Divergence-Based Robust Inference for the Marshall-Olkin Bivariate Exponential Distribution

Sanjay Kumar[1, a*], Debasis Kundu[2, a] and Sharmishtha Mitra[3, a]

[a]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, Pin 208016, India.

*Corresponding author(s). E-mail(s): [1]sanjaymk@iitk.ac.in;
Contributing authors: [2]kundu@iitk.ac.in; [3]smitra@iitk.ac.in;

Abstract

Statistical modeling of bivariate data sets with ties is an active research area in the literature. The most common model fitted to such data sets is the Marshall-Olkin bivariate exponential (MOBE) distribution proposed by Marshall and Olkin [1]. As bivariate data sets often contain outliers and maximum likelihood estimation (MLE) for the MOBE model is highly sensitive to outliers, it is crucial to adopt a robust parameter estimation procedure. In this article, we extend the widely adopted minimum density power divergence estimation (MDPDE) method, introduced by Basu et al. [2], for robust and efficient estimation of the MOBE parameters. The MDPDE method is a robust generalization of the MLE and is widely applicable in practice. We derive analytical expressions for the estimating equations and the asymptotic distributions of the MDPDE. We compare the asymptotic covariance of MDPDE with the Fisher information matrix. Besides, we study the asymptotic relative efficiency of MDPDE. The influence function of MDPDE for the MOBE model is bounded, which ensures the robustness of MDPDE. We discuss a data-driven procedure for the optimal tuning parameter selection. We apply the proposed technique to some simulated studies and a real data set. The results indicate superior performance of the MDPDE compared to the MLE, especially in the presence of outliers.

Keywords: Marshall-Olkin bivariate exponential distribution, Maximum likelihood, Outliers, Density power divergence, Robust and efficient estimation, Optimal tuning parameter.

1 Introduction

Statistical modeling of bivariate data sets, where for some observations both components have same values, i.e., with ties is a crucial and active research area in various disciplines, such as medical sciences, life-testing experiments, environmental studies, reliability engineering, and survival analysis. For example, the Union of European Football Associations (UEFA) Champions League football data, originally analyzed by Meintanis [3], consist of 37 matches in which the home team scored at least one goal and at least one goal was scored from a direct kick, such as a penalty kick, free kick, or foul kick, by either team. The data have a bivariate structure in which ties arise naturally, when the first kick goal is scored by the home team. The most common probability distribution for modeling such data in practice is the Marshall-Olkin bivariate exponential (MOBE) distribution, introduced by Marshall and Olkin [1]. The joint probability density function (PDF) of the MOBE distribution, as given in [1], with parameters $\alpha_1, \alpha_2, \alpha_3 > 0$ is given by

$$f(x_1, x_2; \theta) = \begin{cases} \alpha_1(\alpha_2 + \alpha_3) e^{-\alpha_1 x_1} e^{-(\alpha_2+\alpha_3)x_2} & \text{if } x_1 < x_2 \\ (\alpha_1 + \alpha_3)\alpha_2 e^{-(\alpha_1+\alpha_3)x_1} e^{-\alpha_2 x_2} & \text{if } x_1 > x_2 \\ \alpha_3 e^{-(\alpha_1+\alpha_2+\alpha_3)x} & \text{if } x_1 = x_2 = x, \end{cases} \tag{1}$$

where $\theta = (\alpha_1, \alpha_2, \alpha_3)$. The corresponding joint survival function (SF) is expressed as:

$$S(x_1, x_2; \theta) = e^{-\alpha_1 x_1} e^{-\alpha_2 x_2} e^{-\alpha_3 \max\{x_1, x_2\}};\ x_1, x_2 > 0.$$

We will denote this distribution by $(X_1, X_2) \sim MOBE(\alpha_1, \alpha_2, \alpha_3)$. The MOBE distribution is not absolutely continuous, as it has a singular component on the diagonal. Instead, it is the most popular bivariate singular distribution in the statistical literature, consisting of both an absolutely continuous and a singular part. The marginal distributions of the MOBE distribution are exponential, i.e., $X_1 \sim Exp(\alpha_1 + \alpha_3)$ and $X_2 \sim Exp(\alpha_2 + \alpha_3)$ with $\min\{X_1, X_2\} \sim Exp(\alpha_1 + \alpha_2 + \alpha_3)$, and they are equal with positive probability, i.e., $P(X_1 = X_2) = \alpha_3/(\alpha_1 + \alpha_2 + \alpha_3)$.

Several methods exist for constructing bivariate lifetime distributions, including copula-based, transformation-based, shock models, and Cambanis-type approaches. Copula methods, particularly Archimedean copulas, provide flexible dependence modeling and have been widely applied to bivariate extensions of classical lifetime distributions such as Rayleigh, Weibull, Lindley, Lomax, and Erlang. However, these approaches often lead to complex likelihoods, especially under censoring. In contrast, the Marshall–Olkin construction offers an interpretable shock-based dependence structure with analytical tractability, making it well suited for reliability modeling and robust estimation. This motivates the focus of the present work on the Marshall–Olkin–type bivariate model.

There are several estimation techniques available for the MOBE distribution; see, for example, Bemis et al. [4], Arnold [5], Bhattacharyya [6], and Peña and Gupta [7]. Among these, the maximum likelihood estimation (MLE) method is the most

popular and widely used due to its attractive theoretical properties, such as consistency and asymptotic efficiency. Karlis [8] derived the MLEs for the parameters of the MOBE distribution. Although these estimators are highly efficient, they are also extremely sensitive to the presence of outliers. The presence of outliers significantly affects the MLE, leading to biased, inefficient, and unreliable parameter estimates. The primary motivation for this study arises from the UEFA Champions League football data set, which contains noticeable outliers. Since many real-world bivariate data sets often include such outliers, the development of robust parameter estimation methods becomes essential.

Recent years have seen rising interest in robust estimation techniques, especially the minimum density power divergence estimation (MDPDE) approach. Basu et al. [2] first introduced the MDPDE method and established its properties, and it has since been employed in many univariate settings, e.g., Juárez and Sachucany [9], Hazra [10]. Robust estimation methods of the generalized exponential distribution with outliers discussed by Almongy and Almetwally [11]. Extending the MDPDE approach from one-dimensional to bivariate models, however, is far more challenging. Dependence between variables complicates both the mathematics and the derivation of asymptotic and robustness properties. To the best of our knowledge, the MDPDE approach has not yet been implemented for bivariate or multivariate models in the existing literature. This study seeks to address that gap by applying the MDPDE methodology to a complex bivariate singular model, with a particular focus on robust parameter estimation for the MOBE distribution. The MDPDE method offers several advantages over classical estimation techniques, notably greater robustness to outliers and improved efficiency. The insights gained from this work can also serve as a foundation for developing robust estimation procedures for other bivariate distributions.

We extend the univariate MDPDE methodology, along with its asymptotic properties, to complex bivariate singular distributions. Recently, Kumar et al. [12] introduced the MDPDE approach to absolutely continuous biraite exponential distribution. In this framework, we develop the associated robust inferential procedures and investigate their theoretical robustness properties. Furthermore, we adapt the univariate approach for selecting the optimal tuning parameter to the bivariate context. Extensive simulation studies and real data analysis demonstrate that the proposed MDPDE method outperforms the MLE in terms of both robustness and efficiency, particularly in the presence of outliers.

The derivation of the MDPDE for this bivariate model, along with the establishment of its asymptotic properties, necessitates model-specific modifications and cannot be seen as a direct extension of existing univariate results. Additionally, the influence function analysis and robustness characteristics had to be carefully formulated to suit the unique structure of this bivariate singular framework. Although establishing the theoretical properties of the MDPDE within the MOBE model is complex, the approach largely parallels that of the univariate case. A further significant contribution is the introduction of a data-driven method for selecting the optimal tuning parameter, which effectively balances robustness and efficiency while controlling the impact of outliers. Our findings indicate that, under contamination, the MDPDE consistently outperforms the MLE.

In this context, we implement the MDPDE technique, introduced by Basu et al. [2], for robust and efficient parameter estimation of the MOBE distribution. This approach involves minimizing a suitable divergence measure, known as the density power divergence (DPD), over the parameter space to obtain the parameter estimates. The family of the DPDs proposed by Basu et al. [2], indexed by a tuning parameter γ, between two density functions g and f is given as

$$d_\gamma(g,f) = \int_0^\infty \int_0^\infty \left\{ f^{1+\gamma}(x;\theta) - \left(1+\frac{1}{\gamma}\right) f^\gamma(x;\theta) g(x) + \frac{1}{\gamma} g^{1+\gamma}(x) \right\} dx \qquad (2)$$

for $\gamma > 0$ and for $\gamma = 0$ (equivalently $\gamma \to 0$) as;

$$d_0(g,f) = \lim_{\gamma\to 0} d_\gamma(g,f) = \int_0^\infty \int_0^\infty g(x) \log\left\{ \frac{g(x)}{f(x;\theta)} \right\} dx, \qquad (3)$$

where g is the PDF of the underlying (true) distribution and f is the parametric PDF. Note that $d_0(g,f)$ is the well-known Kullback-Leibler (KL) divergence, see Kullback and Leiber [13]. On the other hand, for $\gamma = 1$, $d_1(g,f)$ is the L_2-squared distance between the densities. Thus, for $0 < \gamma < 1$, the class of the DPDs provides a smooth bridge between the KL divergence and the L_2-square distance. The DPD is a special case of the Bregman divergence, see Jones and Byrne [14]. We primarily focus on smaller values of γ, i.e., in [0, 1]. There is no strong reason to avoid the case $\gamma > 1$, except for the loss of efficiency. The parameter γ is user-specified and controls the trade-off between robustness and efficiency. As γ increases, the robustness of the procedure improves, but its efficiency diminishes. Choices of γ near zero provide considerable robustness while retaining efficiency close to that of the MLE. A significant advantage of the MDPDE approach is that it does not require nonparametric smoothing for the estimation of true density unlike other minimum divergence approaches (Basu and Lindsay, [15], Beran [16]); as a result, the MDPDE is comparatively easy to implement in practice.

In this paper, we analyze football (soccer) data from 37 UEFA Champions League matches played between 2004 and 2006. Using the widely adopted bagplot method, we identify the proportion of outliers in the data set. Due to the possible presence of outliers, we apply the proposed MDPDE method to obtain robust parameter estimates for the MOBE model. The MDPDEs for the MOBE model cannot be obtained in closed form. The MDPDEs are numerically computed by using *optim*() (Nelder-Mead algorithm) function in R environment. For each value of the tuning parameter γ, the MDPDE provides a distinct set of parameter estimates. Thus, the MDPDE at $\gamma = 0$ coincides with the MLE, while for $\gamma > 0$, it serves as a robust generalization of the MLE. We derive the asymptotic distribution of the MDPDE for the MOBE distribution and compare its asymptotic covariance matrix with the Fisher information matrix. Furthermore, we examine the asymptotic relative efficiency (ARE) of the MDPDEs observe that it decreases as γ increases. The stability of the MDPDEs for the MOBE distribution is study by the influence function (IF). We observed that the IF for $\gamma > 0$ is bounded, ensuring the robustness of the MDPDEs, while the IF for $\gamma = 0$, the MLE case, is unbounded. We also investigate how the fitted model varies

with different values of the tuning parameter γ and propose an optimal, data-driven selection of γ by minimizing the empirical asymptotic mean square error. An extensive simulation study is conducted to evaluate the performance of the proposed MDPDE method. The numerical results show that the MDPDEs for the MOBE distribution outperform the MLEs, particularly in the presence of outliers.

The rest of the paper is structured as follows. We discuss the robust and efficient estimation technique in Section 2. The asymptotic properties of the MDPDEs are given in Section 3. For robustness of the MDPDE we discuss the Influence function in Section 4. In section 5, we discuss the choice of the optimal tuning parameter. In Section 6, a simulation study is carried out to demonstrate the effectiveness of the proposed method. A numerical example is presented in Section 7. Finally, in Section 8, we concludes this article.

2 Minimum Density Power Divergence Estimation

In this section, we present robust and efficient estimates, with a particular focus on the computation of the MDPDEs for the unknown parameters of the MOBE model. While the MLE is the most efficient estimation method, it is highly sensitive to the presence of outliers, resulting in biased and inefficient estimates. Therefore, a robust estimation method is essential, especially when outliers are present in the data set. To address this issue, we adopt the minimum density power divergence (MDPD) estimation technique introduced by Basu et al. [2]. Since the Marshall–Olkin distribution is supported partly on a lower-dimensional manifold, the density power divergence is defined with respect to a mixed dominating measure, and the corresponding estimating equations naturally decompose into absolutely continuous and singular components.

Suppose we observe $(x_{11}, x_{21}), (x_{12}, x_{22}), ..., (x_{1n}, x_{2n})$ as n realizations from a underlying true distribution G. We aim to fit the data using a statistical parametric family of the MOBE densities, denoted by $\mathcal{F}_\theta = \{f(x_1, x_2; \theta) : \theta \in \Theta \subseteq \mathbb{R}^3_+\}$. For further development, we use the following notations: $I_1 = \{i : x_{1i} < x_{2i}\}$, $I_2 = \{i : x_{1i} > x_{2i}\}$ and $I_3 = \{i : x_{1i} = x_{2i}\}$. Let $n_j = |I_j|$ denote the number of observations in the set I_j, $j = 1, 2, 3$ such that $n = n_1 + n_2 + n_3$. For a given tuning parameter $\gamma \geq 0$, the MDPD functional (trivariate) $T_\gamma(\cdot)$, defined over the true distribution G, is the point in the parameter space Θ corresponding to the element in $\mathcal{F}_\theta$ that is closest to g such that

$$d_\gamma\big(g(x_1, x_2), f(x_1, x_2; T_\gamma(G))\big) = \arg\min_{\theta \in \Theta} d_\gamma\big(g(x_1, x_2), f(x_1, x_2; \theta)\big) \tag{4}$$

provided that the minimum is attained. Thus, $T_\gamma(G) = \theta$ represents the parameter vector that best fits the true distribution G. However, in practice, the true density g is unknown, and hence the minimizer of $d_\gamma(g, f(x_1, x_2; \theta))$ cannot be obtained directly; alternatively, we use an estimate of g. A significant advantage of the DPD family over other robust divergence measures is that it avoid the nonparametric smoothing (and the associated numerical challenges) when estimating the density g. It just use the empirical distribution function associated with the observed random sample. Consequently, the MDPDE of θ for a fixed tuning parameter $\gamma \in [0, 1]$ is given by

$\hat{\theta}_{\gamma,n} = T_\gamma(G_n) \in \Theta$. Thus, $\hat{\theta}_{\gamma,n}$ minimizes $d_\gamma(g_n(x_1, x_2), f_\theta(x_1, x_2))$ with respect to θ, i.e.,

$$\hat{\theta}_{\gamma,n} = \arg\min_{\theta\in\Theta} H_{\gamma,n}(\theta), \tag{5}$$

where for $\gamma > 0$ the MDPD function is

$$H_{\gamma,n}(\theta) = \int_0^\infty \int_0^\infty f^{1+\gamma}(x_1, x_2; \theta) dx_1 dx_2 - \left(1 + \frac{1}{\gamma}\right) \frac{1}{n} \sum_{i=1}^{n} f^\gamma(x_{1i}, x_{2i}; \theta). \tag{6}$$

For $\gamma = 0$, the MDPD function $H_{\gamma,n}(\theta) = -1/n \sum_{i=1}^{n} \log f(x_{1i}, x_{2i}; \theta)$ is the negative of 1/n times log-likelihood function, and hence the MDPDE, $\hat{\theta}_{0,n} = T_0(G_n)$, coincides with the MLE. The explicit expression of $H_{\gamma,n}(\theta)$ for γ is provided in the following theorem.

Theorem 1 *Let $(x_{11}, x_{21}), (x_{12}, x_{22}), ..., (x_{1n}, x_{2n})$ be n observations from the true distribution $G(x_1, x_2)$. For $\gamma > 0$, the MDPD function for $MOBE(\alpha_1, \alpha_2, \alpha_3)$ model is*

$$\begin{aligned} H_{\gamma,n}(\theta) = {} & \frac{\alpha_1^{1+\gamma}(\alpha_2 + \alpha_3)^\gamma}{(1+\gamma)^2(\alpha_1 + \alpha_2 + \alpha_3)} + \frac{(\alpha_1 + \alpha_3)^\gamma \alpha_2^{1+\gamma}}{(1+\gamma)^2(\alpha_1 + \alpha_2 + \alpha_3)} + \frac{\alpha_3^{1+\gamma}}{(1+\gamma)(\alpha_1 + \alpha_2 + \alpha_3)} \\ & - \left(1 + \frac{1}{\gamma}\right) \frac{1}{n} \Bigg[\alpha_1^\gamma (\alpha_2 + \alpha_3)^\gamma \sum_{i\in I_1} e^{-\gamma\alpha_1 x_{1i}} e^{-\gamma(\alpha_2+\alpha_3)x_{2i}} \\ & + (\alpha_1 + \alpha_3)^\gamma \alpha_2^\gamma \sum_{i\in I_2} e^{-\gamma(\alpha_1+\alpha_3)x_{1i}} e^{-\gamma\alpha_2 x_{2i}} + \alpha_3^\gamma \sum_{i\in I_3} e^{-\gamma(\alpha_1+\alpha_2+\alpha_3)x_i} \Bigg]. \end{aligned} \tag{7}$$

The proof of Theorem 1 is given in Appendix A. The MDPDEs for the unknown parameters α_1, α_2 and α_3 are determined by minimizing the function $H_{\gamma,n}(\theta)$ provided in equation (7). It is evident that the MDPDEs cannot be expressed in closed form, thereby resulting in a three-dimensional optimization problem. These estimates are computed numerically using standard optimization algorithms, such as the Nelder-Mead method, see Nelder and Mead [17], which is available in R programming environment through the *optim()* function.

3 Properties of MDPDE

The unbiased estimating equation for any $\gamma \geq 0$ is derived by differentiating $H_{\gamma,n}(\theta)$, as defined in equation (6), with respect to θ. The resulting equation is given by:

$$U_{\theta,n} = \frac{1}{n} \sum_{i=1}^{n} S_\theta(x_{1i}, x_{2i}) f^\gamma(x_{1i}, x_{2i}; \theta) - \int_0^\infty \int_0^\infty S_\theta(x_1, x_2) f^{\gamma+1}(x_1, x_2; \theta) dx_1 dx_2 = 0, \tag{8}$$

where $S_\theta(x_1, x_2) = \nabla \log[f(x_1, x_2; \theta)]$ is the maximum likelihood score vector of the MOBE model, given in Appendix B. For $\gamma = 0$, equation (8) reduces to the standard estimating equations for the MLE. However, for $\gamma > 0$, the MDPDE yields weighted

score equations, where the weights are given by $f^{\gamma}(x_{1i}, x_{2i}; \theta)$ for (x_{1i}, x_{2i}). These weights are smaller for outlying observations, effectively downweighting their influence and resulting in robust estimates. The equation (8) can also be expressed as

$$\sum_{i=1}^{n} \psi(x_{1i}, x_{2i}, \theta) = 0, \tag{9}$$

where

$$\psi(x_1, x_2, \theta) = S_{\theta}(x_1, x_2) f^{\gamma}(x_1, x_2; \theta) - \int_0^{\infty} \int_0^{\infty} S_{\theta}(x_1, x_2) f^{\gamma+1}(x_1, x_2; \theta) dx_1 dx_2$$

is a vector-valued function of three dimension same as that of the parameter space of the MOBE distribution. From equation (9), it is evident that the MDPDEs are the M-estimators, for further details see Huber [18], and Hample et al. [19].

3.1 Asymptotic Properties

We further explore some theoretical properties of the MDPDE $\hat{\theta}_{\gamma,n}$. When G is not in the model, our best fitting parameter $\theta^g = T_{\gamma}(G)$, in the sense of minimising the discrepancy $d_{\gamma}(g, f(\cdot; \theta))$, where θ denotes the generic element of Θ, will be the root of the theoretical estimating equation

$$\int_0^{\infty} \int_0^{\infty} S_{\theta}(x_1, x_2) f^{\gamma}(x_1, x_2; \theta) dG(x_1, x_2) = \int_0^{\infty} \int_0^{\infty} S_{\theta}(x_1, x_2) f^{\gamma}(x_1, x_2; \theta) dF_{\theta}(x_1, x_2),$$

where $S_{\theta}(x_1, x_2)$ denotes the score function provided in Appendix B. Let the 3×3 symmetric matrices $J_{\gamma}(\theta)$ and $K_{\gamma}(\theta)$ be defined as follows.

$$\begin{aligned} J_{\gamma}(\theta) &= \int_0^{\infty} \int_0^{\infty} S_{\theta}(x_1, x_2) S_{\theta}(x_1, x_2)^T f^{1+\gamma}(x_1, x_2; \theta) dx_1 dx_2 \\ &+ \int_0^{\infty} \int_0^{\infty} \{I_{\theta}(x_1, x_2) - \gamma S_{\theta}(x_1, x_2) S_{\theta}(x_1, x_2)^T\} [g(x_1, x_2) - f(x_1, x_2; \theta)] f^{\gamma}(x_1, x_2; \theta) dx_1 dx_2, \end{aligned} \tag{10}$$

where $I_{\theta}(x_1, x_2) = -\nabla[S_{\theta}(x_1, x_2)]$ is the information function of the model given in Appendix B, which is positive definite for all required θ, and

$$K_{\gamma}(\theta) = \int_0^{\infty} \int_0^{\infty} S_{\theta}(x_1, x_2) S_{\theta}(x_1, x_2)^T f^{2\gamma}(x_1, x_2; \theta) g(x_1, x_2) dx_1 dx_2 - \xi_{\gamma}(\theta) \xi_{\gamma}(\theta)^T \tag{11}$$

$$\text{with} \quad \xi_{\gamma}(\theta) = \int_0^{\infty} \int_0^{\infty} S_{\theta}(x_1, x_2) f^{\gamma}(x_1, x_2; \theta) g(x_1, x_2) dx_1 dx_2.$$

The MOBE distribution can also be written as convex combination of two-dimensional absolutely continuous distribution and one-dimensional absolutely continuous distribution as given below

$$f(x_1, x_2, \theta) = \frac{\alpha_1 + \alpha_2}{\alpha_1 + \alpha_2 + \alpha_3} f_{ac}(x_1, x_2; \theta) + \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3} f_{si}(x; \theta), \tag{12}$$

where the bivariate absolutely continuous distribution is

$$f_{ac}(x_1, x_2; \theta) = c \begin{cases} \alpha_1(\alpha_2 + \alpha_3) e^{-\alpha_1 x_1} e^{-(\alpha_2+\alpha_3)x_2} & \text{if } x_1 < x_2 \\ (\alpha_1 + \alpha_3)\alpha_2 e^{-(\alpha_1+\alpha_3)x_1} e^{-\alpha_2 x_2} & \text{if } x_1 > x_2, \end{cases}$$

with $c = \frac{\alpha_1+\alpha_2+\alpha_3}{\alpha_1+\alpha_2}$, and the univariate continuous distribution is

$$f_{si}(x; \theta) = (\alpha_1 + \alpha_2 + \alpha_3) e^{-(\alpha_1+\alpha_2+\alpha_3)x}; \qquad x_1 = x_2 = x > 0.$$

This formulation guarantees that the density is smooth and differentiable with respect to the parameter vector θ, and that the integral of the density over the sample space is well-defined. Moreover, it ensures that the MDPD objective function is differentiable, that differentiation and integration can be interchanged, and that the expectations of the required derivatives remain finite. The following theorem addresses the asymptotic properties of the MDPDE for the MOBE model.

Theorem 2 *For a fixed $\gamma > 0$, under the conditions $(D1) - (D5)$ outlined by Basu et al. (page 304) [20], as $n \to \infty$ the MDPDE $\hat{\theta}_{\gamma,n}$ is consistent for θ^g, i.e., $\hat{\theta}_{\gamma,n} \xrightarrow{p} \theta^g$, and follows an asymptotic trivariate normal distribution, i.e., $\sqrt{n}(\hat{\theta}_{\gamma,n} - \theta^g) \sim \mathcal{AN}_3(0, \Sigma_{\theta^g})$, where $\Sigma_{\theta^g} = J_\gamma(\theta^g)^{-1} K_\gamma(\theta^g) J_\gamma(\theta^g)^{-1}$.*

Proof The joint PDF $f(x_1, x_2; \theta)$ of the MOBE distribution as given in (12) is constructed as a mixture of absolutely continuous distributions. Hence it is very easy to show that the MOBE distribution satisfies all the regularity conditions $(D1) - (D5)$ outlined by Basu et al. (page 304) [20]. The proof follows a similar approach to the one used in the proof of Theorem 9.2 (Basu et al., page 304) [20]. □

3.1.1 Estimate of covariance matrix of MDPDE

When the true distribution G belongs to the model family, i.e. $G = F(\cdot; \theta^0)$ for some $\theta^0 \in \Theta$, then the matrices $J_\gamma(\theta)$, $K_\gamma(\theta)$ and $\xi_\gamma(\theta)$ can be written as follws;

$$J_\gamma(\theta^0) = \int_0^\infty \int_0^\infty S_{\theta^0}(x_1, x_2) S_{\theta^0}(x_1, x_2)^T f^{1+\gamma}(x_1, x_2; \theta^0) dx_1 dx_2, \tag{13}$$

$$K_\gamma(\theta^0) = J_{2\gamma}(\theta^0) - \xi_\gamma(\theta^0)\xi_\gamma(\theta^0)^T, \text{ and } \xi_\gamma(\theta^0) = \int S_{\theta^0}(x_1, x_2) f^{1+\gamma}(x_1, x_2; \theta^0) dx_1 dx_2. \tag{14}$$

Note, as $\gamma \to 0$, $J_\gamma(\theta^0)$ and $K_\gamma(\theta^0)$ both are equal and tend to the Fisher information matrix.

Theorem 3 *Let $\theta^0 = (\alpha_1^0, \alpha_2^0, \alpha_3^0) \in \Theta \subset \mathbb{R}_+^3$ represent the target parameter in the MOBE model. Assume that for a fixed $\gamma > 0$, the integrability conditions (B1) - (B3) in Appendix B are satisfied. Then, as $n \to \infty$, we have $\hat{\theta}_{\gamma,n} \xrightarrow{p} \theta^0$ and $\sqrt{n}(\hat{\theta}_{\gamma,n} - \theta^0) \sim \mathcal{AN}_3(0, \Sigma_{\theta^0})$, where $\Sigma_{\theta^0} = J_\gamma(\theta^0)^{-1} K_\gamma(\theta^0) J_\gamma(\theta^0)^{-1}$.*

The elements of the matrices $J_\gamma(\theta^0)$ and $K_\gamma(\theta^0)$ are derived analytically and are provided in Appendix C. The final expressions for the elements of Σ_{θ^0} depend on the model parameters $\alpha_1, \alpha_2, \alpha_3$ and the tuning parameter γ. The natural consistent estimate of the asymptotic covariance of the MDPDE $\hat{\theta}_{\gamma,n}$ is

$$\widehat{Var}(\hat{\theta}_{\gamma,n}) = n^{-1} J_\gamma^{-1}(\hat{\theta}_{\gamma,n}) K_\gamma(\hat{\theta}_{\gamma,n}) J_\gamma^{-1}(\hat{\theta}_{\gamma,n}). \tag{15}$$

This approach is utilized in a data-driven manner to select the optimal tuning parameter γ, as discussed in Section 5. We plot $\Sigma_{\theta^0}^{(i,j)}$, the $(i,j)^{th}$ element of Σ_{θ^0} for $i, j = 1, 2, 3$, for various values of γ, as shown in Figure 1. The diagonal elements $\Sigma_{\theta^0}^{(i,i)}$ (or Σii) represent the asymptotic variance of $\sqrt{n}\hat{\alpha}_i$, and they increase as γ increases. The asymptotic covariance $\Sigma_{\theta^0}^{(1,2)}$ (or $\Sigma 12$) between $\sqrt{n}\hat{\alpha}_1$ and $\sqrt{n}\hat{\alpha}_2$ decreases with increasing γ, while the asymptotic covariances $\Sigma_{\theta^0}^{(1,3)}$ (or $\Sigma 13$) between $\sqrt{n}\hat{\alpha}_1$ and $\sqrt{n}\hat{\alpha}_3$, and $\Sigma_{\theta^0}^{(2,3)}$ (or $\Sigma 23$) between $\sqrt{n}\hat{\alpha}_2$ and $\sqrt{n}\hat{\alpha}_3$ increase as γ increases. Thus the MDPDE with higher γ is more robust to outliers but less efficient.

3.2 Asymptotic Relative Efficiency

We study the performances of the proposed MDPDEs that can be expected through their theoretical properties. The first measure of the correctness of any estimator is its standard error or variance. It is difficult to obtain an analytic expression of the exact sampling distribution of the MDPDEs, in general, and also in the case of the MLE. It is easy to verify that the asymptotic variance of the MDPDE is minimum when $\gamma = 0$. Thus, the asymptotic variance of the MDPDE is larger than that of the MLE, and considering an estimator with a smaller variance to be preferred in general, it is important to study the asymptotic relative efficiency (ARE), the ratio of the asymptotic variances of the MLE over that of the MDPDE assuming there is no outlier in the data. A value of the ARE close to one indicates that the standard errors of the MLE and the MDPDE are comparable and hence, we can achieve robustness with only a little compromise in variance. By definition, the ARE of the MDPDE at $\gamma = 0$ is one for the MOBE model.

To compare the performance of the MLE and the MDPDE for the MOBE model we performed an extensive simulation. We generate 1000 samples of size 100 from the

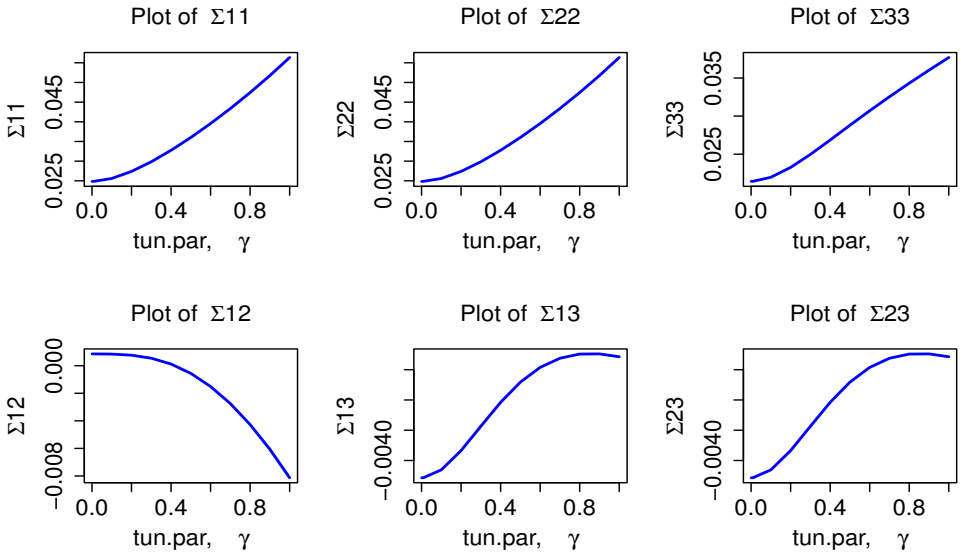


Fig. 1: Variation of the elements Σij of the covariance matrix Σ_{θ^0} for different values of the tuning parameter γ with MOBE(1, 1, 1) model.

MOBE model $MOBE(\alpha_1, \alpha_2.\alpha_3)$. For each sample, we evaluate the AREs of MDPDEs with different values γ reported in Table 1. From Table 1, we observe that the AREs of the MDPDEs decrease with increasing values of γ, but the loss is not quite significant for smaller values of $\gamma > 0$ for more details see Figure 2. Thus, the asymptotic variance of MDPDE under pure data (no outlier) assumption is comparable with that of the MLE at least for small values of γ.

Table 1: Asymptotic relative efficiency of the MDPDEs at different γ with respect to the MLE.

	tuning parameter $\gamma \rightarrow$											
Par.	0.001	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
MOBE(1.0, 1.0, 1.0)												
α_1	1.00	1.00	0.98	0.92	0.85	0.77	0.71	0.65	0.59	0.54	0.50	0.46
α_2	1.00	1.00	0.98	0.92	0.84	0.77	0.70	0.64	0.59	0.54	0.50	0.46
α_3	1.00	1.00	0.98	0.93	0.87	0.81	0.75	0.71	0.67	0.63	0.60	0.57
MOBE(0.5, 0.4, 0.6)												
α_1	1.00	1.00	0.98	0.92	0.85	0.78	0.72	0.66	0.60	0.56	0.51	0.47
α_2	1.00	1.00	0.98	0.93	0.87	0.80	0.74	0.68	0.63	0.58	0.54	0.50
α_3	1.00	1.00	0.97	0.90	0.83	0.76	0.70	0.65	0.61	0.58	0.56	0.53
MOBE(1.5, 1.4, 1.6)												
α_1	1.00	1.00	0.98	0.92	0.85	0.78	0.71	0.65	0.60	0.55	0.51	0.47
α_2	1.00	1.00	0.97	0.91	0.84	0.77	0.70	0.64	0.59	0.54	0.50	0.46
α_3	1.00	1.00	0.98	0.93	0.87	0.81	0.75	0.70	0.66	0.62	0.58	0.55

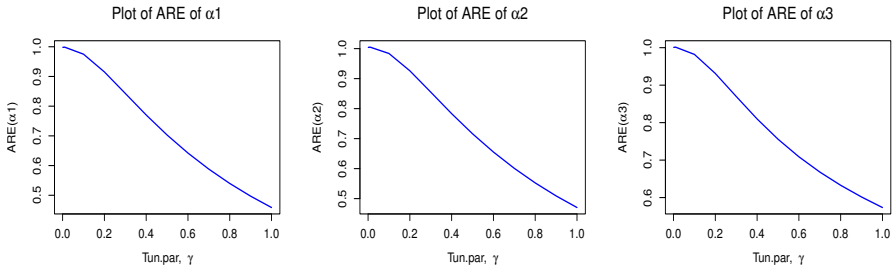


Fig. 2: Variation of ARE for the estimates of α_1, α_2 and α_3 for different values of the tuning parameter γ with MOBE(1, 1, 1) model.

4 Influence Function

In the previous sections, we have developed MDPDEs for the MOBE distribution, as a robust alternative to classical MLE. In this section, we theoretically justify the robustness of the proposed estimators through the study of its influence function (IF), proposed by Hample [21].

When the model is correctly specified, i.e., $G = F(\cdot;\theta^0)$ for some $\theta^0 = (\alpha_1^0, \alpha_2^0, \alpha_3^0) \in \Theta$, following [2], the IF of the MDPD functional $T_\gamma(\cdot)$ with tuning parameter $\gamma > 0$ for a single observation is

$$IF(x_1, x_2; T_\gamma, F(\cdot;\theta^0)) = J_\gamma^{-1}(\theta^0)\big[S_{\theta^0}(x_1, x_2) f^\gamma(x_1, x_2; \theta^0) - \xi_\gamma(\theta^0)\big]. \tag{16}$$

where $S_{\theta^0}(x_1, x_2)$ is the score function given in Appendix B and $J_\gamma(\theta)$ and $\xi_\gamma(\theta)$ are defined in equation (13) and (14) respectively. The IF with respect to all the observations is given by

$$IF(x_1, x_2; T_\gamma, F(\cdot;\theta^0)) = J_\gamma^{-1}(\theta^0) \sum_{i=1}^{n} \big\{S_{\theta^0}(x_{1i}, x_{2i}) f^\gamma(x_{1i}, x_{2i}; \theta^0) - \xi_\gamma(\theta^0)\big\}.$$

Based on this, it may be mentioned that conditions for boundedness of the influence functions presented in this article, either with respect to an observation or with respect to all the observations, are bounded if $\gamma > 0$, but if $\gamma = 0$, the norm of the IFs can be very large, in comparison to $\gamma > 0$, since it can be deduced that $IF(x_1, x_2; T_0, F(\cdot;\theta^0)) = J_\gamma^{-1}(\theta^0) S_{\theta^0}(x_1, x_2)$, where $\xi_0(\theta^0) = 0$. This implies that in the following theorem the proposed MDPDEs with $\gamma > 0$ are robust against outliers, but the classical MLE is clearly non-robust.

Theorem 4 *Assume that true parameters α_1^0, α_2^0 and α_3^0 of the MOBE distribution are finite. Then each component of the trivariate function $IF(x_1, x_2; T_\gamma, F(\cdot;\theta^0))$ given in equation (16) is bounded for $\gamma > 0$ and is unbounded for $\gamma = 0$, the MLE case.*

The proof of Theorem 4 is given Appendix A.

Theorem 5 *For fixed* (x_1, x_2)*, the function* $IF(x_1, x_2; T_\gamma, F(\cdot; \theta^0))$*, given in equation* (16)*, is a decreasing function of tuning parameter* $\gamma > 0$*.*

Proof The proof is very simple due to the exponential terms in the IF, see equation (16), which are decreasing functions of γ. These exponential terms dominate all remaining functions of γ. Thus, the first component of the IF is a decreasing function of γ. Similarly, we can easily see this for the remaining components. Hence, the IF is a decreasing function of γ. □

Using the IF, the asymptotic variance of the MDPDE under model condition is obtained as

$$Var(\hat{\theta}_{\gamma,n}) = \int_0^\infty \int_0^\infty IF(x_1, x_2; T_\gamma, F(\cdot; \theta^0)) IF(x_1, x_2; T_\gamma, F(\cdot; \theta^0))^T dx_1 dx_2.$$

To study the robustness, Hampel [21] introduced three important summary values of the IF apart from its expected square. The first and most important one is the supremum of the absolute value, therefore the gross-error-sensitivity (GES) of T_γ at G

$$GES = \sup_{(x_1,x_2)} ||IF(x_1, x_2; T_\gamma, G)||, \tag{17}$$

where $||...||$ is the Euclidean norm and the supremum being taken over all (x_1, x_2) where $IF(x_1, x_2; T_\gamma, G)$ exists. If GES is finite (or infinite), the MDPDE is said to be robust (or non-robust). For MOBE model, since the IF $\gamma > 0$ is bounded, the GES is finite, indicating that the MDPDEs are robust. On the other hand, since the IF for $\gamma = 0$ is unbounded, the GES for the MLE is infinite. Furthermore, as the value of γ, the GES decreases, thereby increasing the robustness of the MDPDEs.

5 Choice of Tuning Parameter

The tuning parameter γ offers a balance between the efficiency and robustness of the MDPDEs. A larger γ results in greater robustness, but more loss in efficiency. The MLE is known to be fully efficient when there is no outliers in the data. However, since the actual outliers proportion in the data is unknown, a data-driven algorithm is required to select an optimal tuning parameter γ for practical applications of the MDPDE. This optimal choice of γ should ideally balance efficiency and robustness.

To achieve this, we consider the mean square error (MSE), which depends on both the asymptotic variance and bias. The asymptotic empirical asymptotic covariance matrix of the MDPDE with tuning parameter γ, given in equation (15), is a function of γ only. We denote this by $\hat{V}(\gamma) = \widehat{Var}(\hat{\theta}_{\gamma,n})$. Initially, Hong and Kim [22] suggested that the optimal γ is obtained by minimizing the $tr\big(\hat{V}(\gamma)\big)$, i.e.,

$$\hat{\gamma} = \arg\min_{\gamma} tr\big(\hat{V}(\gamma)\big), \tag{18}$$

where $tr(\cdot)$ is the trace of the matrix. We refer to this method as the HK algorithm. It ignores the bias of the estimates, which can sometimes play a significant role in estimation, especially in the presence of outliers. Warwick and Johnson [23] proposed a modified version of the Hong and Kim approach by incorporating the bias term. Consequently, the optimal γ is determined by minimizing the empirical mean square error, which is a function of only γ, as given by:

$$MSE(\gamma) = tr\big(\hat{V}(\gamma)\big) + (\hat{\theta}_{\gamma,n} - \theta_p)^T(\hat{\theta}_{\gamma,n} - \theta_p), \tag{19}$$

where θ_p is the initial (or pilot) MDPDE with a tuning parameter p. This approach is referred to as the WJ algorithm for γ-selection. However, However, this algorithm suffers from the drawback of choosing the initial tuning parameter p. Warwick and Johnson suggested to choose p as 1.0. But for $\gamma = 1$, the MDPDE is Square L_2 minimum distance estimate, which is inefficient. Ghosh and Basu [24] proposed a one-step algorithm using $\gamma = 0.5$ as the pilot choice. Despite this improvement, the fundamental issue of pilot dependence remains unresolved in both cases. To overcome this limitation, Basak et al. [25] proposed an iterative algorithm, where the WJ algorithm is used as the first iteration. This iterative approach is referred to as the IWJ algorithm. Below, we outline the steps of the IWJ algorithm. Now, we describe the k^{th} for $k = 1, 2, ...,$ iteration of the IWJ algorithm. Suppose we first choose an initial pilot $p^{(0)} \in [0, 1]$. At the $(k-1)^{th}$ iteration, the pilot tuning parameter is $p^{(k-1)}$.

1. Calculate pilot MDPDE with the tuning parameter $p^{(k-1)} = \gamma_0$, denoted as $\theta_{p^{(k-1)}}$.
2. Find MDPDE $\hat{\theta}_{\gamma,n}$ with different γ, like $\gamma \in \{0.001, 0.002, ...\}$.
3. Find $MSE(\gamma) = tr\big(\hat{V}(\gamma)\big) + (\hat{\theta}_{\gamma,n} - \theta_{p^{(k-1)}})^T(\hat{\theta}_{\gamma,n} - \theta_{p^{(k-1)}})$.
4. Find $\hat{\gamma}^{(k)} = \arg\min_\gamma MSE(\gamma)$.
5. Update $p^{(k)} = \hat{\gamma}^{(k)}$ and find the corresponding MSE $MSE(\hat{\gamma}^{(k)})$.

Stop the algorithm at the k^{th} iteration, where $k = 2, 3, 4, ...,$ when $MSE(\hat{\gamma}^{(k)}) - MSE(\hat{\gamma}^{(k-1)}) < \epsilon$, where ϵ is a positive small number. The IWJ algorithm requires at least two iterations. We suggest initializing p in the IWJ algorithm with the optimal γ obtained from the HK algorithm, as this choice typically requires least number of iterations.

6 Simulation Study

In this section, we present simulation results to evaluate and compare the performance of the proposed MDPDE for different values of the tuning parameter γ, along with the MLE, in the presence of outliers in the data. Seven levels of outlier proportions were considered: 0%, 1%, 5%, 10%, 20%, 25%, and 30%. We generated 1000 samples with outliers from the mixture distribution $(1-\pi)MOBE(\alpha_1, \alpha_2, \alpha_3) + \pi H(\cdot)$, where $\pi \in [0, 1]$ is the outlier proportion, $H(\cdot)$ is the bivariate outlier distribution. We assume that the outliers follows Marshall–Olkin bivariate Weibull (MOBW) distribution, denoted as $MOBW(0.01, 0.01, 0.01, 1.5)$. The MLEs and MDPDEs of the three unknown parameters of the MOBE model are obtained numerically using the Nelder-Mead method, available in different statistical software environment. In the

R environment the method is available under the "*optim()*" function. The average bias (AB) and mean square error (MSE) are obtained based on 1000 replications using the R statistical software. They give us an idea about the consistency of the estimators. The average length of confidence interval (ALCI) of parameters are obtained to check the behavior of estimators. The coverage probability (CP) based on 95% confidence interval (CI) is the proportion of times the CIs contain the true parameter value. These metrics provide insights into the consistency and reliability of the estimators. For the samples of size $n = 100$ generated from the $(1-\pi)MOBE(1.5,,1.4,1.6)+\pi MOBW(0.01,0.01,0.01,1.5)$ at different outlier proportions: $\pi = 0, 0.01, 0.05, 0.1, 0.15, 0.20, 0.25, 0.30$, the AB and MSE results are reported in Table 2 and the ALCIs and CPs are given in Table 3. We have also tried for the different outlier model, for different sample sizes and different sets of parameter values, but not reported here. We can also add the outliers in the data by using the degenerate distribution. Table 2 shows that, the ABs and MSEs of MDPDEs and MLEs are increase as the outlier proportion increases. In most of the cases the MDPDEs and MLEs have negative bias. The MDPDEs with optimal γ exhibit smaller biases and MSEs than MLEs and MDPDEs with another γ, indicating superior performance. Hence the MDPDE with optimal γ outperforms than other methods as well as the other choices of γ in terms of the AB and MSE. Table 2 demonstrate the effectiveness of MDPDE over the MLE, even in the presence of a small percentage of outliers. Table 3 shows, as the proportion of outliers increases, ALCI decreases, and CP declines to zero. The tuning parameter γ is not very significant when an uncontaminated data set is considered, while in the case of contaminated data, estimates and confidence intervals based on MLE get improved by those based on $\gamma > 0$. However, once the outlier is present in the data set, the MDPDE with optimal γ outperforms other alternatives. For high contamination percentage, MDPDE is not necessarily achieve better robustness whenever γ increases, and selecting an optimal is crucial. Thus MDPDE with optimal γ is quite effectively used to analyzed the data when outliers present in the data, and is gives robust estimate. The experimental results demonstrate that the MDPDE is highly effective for robust and efficient estimation of the MOBE model parameters α_1, α_2 and α_3.

7 Real Data Analysis

In this section, we analyze a real data set to demonstrate the advantages of the proposed MDPDE method for illustrative purposes. We consider the UEFA Champion's League football data set, which includes 37 matches where in each match the home team scored at least one goal, and at least one goal was scored directly from a kick (penalty kicks, free kicks, foul kicks, or other direct kicks) by any team. The data set, provided in Table 4, was originally analyzed by Meintanis [3] using the MOBE model. Several authors have since reanalyzed the data using the conventional MLE method for various bivariate models, e.g., Kundu and Dey [26] applied the MLE to estimate the parameters of Marshall-Olkin bivariate Weibull distribution. In this study, we apply the robust MDPDE method to the MOBE model and compare the results

Table 2: The average bias (AB) and mean square error (MSE) of estimates for the MOBE model with $\alpha_1 = 1.5, \alpha_2 = 1.4, \alpha_3 = 1.6$ with different proportions of outliers from MOBW(0.01, 0.01, 0.01, 1.5).

			Outlier proportion π							
Method	Par.	Est.	0%	1%	5%	10%	15%	20%	25%	30%
MLE	α_1	AB	0.0105	-0.3211	-0.8977	-1.1474	-1.2565	-1.3117	-1.3502	-1.3725
($\gamma = 0$)		MSE	0.0585	0.2375	0.8757	1.3341	1.5844	1.7233	1.8244	1.8846
	α_2	AB	0.0212	-0.2806	-0.8186	-1.0656	-1.1619	-1.2181	-1.2543	-1.2761
		MSE	0.0577	0.1968	0.7331	1.1501	1.3559	1.4864	1.5747	1.6292
	α_3	AB	0.0212	-0.3067	-0.9120	-1.1955	-1.3127	-1.3789	-1.4240	-1.4492
		MSE	0.0514	0.2292	0.9023	1.4470	1.7303	1.9045	2.0292	2.1012
MDPDE	α_1	AB	0.0105	-0.2401	-0.8002	-1.0998	-1.2322	-1.2979	-1.3420	-1.3669
($\gamma = 0.01$)		MSE	0.0587	0.1556	0.7173	1.2329	1.5257	1.6880	1.8027	1.8694
	α_2	AB	0.0210	-0.2074	-0.7283	-1.0207	-1.1387	-1.2050	-1.2465	-1.2708
		MSE	0.0577	0.1297	0.5987	1.0616	1.3043	1.4553	1.5553	1.6159
	α_3	AB	0.0207	-0.2214	-0.8067	-1.1424	-1.2852	-1.3638	-1.4151	-1.4433
		MSE	0.0513	0.1445	0.7265	1.3292	1.6613	1.8635	2.0043	2.0843
MDPDE	α_1	AB	0.0120	-0.0070	-0.0559	-0.1214	-0.2121	-0.3605	-0.6106	-0.9212
($\gamma = 0.1$)		MSE	0.0615	0.0589	0.0709	0.0776	0.1180	0.2212	0.5364	1.0368
	α_2	AB	0.0198	-0.0006	-0.0392	-0.1099	-0.1985	-0.3327	-0.5571	-0.8611
		MSE	0.0591	0.0528	0.0562	0.0743	0.1037	0.1913	0.4542	0.8981
	α_3	AB	0.0182	0.0104	-0.0338	-0.0937	-0.1834	-0.3429	-0.6091	-0.9600
		MSE	0.0517	0.0494	0.0567	0.0639	0.0934	0.2013	0.5477	1.1305
MDPDE	α_1	AB	0.0186	0.0023	-0.0213	-0.0408	-0.0693	-0.1153	-0.1571	-0.2009
($\gamma = 0.3$)		MSE	0.0748	0.0687	0.0765	0.0712	0.0800	0.0828	0.0923	0.1099
	α_2	AB	0.0217	0.0074	-0.0029	-0.0370	-0.0686	-0.1026	-0.1399	-0.1979
		MSE	0.0693	0.0621	0.0624	0.0695	0.0738	0.0738	0.0817	0.1001
	α_3	AB	0.0197	0.0179	-0.0116	-0.0339	-0.0774	-0.1383	-0.1867	-0.2410
		MSE	0.0587	0.0560	0.0638	0.0595	0.0685	0.0755	0.0927	0.1147
MDPDE0	α_1	AB	0.0263	0.0072	-0.0263	-0.0536	-0.0905	-0.1473	-0.1972	-0.2455
($\gamma = 0.5$)		MSE	0.0928	0.0849	0.0894	0.0841	0.0924	0.0985	0.1104	0.1317
	α_2	AB	0.0247	0.0109	-0.0042	-0.0496	-0.0893	-0.1292	-0.1768	-0.2414
		MSE	0.0845	0.0758	0.0749	0.0810	0.0869	0.0869	0.0976	0.1214
	α_3	AB	0.0260	0.0235	-0.0225	-0.0591	-0.1214	-0.1977	-0.2669	-0.3346
		MSE	0.0697	0.0651	0.0731	0.0669	0.0835	0.0976	0.1288	0.1661
MDPDE	α_1	AB	0.0335	0.0128	-0.0284	-0.0605	-0.1027	-0.1673	-0.2225	-0.2735
($\gamma = 0.7$)		MSE	0.1133	0.1048	0.1050	0.0994	0.1068	0.1151	0.1280	0.1518
	α_2	AB	0.0271	0.0148	-0.0040	-0.0573	-0.1026	-0.1453	-0.1996	-0.2694
		MSE	0.1022	0.0923	0.0903	0.0949	0.1017	0.1016	0.1136	0.1418
	α_3	AB	0.0341	0.0320	-0.0279	-0.0753	-0.1502	-0.2369	-0.3193	-0.3949
		MSE	0.0819	0.0754	0.0830	0.0750	0.0981	0.1180	0.1607	0.2104
MDPDE	α_1	AB	0.0425	0.0208	-0.0266	-0.0601	-0.1048	-0.1761	-0.2336	-0.2846
($\gamma = 1$)		MSE	0.1474	0.1392	0.1325	0.1258	0.1309	0.1385	0.1490	0.1712
	α_2	AB	0.0296	0.0207	-0.0004	-0.0596	-0.1077	-0.1499	-0.2095	-0.2828
		MSE	0.1321	0.1211	0.1179	0.1189	0.1263	0.1234	0.1336	0.1641
	α_3	AB	0.0496	0.0490	-0.0253	-0.0814	-0.1655	-0.2606	-0.3526	-0.4328
		MSE	0.1019	0.0929	0.0980	0.0872	0.1151	0.1371	0.1879	0.2461

with those obtained using the MLE. The MDPDEs for the MOBE model parameters are calculated numerically using standard direct optimization techniques, such as the Nelder-Mead algorithm, see Nelder-Mead [17], which is implemented through the built-in *optim()* function available in R software. In this data set, X_1 represents the time (in minutes) when the first kick goal is scored by any team, and X_2 denotes the time (in minutes) at which the first goal of any type is scored by the home team. The data set includes all three possible cases: $n_1 = 6$ observations where $X_1 < X_2$, $n_2 = 17$ observations where $X_1 > X_2$ and $n_3 = 14$ observations where $X_1 = X_2$ such that $n_1 + n_2 + n_3 = 37$. The scatter plot of the data is given in Figure 3.

Table 3: The average length of 95% confidence interval (ALCI) and coverage probability (CP) of estimates for the MOBE model $\alpha_1 = 1.5, \alpha_2 = 1.4, \alpha_3 = 1.6$ with different proportions of outliers from MOBW(0.01, 0.01, 0.01, 1.5).

			Outlier proportion π							
Methods	Par.	Est.	0%	1%	5%	10%	15%	20%	25%	30%
MLE	α_1	ALCI	0.9434	0.7402	0.3678	0.2177	0.1504	0.1172	0.0956	0.0801
$\gamma = 0$		CP	0.94	0.53	0.04	0.00	0.00	0.00	0.00	0.00
	α_2	ALCI	0.9098	0.7148	0.3702	0.2113	0.1504	0.1163	0.0951	0.0789
		CP	0.94	0.55	0.05	0.00	0.00	0.00	0.00	0.00
	α_3	ALCI	0.8970	0.7059	0.3703	0.2168	0.1537	0.1200	0.0976	0.0826
		CP	0.96	0.51	0.04	0.00	0.00	0.00	0.00	0.00
MDPDE	α_1	ALCI	0.9430	0.7936	0.4359	0.2504	0.1682	0.1285	0.1029	0.0852
$\gamma = 0.01$		CP	0.95	0.65	0.07	0.00	0.00	0.00	0.00	0.00
	α_2	ALCI	0.9100	0.7656	0.4292	0.2422	0.1665	0.1260	0.1015	0.0834
		CP	0.94	0.66	0.08	0.00	0.00	0.00	0.00	0.00
	α_3	ALCI	0.8961	0.7534	0.4239	0.2423	0.1658	0.1262	0.1009	0.0843
		CP	0.96	0.63	0.06	0.00	0.00	0.00	0.00	0.00
MDPDE	α_1	ALCI	0.9564	0.9495	0.9212	0.8695	0.8128	0.7292	0.5739	0.3642
$\gamma = 0.1$		CP	0.95	0.94	0.92	0.82	0.71	0.51	0.29	0.13
	α_2	ALCI	0.9225	0.9145	0.8864	0.8375	0.7918	0.7004	0.5565	0.3527
		CP	0.94	0.95	0.92	0.85	0.72	0.52	0.32	0.12
	α_3	ALCI	0.9063	0.8964	0.8690	0.8247	0.7798	0.6971	0.5513	0.3540
		CP	0.95	0.93	0.92	0.86	0.73	0.52	0.30	0.12
MDPDE	α_1	ALCI	1.0370	1.0309	1.0180	0.9912	0.9667	0.9488	0.9143	0.8774
$\gamma = 0.3$		CP	0.94	0.93	0.94	0.91	0.88	0.87	0.81	0.74
	α_2	ALCI	0.9972	0.9904	0.9759	0.9516	0.9416	0.9078	0.8832	0.8443
		CP	0.94	0.96	0.93	0.91	0.90	0.87	0.83	0.77
	α_3	ALCI	0.9678	0.9592	0.9421	0.9189	0.9026	0.8783	0.8473	0.8186
		CP	0.94	0.94	0.93	0.92	0.90	0.86	0.77	0.70
MDPDE	α_1	ALCI	1.1456	1.1346	1.1150	1.0776	1.0403	1.0138	0.9688	0.9215
$\gamma = 0.5$		CP	0.94	0.93	0.94	0.90	0.87	0.86	0.78	0.70
	α_2	ALCI	1.0976	1.0879	1.0659	1.0318	1.0143	0.9687	0.9344	0.8849
e		CP	0.94	0.95	0.93	0.90	0.90	0.86	0.81	0.74
	α_3	ALCI	1.0463	1.0344	1.0084	0.9745	0.9462	0.9110	0.8686	0.8278
		CP	0.95	0.94	0.93	0.91	0.87	0.80	0.70	0.58
MDPDE	α_1	ALCI	1.2621	1.2457	1.2203	1.1737	1.1250	1.0915	1.0376	0.9814
$\gamma = 0.7$		CP	0.94	0.93	0.94	0.91	0.86	0.85	0.78	0.69
	α_2	ALCI	1.2055	1.1937	1.1647	1.1223	1.0980	1.0423	0.9996	0.9411
		CP	0.94	0.95	0.93	0.90	0.90	0.85	0.80	0.73
	α_3	ALCI	1.1245	1.1091	1.0750	1.0316	0.9924	0.9481	0.8961	0.8455
		CP	0.95	0.94	0.93	0.90	0.85	0.76	0.65	0.51
MDPDE	α_1	ALCI	1.4416	1.4169	1.3856	1.3301	1.2688	1.2291	1.1656	1.1002
$\gamma = 1.0$		CP	0.94	0.94	0.94	0.92	0.87	0.86	0.79	0.72
	α_2	ALCI	1.3733	1.3591	1.3224	1.2721	1.2405	1.1736	1.1218	1.0535
		CP	0.93	0.95	0.93	0.91	0.90	0.85	0.81	0.74
	α_3	ALCI	1.2399	1.2179	1.1750	1.1209	1.0685	1.0144	0.9519	0.8907
		CP	0.95	0.94	0.93	0.90	0.85	0.76	0.64	0.47

Before proceeding further, we first identify outliers using a bagplot for the bivariate data set (X_1, X_2). The bagplot provides insight into the presence of outliers in the football data. A bagplot can be considered a bivariate version of a boxplot: the orange central region represents the bivariate median, while the dark blue area known as 'the bag' corresponds to the bivariate interquartile range (IQR), containing the 50% of the most central points. The light blue region, referred to as 'the fence', includes points that are farther away but not far enough to be classified as outliers. Points that fall outside the fence are considered outliers according to the bagplot criteria; for further details, see Rousseeuw et al. [27]. From the bagplot shown in Figure 4, we observe one

Table 4: UEFA Champion's League Data.

2005-2006	X_1	X_2	2004-2005	X_1	X_2
Lyon-Real Madrid	26	20	Internazionale-Bremen	34	34
Milan-Fenerbahce	63	18	Real Madrid-Roma	53	39
Chelsea-Anderlecht	19	19	Man. United-Fenerbache	54	7
Club Brugge-Juventus	66	85	Bayern-Ajax	51	28
Fenerbache-PSV	40	40	Moscow-PSG	76	64
Internazionale-Rangers	49	49	Barcelona-Shakhtar	64	15
Panathinaikos-Bremen	8	8	Leverkusen-Roma	26	48
Ajax-Arsenal	69	71	Arsenal- Panathinaikos	16	16
Man. United-Benfica	39	39	Dynamo Kyiv-Real Madrid	44	13
Real Madrid–Rosenborg	82	48	Man. United-Sparta	25	14
Villarreal-Benfica	72	72	Bayern-M. TelAviv	55	11
Juventus-Bayern	66	62	Bremen-Internazionale	49	49
Club Brugge-Rapid	25	9	Anderlecht-Valencia	24	24
Olympiacos-Lyon	41	3	Panathinaikos-PVS	44	30
Internazionale-Porto	16	75	Arsenal-Rosenborg	42	3
Schalke-PSV	18	18	Liverpool-Olympiacos	27	47
Barcelona-Bremen	22	14	M. TelAviv-Juventus	28	28
Milan-Schalke	42	42	Bremen-Panathinaikos	2	2
Rapid-Juventus	36	52			

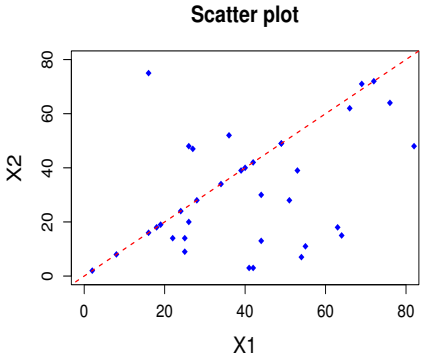


Fig. 3: Scatter plot of UEFA data.

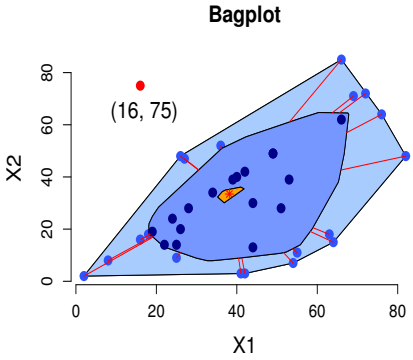


Fig. 4: Bagplot of UEFA data.

outlier point $(16, 75)$ in the data set. The MLE method is not suitable for this data set due to its excessive sensitivity to outliers, a robust parameter estimation technique is necessary for analyzing the data set using the MOBE model. To estimate the model parameters, we apply MDPDE with an appropriate optimal tuning parameter as described in Section 5.

In this case, the scaled total time on test (TTT) transform for the football data is shown in Figure 5. The results clearly indicate that both marginals X_1, X_2 and $\min\{X_1, X_2\}$ exhibit increasing hazard rates. Therefore, the MOBE model is not appropriate for analyzing this data. We further justify it using the following numerical results. First, the exponential distribution with parameter α is fitted separately

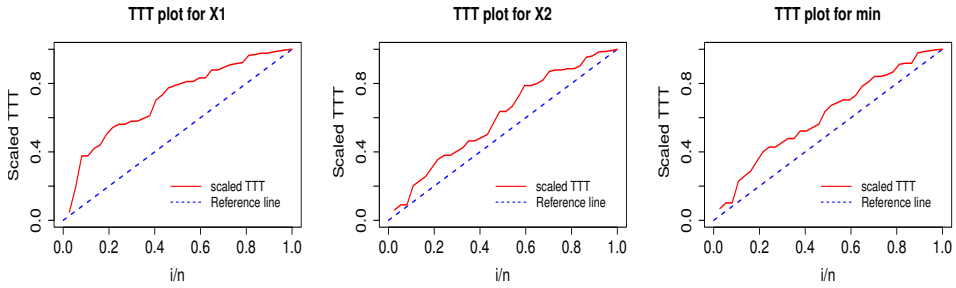


Fig. 5: Scaled TTT plots of X_1, X_2 and $\min\{X_1, X_2\}$ for the football data.

to X_1, X_2, and $\min\{X_1, X_2\}$ using both the MLE and the MDPDE methods, along with model checking, to obtain initial estimates of the parameters α_1, α_2, and α_3. The corresponding marginals numerical results are presented in Table 5. The optimal tuning parameter γ, selected using the IWJ algorithm, is found to be 0.2693. We then fit the MOBE$(\alpha_1, \alpha_2, \alpha_3)$ model to the UEFA data. The resulting MLEs, MDPDEs with the optimal γ, the bivariate Kolmogorov–Smirnov (KS) distance (see Kumar et al. [28]), and the corresponding parametric bootstrap p-values are reported in Table 6. It is clear from the KS statistics and the associated p-values for both the MLE and MDPDE methods that the MOBE model does not provide an adequate fit to the original data. Therefore, for illustrative purposes, we subsequently transform the UEFA data using a suitable transformation, as discussed below.

Table 5: The MDPDEs, MLEs, KS distances and p-values for UEFA data.

	MDPDE				MLE		
Variables	$\hat{\gamma}$	$\hat{\alpha}$	KS	p-value	$\hat{\alpha}$	KS	p-value
X_1	0.345	0.0196	0.2154	0.0646	0.0245	0.2698	0.0092
X_2	0.332	0.0263	0.1118	0.7441	0.0304	0.1375	0.4862
$\min\{X_1, X_2\}$	0.326	0.0296	0.1300	0.5594	0.0343	0.1708	0.2309

Table 6: Estimates of parameters with SEs and KS distances with associated p-value for UEFA data.

Method	Par.	Est.	SE	KS	p-value
MLE	α_1	0.0072	0.0027	0.2661	0.0016
$\gamma = 0$	α_2	0.0164	0.0038		
	α_3	0.0177	0.0037		
MDPDE	α_1	0.0058	0.0022	0.1909	0.0446
$\gamma^* = 0.2693$	α_2	0.0145	0.0038		
	α_3	0.0125	0.0033		

Based on some previous calculations we have seen that the MOBE model is not suitable to fit the football data. That means the data does not follow the MOBE distribution. Thus before analyzing the data we need to transform all the data points. After some prior analysis using the KS distance we observe that the transformation $X_1^* = (X_1/100)^{1.65}$ and $X_2^* = (X_2/100)^{1.65}$ is more appropriate for the analysis of data by using the MOBE distribution. We make this transformation mainly for computational and illustration purposes. We start with checking the MOBE distribution could be used to fit the transformed bivariate UEFA data. For this purpose we implemented "In-and-out-sample" (IOS) GoF test as introduced by Presnell and Boos [29]. The methodology for the IOS GoF test in this case can be stated as; suppose we have n observations $(x_{11}, x_{21}), (x_{12}, x_{22}), ..., (x_{1n}, x_{2n})$ from $f(x_1, x_2; \theta)$, where $\theta = (\alpha_1, \alpha_2, \alpha_3)$. Let $\hat{\theta}$ is the MLE of θ based on complete sample. The In-sample log-likelihood is given as $\sum_{i=1}^{n} l(\hat{\theta}; x_{1i}, x_{2i})$, where $l(\theta; x_{1i}, x_{2i}) = \log f(x_{1i}, x_{2i}; \theta)$. Let $\hat{\theta}_{(-i)}$ be the MLE when i^{th} observation is deleted from the sample. The Out-sample log-likelihood is given as $\sum_{i=1}^{n} l(\hat{\theta}_{(-i)}; x_{1i}, x_{2i})$. The IOS statistics is given by

$$IOS = \sum_{i=1}^{n} \big[l(\hat{\theta}; x_{1i}, x_{2i}) - l(\hat{\theta}_{(-i)}; x_{1i}, x_{2i}) \big].$$

Note that the IOS is always non-negative. Finally with the help of parametric bootstrap p-value decision is made.

Before fitting the MOBE distribution to the data set, we first fit the exponential distribution to X_1^*, X_2^* and $\min\{X_1^*, X_2^*\}$. Apart from model checking, it will help us to guess the initial values of unknown parameters. We calculate the Kolmogorov-Smirnov (KS) distance between the empirical cumulative distribution function (CDF) and the fitted CDF to assess the goodness-of-fit (GoF) of the exponential model for X_1^*, X_2^* and $\min\{X_1^*, X_2^*\}$. The MDPDE for the exponential distribution with parameter α is obtained by minimizing the following function

$$ll_{\gamma,n}(\alpha) = \frac{\alpha^{\gamma}}{(1+\gamma)} - \left(1 + \frac{1}{\gamma}\right) \frac{1}{n} \sum_{i=1}^{n} \alpha^{\gamma} e^{-\alpha\gamma x_i}$$

for $\gamma > 0$. The optimal γ is obtained by the empirical Cramer-von Mises (ECVM) distance. Taking the idea of cross-validation, proposed by Fujisawa and Eguchi [30], where the optimal γ is chosen by minimizing the ECVM distance as

$$\hat{\gamma} = \underset{\gamma}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left[\frac{n_i}{n+1} - F\big(x_i; \hat{\theta}_{\gamma,n}^{(-i)}\big) \right]^2, \tag{20}$$

where n_i is the number of observations x_j such that $x_j \leq x_i$; $j = 1, 2, ..., n$, that is $n_i = \sum_{j=1}^{n} \mathbb{1}(x_j \leq x_i)$, where $\mathbb{1}(\cdot)$ is the indicator function, and $\hat{\theta}_{\gamma,n}^{(-i)}$ is the MDPDE with tuning parameter γ obtained after removing i^{th} observation x_i from the sample. The MLEs of the shape parameter is obtained by the usual routine. The KS distances, and the associated p-values are obtained for GoF in each case. For the football data,

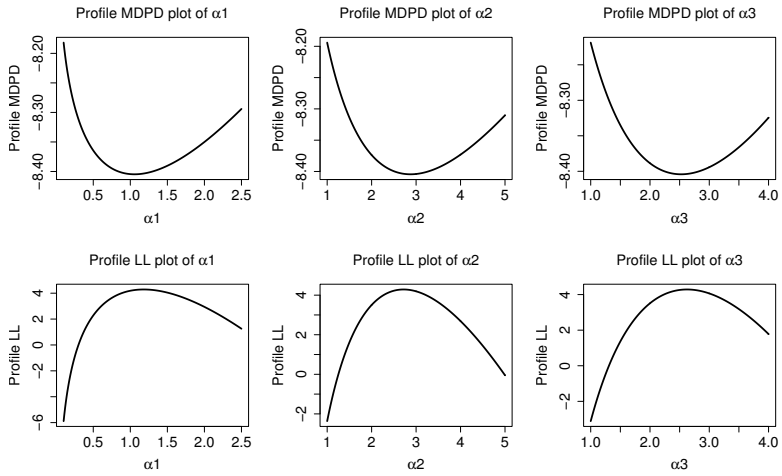
the results are summarized in Table 7. Based on the p-values, the MDPDE better fit the data than MLE. Thus we find that the exponential distribution cannot be rejected for the marginals as well as for the minimum when using the MDPDE method. Therefore, we can reasonably assume that the observations are distributed according to the exponential distribution for both marginals X_1^*, X_2^* and $\min\{X_1^*, X_2^*\}$ also. This indicates that it is appropriate to fit the MOBE model to the football data.

Table 7: The MDPDEs, MLEs, KS distances and p-values for UEFA data.

	MDPDE				MLE		
Variables	$\hat{\gamma}$	$\hat{\alpha}$	KS	p-value	$\hat{\alpha}$	KS	p-value
X_1^*	0.328	3.4133	0.0989	0.8626	3.8758	0.1270	0.5872
X_2^*	0.001	5.0162	0.1354	0.5061	5.0162	0.1354	0.5061
$\min\{X_1^*, X_2^*\}$	0.411	6.6048	0.0882	0.9358	6.1678	0.1078	0.7831

Using the initial values (1.1516, 2.2920, 2.7242) of $(\alpha_1, \alpha_2, \alpha_3)$ obtained from the MLEs given in Table 7 by solving three linear equations. The value of IOS test statistics is obtained as 2.8514 with associated p-value is 0.768. Therefore, we can conclude that the MOBE model fits the transformed data quite well. The plots of the profile MDPD function given in equation (7) and the profile log-likelihood function for each parameter of the MOBE distribution are given in Figure 6. These plots confirm the existence of unique estimates for the parameters α_1, α_2, and α_3.

Fig. 6: Plots of profile minimum density power divergence (MDPD) and log-likelihood (LL) for the football data.



Now we obtain the optimal γ, discussed in Section 6 by using the initial values (1.5887, 3.1916, 1.8246) for the parameters $(\alpha_1, \alpha_2, \alpha_3)$ from the MDPDEs obtained

above, see Table 7, by solving three linear equations. Now we select the optimal γ based on the empirical asymptotic MSE by using the IWJ algorithm. First we obtained the initial value of the pilot tuning parameter γ by the HK algorithm, which is obtained as $\gamma = 0.0206$ with Var $= 0.9183$. The optimal γ by the IWJ is obtained as $\gamma^* = 0.0206$ with MSE is 0.9183. From the Table 8 it is clear that the IWJ algorithm is consistent with the choice of initial γ and it require minimum number of iteration when we use the optimal γ obtained from the HK algorithm as initial value.

Table 8: The Optimal γ^* by the IWJ algorithm for different initial values of γ.

γ^0	γ^*	k	MSE	γ^0	γ^*	k	MSE
0.001	0.0206	4	0.9183	0.005	0.0206	4	0.9183
0.01	0.0206	4	0.9183	0.0206	0.0206	2	0.9183
0.05	0.0206	3	0.9183	0.1	0.0206	4	0.9183
0.2	0.0206	5	0.9183	0.3	0.0206	6	0.9183
0.4	0.0206	6	0.9183	0.5	0.0206	6	0.9183
0.6	0.0206	6	0.9183	0.7	0.0206	6	0.9183
0.8	0.0206	6	0.9183	0.9	0.0206	6	0.9183
1.0	0.0206	6	0.9183				

We calculate the total variation (Tvar), which is sum of asymptotic variances of estimates, to check the efficiency of the estimators. The GES given in equation (17) is obtained to see the robustness of estimators. For the transformed UFEA data, the estimate of parameters α_1, α_2 and α_3 of the MOBE model with standard error (SE), 95% CIs, Tvar and GES are given in Table 9. From Table 9 it is quite clear that the MLE is most efficient but not robust. For small γ the MDPDE is more robust but less efficient. Thus the MDPDE with optimal tuning parameter $\gamma^* = 0.0206$ provides the appropriate balance between robustness and efficient.

Now the natural question is the MDPDE with $\gamma^* = 0.0206$ fits the data well or not. For this purpose we consider the bivariate GoF test based on the statistical equivalent blocks (SEBs), for more details see Kumar et al. [28]. We also consider the univariate GoF KS test to get just idea about the marginal fits. The KS distances for the bivariate GoF with the associated p-value obtained by the parametric bootstrap method and the univariate GoF given in Table 10. Thus from the Table 10, based on the bivariate GoF the MDPDE with optimal $\gamma^* = 0.0206$ fits data better than MLE and MDPDE with other γ, because it has the smallest KS distance and the largest p-value. Based on the univariate GoF analysis, including the KS plot (Figure 7) and the percentile-percentile (PP) plot (Figure 8), the MDPDE with $\gamma^* = 0.0206$ also fits the data marginally quite well.

8 Conclusion

In this study, we have developed a robust estimation framework for the Marshall-Olkin Bivariate Exponential (MOBE) distribution, using the Minimum Density Power

Table 9: Estimates of parameters with SE and 95% CI for the UEFA data.

Method	Par.	Est.	SE	95% CI	Tvar	GES
MLE	α_1	1.1813	0.4442	(0.3106, 2.0519)	0.9156	1785326
$\gamma = 0$	α_2	2.7185	0.6329	(1.4780, 3.9589)		
	α_3	2.6246	0.5637	(1.5197, 3.7295)		
MDPDE	α_1	1.1564	0.4201	(0.3330, 1.9798)	0.9182	51.0621
$\gamma^* = 0.0206$	α_2	2.7375	0.6610	(1.4419, 4.0331)		
	α_3	2.6095	0.5521	(1.5274, 3.6916)		
MDPDE	α_1	0.7102	0.3974	(0.0000, 1.4891)	1.6411	6.2225
$\gamma = 0.5$	α_2	3.7869	1.0350	(1.7583, 5.8155)		
	α_3	2.6119	0.6418	(1.3540, 3.8698)		
MDPDE	α_1	0.2889	0.3412	(0.0000, 0.9577)	3.2269	8.7597
$\gamma = 1.0$	α_2	5.3022	1.5893	(2.1872, 8.4172)		
	α_3	2.5502	0.7646	(1.0516, 4.0488)		

Table 10: The KS distances with associated p-values in the parenthesis.

	Bivariate GoF	Univariate GoF		
		X_1	X_2	$\min\{X_1, X_2\}$
γ	KS (p-value)	KS (p-value)	KS (p-value)	KS (p-value)
0	0.0715 (0.7459)	01204 (0.6568)	0.1194 (0.6667)	0.0918 (0.9144)
0.0206	0.0706 (0.7510)	0.1166 (0.6957)	0.1193 (0.6686)	0.0927 (0.9083)
0.5	0.0978 (0.5165)	0.0951 (0.8915)	0.1740 (0.2125)	0.1019 (0.8366)
1.0	0.1442 (0.1877)	0.1378 (0.4836)	0.2424 (0.0259)	0.1456 (0.4132)

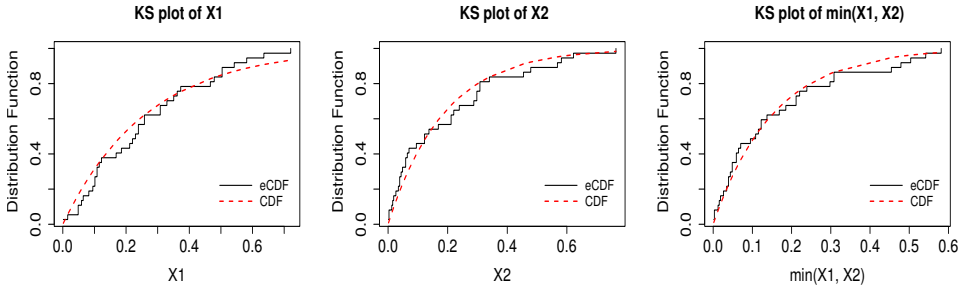


Fig. 7: KS plots of X_1, X_2 and $\min\{X_1, X_2\}$ for the UEFA data.

Divergence Estimator (MDPDE) as a divergence-based robust inference technique. Given the presence of outliers in real-world data sets, such as those in UEFA data, the MDPDE method offers a valuable alternative to classical estimation methods, particularly the Maximum Likelihood Estimator (MLE), which is highly sensitive to outliers. We derived the theoretical properties of MDPDE for the MOBE model, including closed-form estimating equations and asymptotic distributions, and demonstrated its robustness through an influence function analysis. Our findings indicate

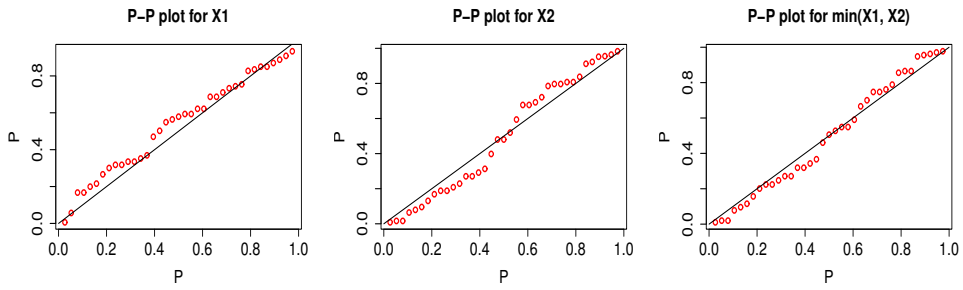


Fig. 8: PP plots of X_1, X_2 and $\min\{X_1, X_2\}$ for the football data.

that, for $\gamma > 0$, the MDPDE achieves a bounded influence function, highlighting its robustness against outliers. Furthermore, our analytical study on asymptotic relative efficiency (ARE) confirms that MDPDE maintains efficiency in estimation across various parameter settings. A comprehensive simulation study and application to a UEFA data set support the practical advantages of MDPDE over MLE, especially in scenarios with high contamination levels. By adjusting the tuning parameter γ, MDPDE can be optimized to balance robustness and efficiency, with data-driven selection methods enhancing its applicability in real data analysis.

In conclusion, the MDPDE framework presents a robust and efficient alternative to traditional methods for estimating MOBE parameters, particularly in data sets with outliers. This research underscores the potential of divergence-based estimation methods in enhancing the reliability of statistical inference in contaminated data environments. Future work could extend this approach to other multivariate lifetime distributions and further explore data-driven tuning parameter selection in diverse practical contexts.

Acknowledgment

The author would like to thank the anonymous reviewers and the Associate Editor for their constructive comments and valuable suggestions, which greatly improved the quality and clarity of this manuscript.

Appendix A Proof of thorems

Proof of Theorem 1. In the expression of $H_{\gamma,n}(\theta)$, as given in equation (6), $f(x_1, x_2; \theta)$ is the PDF of the MOBE distribution given in equation (1). For $\gamma > 0$, the first term of $H_{\gamma,n}(\theta)$ is obtained as

$$\begin{aligned}
&\int_0^\infty \int_0^\infty f_{X_1,X_2}^{1+\gamma}(x_1, x_2; \theta) dx_1 dx_2 \\
&= \int_0^\infty \int_{x_1}^\infty \alpha_1^{1+\gamma} (\alpha_2 + \alpha_3)^{1+\gamma} e^{-\alpha_1(1+\gamma)x_1} e^{-(\alpha_2+\alpha_3)(1+\gamma)x_2} dx_2 dx_1 \\
&+ \int_0^\infty \int_{x_2}^\infty (\alpha_1 + \alpha_2)^{1+\gamma} \alpha_2^{1+\gamma} e^{-(\alpha_1+\alpha_2)(1+\gamma)x_1} e^{-\alpha_2(1+\gamma)x_2} dx_1 dx_2
\end{aligned}$$

$$+\int_0^\infty \alpha_3^{1+\gamma} e^{-(\alpha_1+\alpha_2+\alpha_3)(1+\gamma)x} dx$$

$$= \frac{\alpha_1^{1+\gamma}(\alpha_2+\alpha_3)^\gamma}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} + \frac{(\alpha_1+\alpha_3)^\gamma \alpha_2^{1+\gamma}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} + \frac{\alpha_3^{1+\gamma}}{(1+\gamma)(\alpha_1+\alpha_2+\alpha_3)}$$

and the second term is given by

$$\sum_{i=1}^n f_{X_1,X_2}^\gamma(x_{1i}, x_{2i}; \theta) = \alpha_1^\gamma(\alpha_2+\alpha_3)^\gamma \sum_{i\in I_1} e^{-\alpha_1\gamma x_{1i}} e^{-(\alpha_2+\alpha_3)\gamma x_{2i}}$$
$$+ (\alpha_1+\alpha_3)^\gamma \alpha_2^\gamma \sum_{i\in I_2} e^{-(\alpha_1+\alpha_3)\gamma x_{1i}} e^{-\alpha_2\gamma x_{2i}} + \alpha_3^\gamma \sum_{i\in I_3} e^{-(\alpha_1+\alpha_2+\alpha_3)\gamma x_i}.$$

Hence prooves the theorem. □

Proof of Theorem 4. Here $J^{-1}(\theta^0)$ and $\xi_\gamma(\theta^0)$ in the $IF(x_1, x_2; T_\gamma, F(\cdot;\theta^0))$ are independent of (x_1, x_2). Thus, only the vector $S_{\theta^0}(x_1, x_2) f^\gamma(x_1, x_2; \theta^0)$ determines the boundedness. All the three elements of $S_{\theta^0}(x_1, x_2) f^\gamma(x_1, x_2; \theta^0)$ needs to be bounded for $IF(x_1, x_2; T_\gamma, F(\cdot;\theta^0))$ being bounded.
Case 1. When $\gamma > 0$
The first component of $\psi(x_1, x_2; \theta^0) = S_{\theta^0}(x_1, x_2) f_{\theta^0}^\gamma(x_1, x_2)$ is

$$\psi_1(x_1, x_2; \theta^0) = \begin{cases} [\alpha_1^0(\alpha_2^0+\alpha_3^0)]^\gamma (1/\alpha_1^0 - x_1) e^{-\alpha_1^0\gamma x_1} e^{-(\alpha_2^0+\alpha_3^0)\gamma x_2} & \text{if } x_1 < x_2 \\ [(\alpha_1^0+\alpha_3^0)\alpha_2^0]^\gamma (1/(\alpha_1^0+\alpha_3^0) - x_1) e^{-(\alpha_1^0+\alpha_2^0)\gamma x_1} e^{-\alpha_2^0\gamma x_2} & \text{if } x_1 > x_2 \\ -[\alpha_3^0]^\gamma x_1 e^{-(\alpha_1^0+\alpha_2^0+\alpha_3^0)\gamma x_1} & \text{if} x_1 = x_2 \end{cases}$$

The constant part of $\psi_1(x_1, x_2; \theta^0)$ does not affect boundedness. The term $(c - x_1)$, where $c = 1/\alpha_1^0, 1/(\alpha_1^0+\alpha_3^0), 0$, is linear in x_1, and for small x_1, say when $x_1 < c$, it remains positive. However, as $x_1 \to \infty$, it becomes negative and increasingly large in magnitude. However, the exponential term $e^{-\alpha_1\gamma x_1}$, where $\alpha_1^0 > 0$, decays to 0 exponentially fast as $x_1 \to \infty$. Thus, although $(c - x_1)$ becomes large and negative as x_1 increases, the exponential decay $e^{-\alpha_1^0\gamma x_1}$ dominates the growth, forcing the entire function to tend to 0 as $x_1 \to \infty$. Also we have as $x_2 \to \infty$, the exponential term $e^{-\alpha_2^0\gamma x_2}$ decays to 0 exponentially fast. Therefore as $x_1 \to \infty$ and $x_2 \to \infty$, the first component $\psi_1(x_1, x_2; \theta^0)$ tends to 0. Similarly, the second component $\psi_2(x_1, x_2; \theta^0)$ and the third $\psi_3(x_1, x_2; \theta^0)$ are controlled by exponential decay terms in both x_1 and x_2, which ensure that it approaches 0 as $x_1 \to \infty$ or $x_2 \to \infty$. Since there are no terms in the functions that cause these to grow unbounded in any direction, the functions are bounded. Thus $S_{\theta^0}(x_1, x_2) f^\gamma(x_1, x_2; \theta^0)$ is bounded for all $x_1 > 0$ and $x_2 > 0$ and hence the function $IF(x_1, x_2; T_\gamma, F(\cdot;\theta^0))$ is bounded.
Case 2. When $\gamma = 0$, we have

$$\psi_1(x_1, x_2; \theta^0) = \begin{cases} (1/\alpha_1^0 - x_1) & \text{if } x_1 < x_2 \\ (1/(\alpha_1^0+\alpha_3^0) - x_1) & \text{if } x_1 > x_2 \\ -x_1 & \text{if} x_1 = x_2 \end{cases}$$

As $x_1 \to \infty$ and $x_2 \to \infty$, $|\psi_1(x_1, x_2; \theta^0)| \to \infty$. Thus the first component of IF in unbounded. Similarly we can see for second and third components. Hence for $\gamma = 0$, the MLE case, $IF(x_1, x_2; T_\gamma, F(\cdot;\theta^0))$ is unbounded. □

Appendix B Integrability conditions

The maximum likelihood score function of the MOBE model is $S_\theta(x_1, x_2) = \nabla[\log f(x_1, x_2; \theta)]$, where $\theta = (\alpha_1, \alpha_2, \alpha_3)$ and its j^{th} component is given below $S_{\theta_j}(x_1, x_2) = \partial \log f(x_1, x_2; \theta)/\partial\theta_j$, for $j = 1, 2, 3$. Then we have

$$S_{\alpha_1}(x_1, x_2) = \begin{cases} \frac{1}{\alpha_1} - x_1 & \text{if } x_1 < x_2 \\ \frac{1}{\alpha_1+\alpha_3} - x_1 & \text{if } x_1 > x_2 \\ -x_1 & \text{if } x_1 = x_2, \end{cases} \qquad S_{\alpha_2}(x_1, x_2) = \begin{cases} \frac{1}{\alpha_2+\alpha_3} - x_2 & \text{if } x_1 < x_2 \\ \frac{1}{\alpha_2} - x_2 & \text{if } x_1 > x_2 \\ -x_2 & \text{if } x_1 = x_2, \end{cases}$$

$$\text{and} \qquad S_{\alpha_3}(x_1, x_2) = \begin{cases} \frac{1}{\alpha_2+\alpha_3} - x_2 & \text{if } x_1 < x_2 \\ \frac{1}{\alpha_1+\alpha_3} - x_1 & \text{if } x_1 > x_2 \\ \frac{1}{\alpha_3} - x_1 & \text{if } x_1 = x_2 \end{cases}$$

The information matrix of the MOBE model is $I_\theta = -\nabla[S_\theta(x_1, x_2)]$, the negative of Hessian matrix. The elements are given as;

$$I_{11}(x_1, x_2) = \begin{cases} \frac{1}{\alpha_1^2} & \text{if } x_1 < x_2 \\ \frac{1}{(\alpha_1+\alpha_3)^2} & \text{if } x_1 > x_2 \\ 0 & \text{if } x_1 = x_2, \end{cases} \qquad I_{13}(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 < x_2 \\ \frac{1}{(\alpha_1+\alpha_3)^2} & \text{if } x_1 > x_2 \\ 0 & \text{if } x_1 = x_2, \end{cases}$$

$$I_{12} = I_{21} = 0,$$

$$I_{22}(x_1, x_2) = \begin{cases} \frac{1}{(\alpha_2+\alpha_3)^2} & \text{if } x_1 < x_2 \\ \frac{1}{\alpha_2^2} & \text{if } x_1 > x_2 \\ 0 & \text{if } x_1 = x_2, \end{cases} \qquad I_{23}(x_1, x_2) = \begin{cases} \frac{1}{(\alpha_2+\alpha_3)^2} & \text{if } x_1 < x_2 \\ 0 & \text{if } x_1 > x_2 \\ 0 & \text{if } x_1 = x_2, \end{cases}$$

$$\text{and} \qquad S_{\alpha_3}(x_1, x_2) = \begin{cases} \frac{1}{(\alpha_2+\alpha_3)^2} & \text{if } x_1 < x_2 \\ \frac{1}{(\alpha_1+\alpha_3)^2} & \text{if } x_1 > x_2 \\ \frac{1}{\alpha_3^2} & \text{if } x_1 = x_2. \end{cases}$$

The consistency and asymptotic normality of the MDPDE when the true distribution G is in the parametric family of the MOBE models follow from the integrability conditions, see Basu et al. [20]

$$\int_0^\infty \int_0^\infty f^{2\gamma}(x_1, x_2; \theta) g(x_1, x_2) dx_1 dx_2 < \infty \tag{B1}$$

$$\int_0^\infty \int_0^\infty S_{\alpha_i}(x_1,x_2) f^{2\gamma}(x_1,x_2;\theta) g(x_1,x_2) dx_1 dx_2 < \infty \qquad i=1,2,3 \tag{B2}$$

$$\int_0^\infty \int_0^\infty \left| \frac{\partial^2 log f(x_1,x_2;\theta)}{\partial \alpha_i \partial \alpha_j} \right| f^{2\gamma}(x_1,x_2;\theta) g(x_1,x_2) dx_1 dx_2 < \infty \qquad i \neq j = 1,2,3. \tag{B3}$$

Appendix C Asymptotic covariance matrix of MDPDE

The components of the $\xi_\gamma(\theta)$ are given as

$$\begin{aligned}\xi_\gamma(\alpha_1) = {} & \frac{\alpha_1^\gamma(\alpha_2+\alpha_3)^\gamma}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)}\left[1-\frac{\alpha_1}{(1+\gamma)(\alpha_1+\alpha_2+\alpha_3)}\right] + \frac{(\alpha_1+\alpha_3)^{\gamma-1}\alpha_2^{1+\gamma}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} \\ & - \frac{(\alpha_1+\alpha_3)^{1+\gamma}\alpha_2^\gamma}{(1+\gamma)^3}\left[\frac{1}{(\alpha_1+\alpha_3)^2} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right] - \frac{\alpha_3^{1+\gamma}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)^2}\end{aligned}$$

$$\begin{aligned}\xi_\gamma(\alpha_2) = {} & \frac{\alpha_1^{1+\gamma}(\alpha_2+\alpha_3)^{\gamma-1}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} - \frac{\alpha_1^\gamma(\alpha_2+\alpha_3)^{1+\gamma}}{(1+\gamma)^3}\left[\frac{1}{(\alpha_2+\alpha_3)^2} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right] \\ & + \frac{(\alpha_1+\alpha_3)^\gamma\alpha_2^\gamma}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)}\left[1-\frac{\alpha_2}{(1+\gamma)(\alpha_1+\alpha_2+\alpha_3)}\right] - \frac{\alpha_3^{1+\gamma}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)^2}\end{aligned}$$

$$\begin{aligned}\xi_\gamma(\alpha_3) = {} & \frac{\alpha_1^{1+\gamma}(\alpha_2+\alpha_3)^{\gamma-1}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} - \frac{\alpha_1^\gamma(\alpha_2+\alpha_3)^{1+\gamma}}{(1+\gamma)^3}\left[\frac{1}{(\alpha_2+\alpha_3)^2} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right] \\ & + \frac{(\alpha_1+\alpha_3)^{\gamma-1}\alpha_2^{1+\gamma}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} - \frac{(\alpha_1+\alpha_3)^{1+\gamma}\alpha_2^\gamma}{(1+\gamma)^3}\left[\frac{1}{(\alpha_1+\alpha_3)^2} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right] \\ & + \frac{\alpha_3^\gamma}{(1+\gamma)(\alpha_1+\alpha_2+\alpha_3)}\left[1-\frac{\alpha_3}{(1+\gamma)(\alpha_1+\alpha_2+\alpha_3)}\right].\end{aligned}$$

Now we derive the elements of asymptotic covariance matrix of MDPDE for the MOBE model under the model condition, i.e., $G(x_1,x_2) = F(x_1,x_2;\theta)$ for some $\theta \in \Theta$. Te elements of the 3×3 symmetric matrix $J_\gamma(\theta)$ are

$$\begin{aligned}J_{11} = {} & \frac{\alpha_1^{\gamma-1}(\alpha_2+\alpha_3)^\gamma}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)}\left[1-\frac{2\alpha_1}{(1+\gamma)(\alpha_1+\alpha_2+\alpha_3)} + \frac{2\alpha_1^2}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)^2}\right] \\ & + \frac{(\alpha_1+\alpha_3)^{\gamma-2}\alpha_2^{1+\gamma}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} - \frac{2(\alpha_1+\alpha_3)^\gamma\alpha_2^\gamma}{(1+\gamma)^3}\left[\frac{1}{(\alpha_1+\alpha_3)^2} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right] \\ & + \frac{2(\alpha_1+\alpha_3)^{1+\gamma}\alpha_2^\gamma}{(1+\gamma)^4}\left[\frac{1}{(\alpha_1+\alpha_3)^3} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^3}\right] + \frac{2\alpha_3^{1+\gamma}}{(1+\gamma)^3(\alpha_1+\alpha_2+\alpha_3)^3}\end{aligned}$$

$$
\begin{aligned}
J_{12} = & \frac{\alpha_1^{\gamma}(\alpha_2+\alpha_3)^{\gamma-1}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} - \frac{\alpha_1^{1+\gamma}(\alpha_2+\alpha_3)^{\gamma-1}}{(1+\gamma)^3(\alpha_1+\alpha_2+\alpha_3)^2} \\
& - \frac{\alpha_1^{\gamma-1}(\alpha_2+\alpha_3)^{1+\gamma}}{(1+\gamma)^3}\left[\frac{1}{(\alpha_2+\alpha_3)^2} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right] \\
& + \frac{\alpha_1^{1+\gamma}(\alpha_2+\alpha_3)^{\gamma}}{(1+\gamma)^4(\alpha_1+\alpha_2+\alpha_3)^2}\left[\frac{2}{(\alpha_1+\alpha_2+\alpha_3)} + \frac{1}{(\alpha_2+\alpha_3)}\right] + \frac{(\alpha_1+\alpha_3)^{\gamma-1}\alpha_2^{\gamma}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} \\
& - \frac{(\alpha_1+\alpha_3)^{1+\gamma}\alpha_2^{\gamma-1}}{(1+\gamma)^3}\left[\frac{1}{(\alpha_1+\alpha_3)^2} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right] - \frac{(\alpha_1+\alpha_3)^{\gamma-1}\alpha_2^{1+\gamma}}{(1+\gamma)^3(\alpha_1+\alpha_2+\alpha_3)^2} \\
& + \frac{(\alpha_1+\alpha_3)^{\gamma}\alpha_2^{1+\gamma}}{(1+\gamma)^4(\alpha_1+\alpha_2+\alpha_3)^2}\left[\frac{2}{(\alpha_1+\alpha_2+\alpha_3)} + \frac{1}{(\alpha_1+\alpha_3)}\right] + \frac{2\alpha_3^{1+\gamma}}{(1+\gamma)^3(\alpha_1+\alpha_2+\alpha_3)^3}
\end{aligned}
$$

$$
\begin{aligned}
J_{13} = & \frac{\alpha_1^{\gamma}(\alpha_2+\alpha_3)^{\gamma-1}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} - \frac{\alpha_1^{1+\gamma}(\alpha_2+\alpha_3)^{\gamma-1}}{(1+\gamma)^3(\alpha_1+\alpha_2+\alpha_3)^2} \\
& - \frac{\alpha_1^{\gamma-1}(\alpha_2+\alpha_3)^{1+\gamma}}{(1+\gamma)^3}\left[\frac{1}{(\alpha_2+\alpha_3)^2} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right] \\
& + \frac{\alpha_1^{1+\gamma}(\alpha_2+\alpha_3)^{\gamma}}{(1+\gamma)^4(\alpha_1+\alpha_2+\alpha_3)^2}\left[\frac{2}{(\alpha_1+\alpha_2+\alpha_3)} + \frac{1}{(\alpha_2+\alpha_3)}\right] + \frac{(\alpha_1+\alpha_3)^{\gamma-2}\alpha_2^{1+\gamma}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} \\
& - \frac{2(\alpha_1+\alpha_3)^{\gamma}\alpha_2^{\gamma}}{(1+\gamma)^3}\left[\frac{1}{(\alpha_1+\alpha_3)^2} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right] \\
& + \frac{2(\alpha_1+\alpha_3)^{1+\gamma}\alpha_2^{\gamma}}{(1+\gamma)^4}\left[\frac{1}{(\alpha_1+\alpha_3)^3} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^3}\right] \\
& - \frac{\alpha_3^{\gamma}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)^2}\left[1 - \frac{2\alpha_3}{(1+\gamma)(\alpha_1+\alpha_2+\alpha_3)}\right]
\end{aligned}
$$

$$
\begin{aligned}
J_{22} = & \frac{\alpha_1^{1+\gamma}(\alpha_2+\alpha_3)^{\gamma-2}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} - \frac{2\alpha_1^{\gamma}(\alpha_2+\alpha_3)^{\gamma}}{(1+\gamma)^3}\left[\frac{1}{(\alpha_2+\alpha_3)^2} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right] \\
& + \frac{2\alpha_1^{\gamma}(\alpha_2+\alpha_3)^{1+\gamma}}{(1+\gamma)^4}\left[\frac{1}{(\alpha_2+\alpha_3)^3} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^3}\right] \\
& + \frac{(\alpha_1+\alpha_3)^{\gamma}\alpha_2^{\gamma-1}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)}\left[1 - \frac{2\alpha_2}{(1+\gamma)(\alpha_1+\alpha_2+\alpha_3)} + \frac{2\alpha_2^2}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)^2}\right] \\
& + \frac{2\alpha_3^{1+\gamma}}{(1+\gamma)^3(\alpha_1+\alpha_2+\alpha_3)^3}
\end{aligned}
$$

$$
\begin{aligned}
J_{23} = & \frac{\alpha_1^{1+\gamma}(\alpha_2+\alpha_3)^{\gamma-2}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)} - \frac{2\alpha_1^{\gamma}(\alpha_2+\alpha_3)^{\gamma}}{(1+\gamma)^3}\left[\frac{1}{(\alpha_2+\alpha_3)^2} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right] \\
& + \frac{2\alpha_1^{\gamma}(\alpha_2+\alpha_3)^{1+\gamma}}{(1+\gamma)^4}\left[\frac{1}{(\alpha_2+\alpha_3)^3} - \frac{1}{(\alpha_1+\alpha_2+\alpha_3)^3}\right]
\end{aligned}
$$

$$+\frac{(\alpha_1+\alpha_3)^{\gamma-1}\alpha_2^{\gamma}}{(1+\gamma)^2((\alpha_1+\alpha_2+\alpha_3))}\left[1-\frac{\alpha_2}{(1+\gamma)(\alpha_1+\alpha_2+\alpha_3)}\right]$$
$$-\frac{(\alpha_1+\alpha_3)^{1+\gamma}\alpha_2^{\gamma-1}}{(1+\gamma)^3}\left[\frac{1}{(\alpha_1+\alpha_3)^2}-\frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right]$$
$$+\frac{(\alpha_1+\alpha_3)^{\gamma}\alpha_2^{1+\gamma}}{(1+\gamma)^4(\alpha_1+\alpha_2+\alpha_3)^2}\left[\frac{2}{(\alpha_1+\alpha_2+\alpha_3)}+\frac{1}{(\alpha_1+\alpha_3)}\right]$$
$$-\frac{\alpha_3^{\gamma}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)^2}\left[1-\frac{2\alpha_3}{(1+\gamma)(\alpha_1+\alpha_2+\alpha_3)}\right]$$

$$J_{33}=\frac{\alpha_1^{1+\gamma}(\alpha_2+\alpha_3)^{\gamma-2}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)}-\frac{2\alpha_1^{\gamma}(\alpha_2+\alpha_3)^{\gamma}}{(1+\gamma)^3}\left[\frac{1}{(\alpha_2+\alpha_3)^2}-\frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right]$$
$$+\frac{2\alpha_1^{\gamma}(\alpha_2+\alpha_3)^{1+\gamma}}{(1+\gamma)^4}\left[\frac{1}{(\alpha_2+\alpha_3)^3}-\frac{1}{(\alpha_1+\alpha_2+\alpha_3)^3}\right]$$
$$+\frac{(\alpha_1+\alpha_3)^{\gamma-2}\alpha_2^{1+\gamma}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)}-\frac{2(\alpha_1+\alpha_3)^{\gamma}\alpha_2^{\gamma}}{(1+\gamma)^3}\left[\frac{1}{(\alpha_1+\alpha_3)^2}-\frac{1}{(\alpha_1+\alpha_2+\alpha_3)^2}\right]$$
$$+\frac{2(\alpha_1+\alpha_3)^{1+\gamma}\alpha_2^{\gamma}}{(1+\gamma)^4}\left[\frac{1}{(\alpha_1+\alpha_3)^3}-\frac{1}{(\alpha_1+\alpha_2+\alpha_3)^3}\right]$$
$$+\frac{\alpha_3^{\gamma-1}}{(1+\gamma)(\alpha_1+\alpha_2+\alpha_3)}-\frac{2\alpha_3^{\gamma}}{(1+\gamma)^2(\alpha_1+\alpha_2+\alpha_3)^2}+\frac{2\alpha_3^{1+\gamma}}{(1+\gamma)^3(\alpha_1+\alpha_2+\alpha_3)^3}.$$

References

[1] Marshall, A.W., Olkin, I.: A multivariate exponential distribution. Journal of the American Statistical Association **62**(317), 30–44 (1967)

[2] Basu, A., Harris, I.R., Hjort, N.L., Jones, M.: Robust and efficient estimation by minimising a density power divergence. Biometrika **85**(3), 549–559 (1998)

[3] Meintanis, S.G.: Test of fit for Marshall–Olkin distributions with applications. Journal of Statistical Planning and Inference **137**(12), 3954–3963 (2007)

[4] Bemis, B.M., Bain, L.J., Higgins, J.J.: Estimation and hypothesis testing for the parameters of a bivariate exponential distribution. Journal of the American Statistical Association **67**(340), 927–929 (1972)

[5] Arnold, B.C.: Parameter estimation for a multivariate exponential distribution. Journal of the American Statistical Association **63**(323), 848–852 (1968)

[6] Bhattacharyya, A.: Maximum likelihood estimation of a system of parameters. Sankhya: The Indian Journal of Statistics, Series A **33**(1), 1–10 (1971)

[7] Peña, E.A., Gupta, A.K.: Bayes estimation for the Marshall–Olkin exponential distribution. Journal of the Royal Statistical Society Series B: Statistical

Methodology **52**(2), 379–389 (1990)

[8] Karlis, D.: ML estimation for multivariate shock models via an EM algorithm. Annals of the Institute of Statistical Mathematics **55**, 817–830 (2003)

[9] Juárez, S.F., Schucany, W.R.: Robust and efficient estimation for the generalized pareto distribution. Extremes **7**, 237–251 (2004)

[10] Hazra, A.: Minimum density power divergence estimation for the generalized exponential distribution. Communications in Statistics-Theory and Methods **54**(4), 1050–1070 (2025)

[11] Almongy, H.M., Almetwally, E.M.: Robust estimation methods of generalized exponential distribution with outliers. Pak. J. Stat. Oper. Res. **16**, 545–559 (2020)

[12] Kumar, S., Kundu, D., Mitra, S.: Robust and efficient estimation for the Block-Basu bivariate exponential distribution. Statistics and Computing **35**(5), 150 (2025) https://doi.org/10.1007/s11222-025-10687-7

[13] Kullback, S., Leibler, R.A.: On information and sufficiency. The Annals of Mathematical Statistics **22**(1), 79–86 (1951)

[14] Jones, L.K., Byrne, C.L.: General entropy criteria for inverse problems, with applications to data compression, pattern classification, and cluster analysis. IEEE Transactions on Information Theory **36**(1), 23–30 (1990)

[15] Basu, A., Lindsay, B.G.: Minimum disparity estimation for continuous models: efficiency, distributions and robustness. Annals of the Institute of Statistical Mathematics **46**, 683–705 (1994)

[16] Beran, R.: Minimum Hellinger distance estimates for parametric models. The Annals of Statistics **5**(3), 445–463 (1977)

[17] Nelder, J.A., Mead, R.: A simplex method for function minimization. The Computer Journal **7**(4), 308–313 (1965)

[18] Huber, P.J.: Robust Statistical Procedures, 2nd edn. SIAM, Philadelphia, PA (1996). https://doi.org/10.1137/1.9781611970677

[19] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A.: Robust Statistics: The Approach Based on Influence Functions. John Wiley & Sons, New York (2011)

[20] Basu, A., Shioya, H., Park, C.: Statistical Inference: The Minimum Distance Approach. CRC Press, Boca Raton, FL (2011). https://doi.org/10.1201/b10905

[21] Hampel, F.R.: The influence curve and its role in robust estimation. Journal of

the American Statistical Association **69**(346), 383–393 (1974)

[22] Hong, C., Kim, Y.: Automatic selection of the turning parameter in the minimum density power divergence estimation. Journal of the Korean Statistical Society **30**(3), 453–465 (2001)

[23] Warwick, J., Jones, M.: Choosing a robustness tuning parameter. Journal of Statistical Computation and Simulation **75**(7), 581–588 (2005)

[24] Ghosh, A., Basu, A.: Robust estimation for non-homogeneous data and the selection of the optimal tuning parameter: the density power divergence approach. Journal of Applied Statistics **42**(9), 2056–2072 (2015)

[25] Basak, S., Basu, A., Jones, M.: On the 'optimal' density power divergence tuning parameter. Journal of Applied Statistics **48**(3), 536–556 (2021)

[26] Kundu, D., Dey, A.K.: Estimating the parameters of the Marshall–Olkin bivariate Weibull distribution by EM algorithm. Computational Statistics & Data Analysis **53**(4), 956–965 (2009)

[27] Rousseeuw, P.J., Ruts, I., Tukey, J.W.: The bagplot: a bivariate boxplot. The American Statistician **53**(4), 382–387 (1999)

[28] Kumar, S., Kundu, D., Mitra, S.: Estimation of five parameter bivariate modified Weibull singular distribution. Communications in Statistics - Simulation and Computation, 1–24 (2024) https://doi.org/10.1080/03610918.2024.2434709

[29] Presnell, B., Boos, D.D.: The IOS test for model misspecification. Journal of the American Statistical Association **99**(465), 216–227 (2004)

[30] Fujisawa, H., Eguchi, S.: Robust estimation in the normal mixture model. Journal of Statistical Planning and Inference **136**(11), 3989–4011 (2006)