i^*;\theta)}{f_X(x_i^*;\theta^0)} + o_p(1),$$

uniformly in θ. Similarly, uniformly over every compact subset of Θ,

$$(n+1)D_i^{(2)}(\theta) = \frac{f_{Y|X}(y_i^*|x_i^*;\theta)}{f_{Y|X}(y_i^*|x_i^*;\theta^0)} + o_p(1),$$

where (x_i^*, y_i^*) lies between two successive conditional order statistics. Consequently,

$$\log D_i^{(2)}(\theta) = -\log(n+1) + \log \frac{f_{Y|X}(y_i^*|x_i^*;\theta)}{f_{Y|X}(y_i^*|x_i^*;\theta^0)} + o_p(1),$$

uniformly in θ.

Proof: Conditional on the observed X_i's, the variables V_i are Uniform(0,1). By conditioning on the observed X's and applying Theorem 2 of Shao [16], the expansion holds almost surely conditionally. Since the approximation is uniform and the remainder is $o_p(1)$, conditionally, the same order remains after removing the conditioning. Therefore, the conditional spacing process is identical to the ordinary univariate spacing process studied by Shao [16], and the same expansion holds holds conditionally. Integrating over the distribution of X, yields the unconditional result. ∎

Lemma 5: Suppose Assumptions (A1)–(A11) hold. Let

$$S_n(\theta) = \frac{\partial}{\partial\theta} H_n(\theta)$$

denote the gradient of the BMPS objective function with respect to the parameter vector θ, and let

$$\ell(X, Y; \theta) = \frac{\partial}{\partial\theta} \ln f_X(X;\theta) + \frac{\partial}{\partial\theta} \ln f_{Y|X}(Y|X;\theta)$$

denote the corresponding score function. Then, at the true parameter value θ^0,

$$S_n(\theta^0) = \frac{1}{n}\sum_{i=1}^{n} \ell(X_i, Y_i; \theta^0) + o_p(n^{-1/2}).$$

Equivalently,

$$\sqrt{n}\, S_n(\theta^0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \ell(X_i, Y_i; \theta^0) + o_p(1).$$

Proof: From Lemma 4, uniformly over a neighborhood of θ^0,

$$\ln D_i^{(1)}(\theta) = -\log(n+1) + \ln \frac{f_X(x_i^*; \theta)}{f_X(x_i^*; \theta^0)} + o_p(1),$$

where x_i^* lies between the consecutive order statistics $X_{(i-1)}$ and $X_{(i)}$. Differentiating both sides with respect to θ gives

$$\frac{\partial}{\partial\theta} \ln D_i^{(1)}(\theta) = \frac{\partial}{\partial\theta} \ln f_X(x_i^*; \theta) + o_p(1),$$

uniformly in a neighborhood of θ^0. Evaluating at $\theta = \theta^0$,

$$\frac{\partial}{\partial\theta} \ln D_i^{(1)}(\theta^0) = \frac{\partial}{\partial\theta} \ln f_X(x_i^*; \theta^0) + o_p(1).$$

Similarly,

$$\frac{\partial}{\partial\theta} \ln D_i^{(2)}(\theta^0) = \frac{\partial}{\partial\theta} \ln f_{Y|X}(y_i^* \mid x_i^*; \theta^0) + o_p(1),$$

where (x_i^*, y_i^*) lies between two successive conditional order statistics. Therefore,

$$S_n(\theta^0) = \frac{1}{n+1}\sum_{i=1}^{n+1} \left\{ \frac{\partial}{\partial\theta} \ln D_i^{(1)}(\theta^0) + \frac{\partial}{\partial\theta} \ln D_i^{(2)}(\theta^0) \right\}.$$

Substituting the above expansions yields

$$S_n(\theta^0) = \frac{1}{n}\sum_{i=1}^{n} \left\{ \frac{\partial}{\partial\theta} \ln f_X(X_i; \theta^0) + \frac{\partial}{\partial\theta} \ln f_{Y|X}(Y_i \mid X_i; \theta^0) \right\} + o_p(n^{-1/2}),$$

where we have used the fact that replacing the intermediate points x_i^* and (x_i^*, y_i^*) by the observations X_i and (X_i, Y_i) introduces only an $o_p(n^{-1/2})$ error under the smoothness assumptions. This completes the proof. ∎

Lemma 6: Suppose Assumptions (A1)–(A11) hold. Let $S_n(\theta)$ be the same as defined before, then,

$$\sqrt{n}\, S_n(\theta^0) \xrightarrow{D} N_p\left(\mathbf{0},\, I(\theta^0)\right),$$

where $I(\theta^0)$ is same as defined in A10.

Proof: From Lemma 5,

$$S_n(\theta^0) = \frac{1}{n}\sum_{i=1}^{n} \ell(X_i, Y_i; \theta^0) + o_p(n^{-1/2}).$$

Therefore,

$$\sqrt{n}\, S_n(\theta^0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \ell(X_i, Y_i; \theta^0) + o_p(1).$$

Now, under Assumption (A11),

$$E_{\theta^0}\left[\ell(X, Y; \theta^0)\right] = \mathbf{0},$$

and

$$E_{\theta^0}\left[\ell(X, Y; \theta^0)\ell(X, Y; \theta^0)^T\right] = I(\theta^0),$$

where the covariance matrix $I(\theta^0)$ is finite and positive definite. Since $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent and identically distributed, it follows that $\ell(X_1, Y_1; \theta^0), \ldots, \ell(X_n, Y_n; \theta^0)$ are also independent and identically distributed random vectors with mean $\mathbf{0}$ and covariance matrix $I(\theta^0)$. Hence, by the multivariate Central Limit Theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \ell(X_i, Y_i; \theta^0) \xrightarrow{D} N_p(\mathbf{0}, I(\theta^0)).$$

Finally, the desired result follows immediately from Slutsky's theorem since

$$o_p(1) \xrightarrow{P} 0.$$

This completes the proof. ∎

Lemma 7: Suppose Assumptions (A1)–(A11) hold. Let

$$J_n(\theta) = \frac{\partial^2}{\partial\theta\partial\theta^\top} H_n(\theta)$$

denote the Hessian matrix of the BMPS objective function with respect to the parameter vector θ. Then, uniformly over every neighborhood of θ^0,

$$J_n(\theta) \xrightarrow{P} J(\theta),$$

where

$$J(\theta) = E_{\theta^0}\left[\frac{\partial^2}{\partial\theta\partial\theta^\top}\ln f_X(X;\theta) + \frac{\partial^2}{\partial\theta\partial\theta^\top}\ln f_{Y|X}(Y|X;\theta)\right].$$

In particular,

$$J_n(\theta^0) \xrightarrow{P} -I(\theta^0),$$

where $I(\theta^0)$ is the Fisher information matrix as defined in A(10).

Proof: From Lemma 4, uniformly over a neighborhood of θ^0,

$$\frac{\partial^2}{\partial\theta\partial\theta^\top}\ln D_i^{(1)}(\theta) = \frac{\partial^2}{\partial\theta\partial\theta^\top}\ln f_X(x_i^*;\theta) + o_p(1),$$

where x_i^* lies between the consecutive order statistics $X_{(i-1)}$ and $X_{(i)}$. Similarly,

$$\frac{\partial^2}{\partial\theta\partial\theta^\top}\ln D_i^{(2)}(\theta) = \frac{\partial^2}{\partial\theta\partial\theta^\top}\ln f_{Y|X}(y_i^* \mid x_i^*;\theta) + o_p(1),$$

where (x_i^*, y_i^*) lies between two successive conditional order statistics. Hence,

$$J_n(\theta) = \frac{1}{n+1}\sum_{i=1}^{n+1}\left[\frac{\partial^2}{\partial\theta\partial\theta^\top}\ln D_i^{(1)}(\theta) + \frac{\partial^2}{\partial\theta\partial\theta^\top}\ln D_i^{(2)}(\theta)\right].$$

Substituting the above expansions gives

$$J_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\theta\partial\theta^\top}\ln f_X(X_i;\theta) + \frac{\partial^2}{\partial\theta\partial\theta^\top}\ln f_{Y|X}(Y_i|X_i;\theta)\right] + o_p(1),$$

uniformly over a neighborhood of θ^0.

By Assumptions (A8)–(A11), the second-order derivatives are continuous in θ and are dominated by integrable functions. Therefore, the Uniform Law of Large Numbers yields

$$\sup_\theta \|J_n(\theta) - J(\theta)\| \xrightarrow{P} 0,$$

where

$$J(\theta) = E_{\theta^0}\left[\frac{\partial^2}{\partial\theta\partial\theta^\top}\ln f_X(X;\theta) + \frac{\partial^2}{\partial\theta\partial\theta^\top}\ln f_{Y|X}(Y|X;\theta)\right].$$

Finally, evaluating at $\theta = \theta^0$,

$$J(\theta^0) = E_{\theta^0}\left[\frac{\partial^2}{\partial\theta\partial\theta^\top}\ln f(X,Y;\theta^0)\right] = -I(\theta^0).$$

Hence,

$$J_n(\theta^0) \xrightarrow{P} -I(\theta^0).$$

This completes the proof. ∎

Lemma 8: Suppose Assumptions (A1)–(A11) hold and let $\widehat{\theta}_n$ denote the BMPS estimator of θ^0. Then there exists a random vector $\widetilde{\theta}_n$ lying on the line segment joining $\widehat{\theta}_n$ and θ^0 such that

$$\mathbf{0} = S_n(\widehat{\theta}_n) = S_n(\theta^0) + J_n(\widetilde{\theta}_n)(\widehat{\theta}_n - \theta^0),$$

where $S_n(\theta)$ and $J_n(\theta)$ denote the score vector and the Hessian matrix of the BMPS objective function, respectively, as defined before. Consequently,

$$\sqrt{n}(\widehat{\theta}_n - \theta^0) = -J_n^{-1}(\widetilde{\theta}_n)\sqrt{n}\,S_n(\theta^0),$$

provided $J_n(\widetilde{\theta}_n)$ is nonsingular.

Proof Since $\widehat{\theta}_n$ maximizes the objective function $H_n(\theta)$ over the parameter space Θ and θ^0 is an interior point of Θ, the first-order optimality condition gives $S_n(\widehat{\theta}_n) = \mathbf{0}$. Since $H_n(\theta)$ is twice continuously differentiable in a neighborhood of θ^0 by Assumptions (A8)–(A11), the multivariate Taylor expansion of $S_n(\theta)$ about θ^0 yields

$$S_n(\widehat{\theta}_n) = S_n(\theta^0) + J_n(\widetilde{\theta}_n)(\widehat{\theta}_n - \theta^0),$$

where $\widetilde{\theta}_n$ lies on the line segment joining θ^0 and $\widehat{\theta}_n$. Since $S_n(\widehat{\theta}_n) = \mathbf{0}$ we obtain

$$\mathbf{0} = S_n(\theta^0) + J_n(\widetilde{\theta}_n)(\widehat{\theta}_n - \theta^0).$$

Assuming that $J_n(\widetilde{\theta}_n)$ is nonsingular, we obtain

$$\widehat{\theta}_n - \theta^0 = -J_n^{-1}(\widetilde{\theta}_n) S_n(\theta^0).$$

Multiplying both sides by $\sqrt{n}$ gives

$$\sqrt{n}(\widehat{\theta}_n - \theta^0) = -J_n^{-1}(\widetilde{\theta}_n)\sqrt{n}\, S_n(\theta^0),$$

which completes the proof. ∎

Theorem 2: Suppose Assumptions (A1)–(A11) hold. Let

$$\widehat{\theta}_n = \arg\max_{\theta \in \Theta} H_n(\theta)$$

denote the BMPS estimator of the true parameter θ^0. Then,

$$\sqrt{n}\,(\widehat{\theta}_n - \theta^0) \xrightarrow{D} N_p\left(\mathbf{0}, I^{-1}(\theta^0)\right),$$

where $I(\theta^0)$ is the Fisher information matrix, as defined in A(10).

Proof Since $\widehat{\theta}_n$ maximizes the BMPS objective function $S_n(\widehat{\theta}_n) = \mathbf{0}$, by Lemma 8, there exists a random vector $\widetilde{\theta}_n$ lying on the line segment joining $\widehat{\theta}_n$ and θ^0 such that

$$\mathbf{0} = S_n(\theta^0) + J_n(\widetilde{\theta}_n)(\widehat{\theta}_n - \theta^0).$$

Therefore,

$$\sqrt{n}(\widehat{\theta}_n - \theta^0) = -J_n^{-1}(\widetilde{\theta}_n)\sqrt{n}\, S_n(\theta^0).$$

Since, $\widehat{\theta}_n \xrightarrow{P} \theta^0$, it implies $\widetilde{\theta}_n \xrightarrow{P} \theta^0$. Now, by Lemma 7, $J_n(\widetilde{\theta}_n) \xrightarrow{P} -I(\theta^0)$, and therefore, by the continuous mapping theorem, $J_n^{-1}(\widetilde{\theta}_n) \xrightarrow{P} -I^{-1}(\theta^0)$. Further, by Lemma 6,

$$\sqrt{n}\, S_n(\theta^0) \xrightarrow{D} N_p\left(\mathbf{0}, I(\theta^0)\right).$$

Finally, an application of Slutsky's theorem yields

$$\sqrt{n}(\widehat{\theta}_n - \theta^0) \xrightarrow{D} I^{-1}(\theta^0)\, Z,$$

where $Z \sim N_p\left(\mathbf{0}, I(\theta^0)\right)$. Since

$$\text{Var}\left(I^{-1}(\theta^0)Z\right) = I^{-1}(\theta^0)I(\theta^0)I^{-1}(\theta^0) = I^{-1}(\theta^0),$$

it follows that

$$\sqrt{n}(\widehat{\theta}_n - \theta^0) \xrightarrow{D} N_p\left(\mathbf{0}, I^{-1}(\theta^0)\right).$$

This completes the proof. ∎

4 Bivariate MPS Estimators of BBBE

,In this section we implement the bivariate MPS estimators for the BBBW model. We need the following results for further development. We have the following results, whose proofs can be obtained in a routine manner.

Theorem 3: Let $(X, Y) \sim \text{BBBW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$, then the marginal and conditional PDFs are as follows:

$$\begin{aligned}
f_X(x;\alpha,\lambda_1,\lambda_2,\lambda_3) &= \frac{\lambda_1+\lambda_2+\lambda_3}{\lambda_1+\lambda_2} f_{WE}(x;\alpha,\lambda_1+\lambda_3) - \frac{\lambda_3}{\lambda_1+\lambda_2} f_{WE}(x;\alpha,\lambda_1+\lambda_2+\lambda_3).\\
f_Y(y;\alpha,\lambda_1,\lambda_2,\lambda_3) &= \frac{\lambda_1+\lambda_2+\lambda_3}{\lambda_1+\lambda_2} f_{WE}(y;\alpha,\lambda_2+\lambda_3) - \frac{\lambda_3}{\lambda_1+\lambda_2} f_{WE}(y;\alpha,\lambda_1+\lambda_2+\lambda_3).\\
f_{X|Y=y}(x;\alpha,\lambda_1,\lambda_2,\lambda_3) &= \begin{cases} \frac{\lambda_2+\lambda_3}{\lambda_2+\lambda_3(1-e^{-\lambda_1 y^\alpha})} f_{WE}(x;\alpha,\lambda_1) & \text{if} \quad x<y \\ \frac{\lambda_2}{e^{-\lambda_3 y^\alpha}(\lambda_2+\lambda_3(1-e^{-\lambda_1 y^\alpha}))} f_{WE}(x;\alpha,\lambda_1+\lambda_3) & \text{if} \quad x>y. \end{cases}\\
f_{Y|X=x}(y;\alpha,\lambda_1,\lambda_2,\lambda_3) &= \begin{cases} \frac{\lambda_1+\lambda_3}{\lambda_1+\lambda_3(1-e^{-\lambda_2 x^\alpha})} f_{WE}(y;\alpha,\lambda_2) & \text{if} \quad x>y \\ \frac{\lambda_1}{e^{-\lambda_3 x^\alpha}(\lambda_1+\lambda_3(1-e^{-\lambda_2 x^\alpha}))} f_{WE}(y;\alpha,\lambda_2+\lambda_3) & \text{if} \quad x<y. \end{cases}
\end{aligned}$$

Theorem 4: Let $(X, Y) \sim \text{BBBW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$, then the marginal and conditional CDFs are as follows:

$$\begin{aligned}
F_X(x;\alpha,\lambda_1,\lambda_2,\lambda_3) &= \frac{\lambda_1+\lambda_2+\lambda_3}{\lambda_1+\lambda_2} F_{WE}(x;\alpha,\lambda_1+\lambda_3) - \frac{\lambda_3}{\lambda_1+\lambda_2} F_{WE}(x;\alpha,\lambda_1+\lambda_2+\lambda_3).\\
F_Y(y;\alpha,\lambda_1,\lambda_2,\lambda_3) &= \frac{\lambda_1+\lambda_2+\lambda_3}{\lambda_1+\lambda_2} F_{WE}(y;\alpha,\lambda_2+\lambda_3) - \frac{\lambda_3}{\lambda_1+\lambda_2} F_{WE}(y;\alpha,\lambda_1+\lambda_2+\lambda_3).
\end{aligned}$$

$$F_{X|Y=y}(x;\alpha,\lambda_1,\lambda_2,\lambda_3) = \begin{cases} \frac{\lambda_2+\lambda_3}{\lambda_2+\lambda_3(1-e^{-\lambda_1 y^\alpha})} F_{WE}(x;\alpha,\lambda_1) & \text{if} \quad x<y \\ \frac{\lambda_2}{e^{-\lambda_3 y^\alpha}(\lambda_2+\lambda_3(1-e^{-\lambda_1 y^\alpha}))} F_{WE}(x;\alpha,\lambda_1+\lambda_3) & \text{if} \quad x>y. \end{cases}$$

$$F_{Y|X=x}(y;\alpha,\lambda_1,\lambda_2,\lambda_3) = \begin{cases} \frac{\lambda_1+\lambda_3}{\lambda_1+\lambda_3(1-e^{-\lambda_2 x^\alpha})} F_{WE}(y;\alpha,\lambda_2) & \text{if} \quad x>y \\ \frac{\lambda_1}{e^{-\lambda_3 x^\alpha}(\lambda_1+\lambda_3(1-e^{-\lambda_2 x^\alpha}))} F_{WE}(y;\alpha,\lambda_2+\lambda_3) & \text{if} \quad x<y. \end{cases}$$

Now we describe how we can implement BMPS in case of the MOBW model. Here $\theta = (\alpha,\lambda_1,\lambda_2,\lambda_3)$, and we have a sample of size n, $\mathcal{D} = \{(x_1,y_1),\ldots,(x_n,y_n)\}$ from MOBW(θ). Let us define

$$u_i(\theta) = F_X(x_i;\theta) \quad \text{and} \quad v_i(\theta) = F_{Y|X=x_i}(y_i;\theta); \quad i=1,\ldots,n.$$

Now order $u_1(\theta),\ldots,u_n(\theta)$ and $v_1(\theta),\ldots,v_n(\theta)$ as

$$u_{(1)}(\theta) < \ldots < u_{(n)}(\theta) \quad \text{and} \quad v_{(1)}(\theta) < \ldots < v_{(n)}(\theta),$$

respectively. Let

$$d_i^1(\theta) = u_{(i)}(\theta) - u_{(i-1)}(\theta) \quad \text{and} \quad d_i^2(\theta) = v_{(i)}(\theta) - v_{(i-1)}(\theta),$$

where $u_{(0)}(\theta) = v_{(0)}(\theta) = 0$ and $u_{(n+1)}(\theta) = v_{(n+1)}(\theta) = 1$. If we write

$$\widetilde{h}_1(\theta) = \frac{1}{n+1}\sum_{i=1}^{n+1} \ln d_i^1(\theta), \quad \widetilde{h}_2(\theta) = \frac{1}{n+1}\sum_{i=1}^{n+1} \ln d_i^2(\theta)$$

and

$$\widetilde{h}(\theta) = \widetilde{h}_1(\theta) + \widetilde{h}_2(\theta),$$

then we compute the BMPS of θ, say $\widehat{\theta}$ as

$$\widehat{\theta} = \arg\max_{\theta\in\Theta} \ \widetilde{h}(\theta). \tag{13}$$

The solution cannot be obtained in explicit forms. Once needs to use some optimization technique to compute the BMPS of the BBBW model parameters. It becomes a four dimensional optimization problem. If we assume α is known, say for example $\alpha = 1$, then it becomes BBBE model or if we assume $\alpha = 2$, then it becomes Block and Basu bivariate Rayleigh (BBBR) model. In case

of BBBE model or BBBR model, we need to solve a three dimensional optimization problem to compute BMPS estimators.

Note that although, we have developed the methodology based on the marginal of X and the conditional of Y given X, similar methodology can be developed based on the marginal of Y and the conditional of X given Y. It is also possible to combine the two.

Comments: In a routine way, it can be verified that the BBBW model satisfies Assumptions (A1) to (A10). Hence, the BMPS estimators of the BBBW model are consistent and asymptotically normally distributed.

5 Numerical Results

In this section we perform some simulation experiments to show how the proposed method works in practice. We have considered three different cases namely (i) BBBE, (ii) BBBR and (iii) BBBW models with the following set of parameters: (i) $\lambda_1 = \lambda_2 = \lambda_3 = 1.0$, (ii) $\lambda_1 = \lambda_2 = \lambda_3 = 1.0$, and (iii) $\lambda_1 = \lambda_2 = \lambda_3 = 1.0$, $\alpha = 2.0$. We have taken different sample sizes namely; n = 25,50,75 and 100. In each case we obtain the BMPS estimators and MLEs of the unknown parameters and compute the average estimates and the corresponding mean squared errors (MSEs) based on 1000 replications. The results are reported in Table 1, Table 3 and Table **??** respectively. We have also calculated the aymptotic confidence intervals and the coverage percentages of the unknown parameters based on both the estimators. The results are reported in Table 2, Table 4 and Table 6, respectively.

Some of the points are quite clear from these simulation experiments. It has been observed that in all the cases as the sample size increases the average biases and the MSEs decrease. It verifies the consistency properties of the BMPS estimators in these three cases. Comparing Table **??** and Table **??** it has been observed that the average biases and the MSEs are more or less same.

Table 1: The average estimates and the corresponding MSEs of (a) BMPS estimators and (b) MLEs of λ_1, λ_2 and λ_3 for BBBE distribution.

	(a)				(b)		
n	λ_1	λ_2	λ_3	n	λ_1	λ_2	λ_3
25	1.0256	1.0356	1.0431	25	0.6901	0.7214	0.7117
	(0.1234)	(0.1321)	(0.1345)		(0.943)	(0.945)	(0.943)
50	1.0221	1.0289	1.0367	50	0.4995	0.5212	0.5016
	(0.0675)	(0.0731)	(0.0778)		(0.946)	(0.952)	(0.946)
75	1.0113	1.0176	1.0211	75	0.4412	0.4398	0.4423
	(0.0501)	(0.0498)	(0.0478)		(0.951)	(0.950)	(0.952)
100	1.0031	1.0078	1.0101	100	0.3512	0.3498	0.3513
	(0.0312)	(0.0367)	(0.0389)		(0.951)	(0.949)	(0.950)

It is not very surprising as computing the BMPS estimators of BBBE and BBBR involve solving a three dimensional optimization problem. Now Comparing Table **??** and Table **??** it has been observed that the average biases and the MSEs for Table **??** are slightly more than those of Table **??**. This is due to the reason that to compute the BMPS estimators of BBBW one needs to solve a four dimensional optimization problem. It seems that the method is working fine for all these models.

6 Data Analysis

In this section we analyze one data set for illustrative purposes. This data set has been obtained from Johnson and Wichern [7] (page 271). It represents the bone mineral density (BMD) measured in g/cm^2 for 24 children one year after their participation in an experimental program. These bivariate data represent the BMD for Dominant Ulna and Ulna bones and they are as follows;

Data set: (0.869 0.964), (0.602 0.689), (0.765 0.738), (0.761 0.698), (0.551 0.619), (0.753 0.515), (0.708 0.787), (0.687 0.715), (0.844 0.656), (0.869 0.789), (0.654 0.726), (0.692 0.526), (0.670 0.580), (0.823 0.773), (0.746 0.729), (0.656 0.506), (0.693 0.740), (0.883 0.785), (0.577 0.627),

Table 2: The average length of the confidence intervals and the associated coverage percetages of (a) BMPS estimators and (b) MLEs of λ_1, λ_2 and λ_3 for BBBE distribution.

	(a)				(b)		
n	λ_1	λ_2	λ_3	n	λ_1	λ_2	λ_3
25	0.6885	0.7110	0.7223	25	1.0249	1.0401	1.0397
	(0.947)	(0.0948)	(0.946)		(0.1257)	(0.1298)	(0.1367)
50	0.5092	0.5123	0.4987	50	1.0197	1.0301	1.0387
	(0.0949)	(0.0955)	(0.948)		(0.0687)	(0.0699)	(0.0773)
75	0.4389	0.4423	0.4298	75	1.0117	1.0201	1.0202
	(0.954)	(0.952)	(0.951)		(0.0598)	(0.0523)	(0.0501)
100	0.3467	0.3512	0.3501	100	1.0047	1.0065	1.0099
	(0.951)	(0.950)	(0.951)		(0.0297)	(0.0401)	(0.0411)

(0.802 0.769), (0.540 0.498), (0.804 0.779), (0.570 0.634), (0.585 0.640).

We have provided the scatter plot of the data set in Figure 1. Just to get an idea about the hazard functions of the marginals we provide the scaled total time on test (TTT) plot as suggested by Aarset [1], which provides an idea of the shape of the hazard function of a distribution. It has been shown by Aarset [1] that the scaled TTT transform is convex (concave) if the hazard rate is decreasing (increasing) and for bathtub (unimodal) shaped hazard rate, the scaled TTT transform is first convex (concave) and then concave (convex). In this case the scale TTT transform of BMDs of Dominant radius and Radius bones are given in Fig. 2. It clearly indicates that both the marginals have increasing hazard rates. The correlation coefficient between the two variables is 0.628 and that shows they are positively correlated. Therefore it is not unreasonable to fit the BBBW model in this case. We have fitted the BBBW model to the data set and obtained the BMPS estimators of the unknown parameters as follows: $\widehat{\alpha} = 7.1687$, $\widehat{\lambda}_1 = 3.5967$, $\widehat{\lambda}_2 = 5.2018$ and $\widehat{\lambda}_3 = 8.0312$. The 95% parametric bootstrap confidence intervals of α, λ_1, λ_2 and λ_3 are (6.2342, 8.1123), (2.0467, 5.1999), (3.2114, 7.1421) and (5.9621, 10.1134), respectively. We have also presented the 95% asymptotic confidence intervals of α, λ_1, λ_2 and λ_3, based on the observed Fisher information matrix as (6.1564,8.1810), (1.1836,6.0098),(3.2005,7.2031),(5.7101,10.3523), re-

Table 3: The average estimates and the corresponding MSEs of (a) BMPS estimators and (b) MLEs of λ_1, λ_2 and λ_3 for BBBR distribution.

(a)

n	λ_1	λ_2	λ_3
25	1.0249	1.0298	1.0417
	(0.1212)	(0.1299)	(0.1357)
50	1.0247	1.0267	1.0299
	(0.0675)	(0.0664)	(0.0701)
75	1.0165	1.0141	1.0201
	(0.0489)	(0.0512)	(0.0501)
100	1.0017	1.0066	1.0099
	(0.0289)	(0.0378)	(0.0311)

(b)

n	λ_1	λ_2	λ_3
25	1.0253	1.0273	1.0399
	(0.1233)	(0.1308)	(0.1398)
50	1.0233	1.0301	1.0305
	(0.0701)	(0.0698)	(0.0689)
75	1.0155	1.0142	1.0199
	(0.0467)	(0.0473)	(0.0511)
100	1.0023	1.0057	1.0106
	(0.0293)	(0.0375)	(0.0399)

spectively.

7 Conclusions

In this paper we have proposed a very natural bivariate maximum product of spacings estimators. It is an extension of the univariate maximum product of spacings estimators. It is mainly based on the marginals and the conditional decomposition of the joint probability density function and the probability integral transformation method. The implementation of the proposed BMPS estimators is quite straight forward. We have established the consistency and the asymptotic normality properties of the proposed estimators under a fairly general set of assumptions. Although, it has been developed mainly for the bivariate model, it can be generalized even to the multivariate case. It will be interesting to develop the necessary properties of the multivariate model. More work is needed in that direction.

Author Contributions:

The author DK has contributed 100% in this paper.

Table 4: The average length of the confidence intervals and the associated coverage percetages of (a) BMPS estimators and (b) MLEs of λ_1, λ_2 and λ_3 for BBBR distribution.

(a)

n	λ_1	λ_2	λ_3
25	0.6824	0.6911	0.6856
	(0.943)	(0.945)	(0.945)
50	0.4825	0.4878	0.4779
	(0.944)	(0.948)	(0.947)
75	0.3939	0.4040	0.4141
	(0.952)	(0.951)	(0.950)
100	0.3411	0.3423	0.3476
	(0.951)	(0.950)	(0.951)

(b)

n	λ_1	λ_2	λ_3
25	0.6896	0.6886	0.7010
	(0.943)	(0.948)	(0.948)
50	0.4901	0.4904	0.4911
	(0.951)	(0.952)	(0.947)
75	0.4010	0.4021	0.4210
	(0.953)	(0.948)	(0.948)
100	0.3423	0.3444	0.3413
	(0.950)	(0.951)	(0.949)

Funding:

The author has not received any funding for preparing this paper.

Acknowlwdgements

The author would like to thank two unknown reviewers for their constructive suggestions which have helped to improve the paper significantly

References

[1] Aarset, M.V. (1987), "How to identify a bathtub shaped hazard rate?", *IEEE Transactions on Reliability*, vol. 36, 106–108.

[2] Block, H.W. and Basu, A.P. (1974), "A continuous bivariate exponential extension", *Journal of the American Statistical Association*, vol. 69, 1031–1037.

Table 5: The average estimates and the corresponding MSEs of (a) BMPS estimators and (b) MLEs of α, λ_1, λ_2 and λ_3 for BBBW distribution.

(a)

n	λ_1	λ_2	λ_3	α
25	1.0451	1.0411	1.0654	2.1412
	(0.2357)	(0.2517)	(0.3621)	(0.5645)
50	1.0376	1.0378	1.0451	2.1011
	(0.1486)	(0.1452)	(0.1701)	(0.3011)
75	1.0267	1.0211	1.0376	2.0891
	(0.0898)	(0.0897)	(0.0817)	(0.1967)
100	1.0156	1.0123	1.0201	2.0312
	(0.0667)	(0.0611)	(0.0623)	(0.1267)

(b)

n	λ_1	λ_2	λ_3	α
25	1.0398	1.0406	1.0599	2.1217
	(0.2412)	(0.2601)	(0.3598)	(0.5498)
50	1.0398	1.0412	1.0513	2.1213
	(0.1498)	(0.1567)	(0.2003)	(0.2987)
75	1.0312	1.0254	1.0415	2.0786
	(0.0876)	(0.0915)	(0.1013)	(0.1917)
100	1.0136	1.0451	1.0187	2.0651
	(0.0596)	(0.0613)	(0.0665)	(0.1287)

[3] Cheng, R.C.H. and Amin, N.A.K. (1983), "Estimating Parameters in Continuous Univariate Distributions with a Shifted Origin", *Journal of the Royal Statistical Society: Series B* (Methodological), vol. 45, pages 394–403.

[4] Darling, D.A. (1953), "On a class of problems related to the random division of an interval", *Annals of Mathematical Statistics*, vol. 24, 239 – 253.

[5] Ekstrom, M. (2008), "Alternatives to maximum likelihood estimation based on spacings and the Kullback-Leibler divergence", *Journal of Statistical Planning and Inference*, Volume 138, Issue 6, 1778–1791.

[6] Ferguson, T.S. (1982), "An inconsistent maximum likelihood estimate", *Journal of the American Statistical Association*, vol. 77, 831 – 834.

[7] Johnson, R.A. and Wichern, D.W. (1999), *Applied Multivariate Statistical Analysis*, 7th Ed., Prentice Hall, New Jersey.

[8] Khan, M.S.R., Hussain, Z., Sher, S., Amjad, M., Baig, F., Ahmad, I. and Ullah, H. (2025), "A comparative analysis of L-moments, maximum likelihood, and maximum

Table 6: The average length of the confidence intervals and the associated coverage percetages of (a) BMPS estimators and (b) MLEs of α, λ_1, λ_2 and λ_3 for BBBW distribution.

(a)

n	λ_1	λ_2	λ_3	α
25	0.9516	0.9612	0.9589	1.4726
	(0.943)	(0.942)	(0.941)	(0.947)
50	0.6728	0.6687	0.6712	1.0423
	(0.946)	(0.947)	(0.947)	(0.948)
75	0.5494	0.5517	0.5487	0.8489
	(0.949)	(0.949)	(0.943)	(0.949)
100	0.4758	0.4801	0.4789	0.7364
	(0.951)	(0.950)	(0.952)	(0.951)

(b)

n	λ_1	λ_2	λ_3	α
25	0.9611	0.9623	0.9598	1.5123
	(0.944)	(0.943)	(0.942)	(0.941)
50	0.6801	0.6798	0.6811	1.0545
	(0.943)	(0.945)	(0.945)	(0.942)
75	0.5487	0.5552	0.5512	0.8510
	(0.948)	(0.948)	(0.947)	(0.947)
100	0.4812	0.4823	0.4812	0.7421
	(0.956)	(0.955)	(0.950)	(0.952)

product of spacing methods for the four-parameter kappa distribution in extreme value analysis", *Sci. Rep. 15*, vol. 164. https://doi.org/10.1038/s41598-024-84056-1

[9] Kuljus, K. and Ranneby, B. (2015), "Generalized maximum spacings estimation for multivariate observations", *Scandinavian Journal of Statistics*, vol. 42, 1092 – 1108.

[10] Kullback, S. and Leibler, R.A. (1951), "On information and sufficiency", *Annals of Mathematical Statistics*, vol. 22, no. 1, 79 – 86.

[11] Kundu, D. (2023), "Bivariate distributions with singular components", *Springer Handbook of Engineering Statistics*, Editor: Hoang Pham, Springer, pp 733 - 761.

[12] Kundu, D. and Gupta, R.D. (2010), "A class of absolute continuous bivariate distribution", *Statistical Methodology*, vol. 7, 464–477.

[13] Marshall,A.W. and Olkin, I. (1967),"A multivariate exponential distribution", *Journal of the American Statistical Association*, vol. 62, 30–44.

[14] Ranneby, B. (1984), "The maximum spacing method. An estimation method related to the maximum likelihood method", *Scandinavian Journal of Statistics*, vol. 11, 93–112.

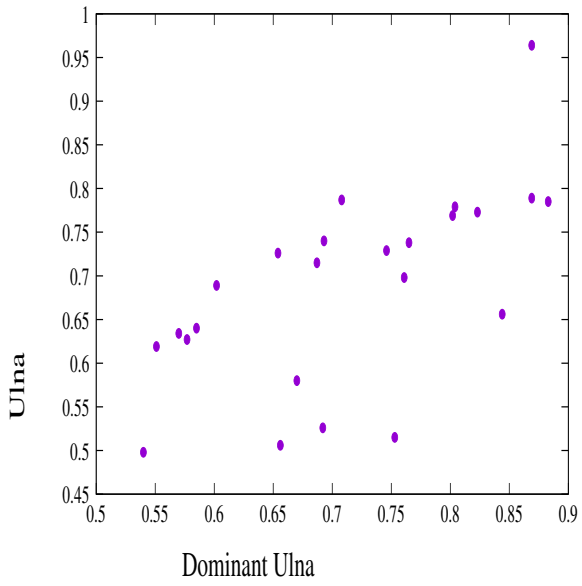


Figure 1: The scatter plot of the bone minerals data of 24 children after one year of their birth.

[15] Ranneby, B., Jammalamadaka, S.R. and Teterukovskiy, A. (2005), "The maximum spacing estimation for multivariate observations", *Journal of Statistical Planning and Inference*, vol. 129, 427–446.

[16] Shao, Y. (2001), "Consistency of the maximum product spacings method and estimation of unimodal distribution", *Statistica Sinica*, vol. 11, 1125 – 1140.

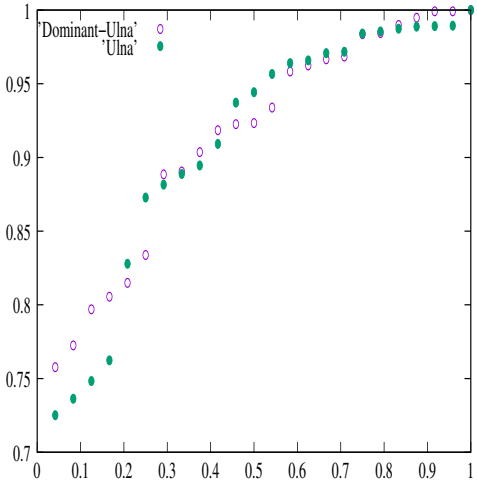


Figure 2: The scaled TTT transforms of the bone minerals data.