stats

MDPI

Article

On the Bivariate Composite Gumbel–Pareto Distribution

Alexandra Badea [1], Catalina Bolancé [2] and Raluca Vernic [3,*]

[1] Doctoral School, Ovidius University of Constanta, 900527 Constanta, Romania
[2] Department of Econometrics, RISKcenter-IREA, Universitat de Barcelona, 08034 Barcelona, Spain
[3] Faculty of Mathematics and Computer Science, Ovidius University of Constanta, 900527 Constanta, Romania
* Correspondence: rvernic@univ-ovidius.ro

Abstract: In this paper, we propose a bivariate extension of univariate composite (two-spliced) distributions defined by a bivariate Pareto distribution for values larger than some thresholds and by a bivariate Gumbel distribution on the complementary domain. The purpose of this distribution is to capture the behavior of bivariate data consisting of mainly small and medium values but also of some extreme values. Some properties of the proposed distribution are presented. Further, two estimation procedures are discussed and illustrated on simulated data and on a real data set consisting of a bivariate sample of claims from an auto insurance portfolio. In addition, the risk of loss in this insurance portfolio is estimated by Monte Carlo simulation.

Keywords: bivariate composite (two-spliced) distribution; Gumbel's bivariate exponential distribution; bivariate Pareto of the first kind distribution; maximum likelihood estimation procedure; risk of loss

Citation: Badea, A.; Bolancé, C.; Vernic, R. On the Bivariate Composite Gumbel–Pareto Distribution. *Stats* **2022**, *5*, 948–969. https://doi.org/10.3390/stats5040055

Academic Editor: Wei Zhu

Received: 31 August 2022
Accepted: 13 October 2022
Published: 16 October 2022

1. Introduction

Dependent multivariate data frequently occur in practice in areas such as insurance, finance, economics, reliability, etc. Therefore, the development of bivariate and multivariate distributions is a very active field of research, especially since—in contrast to univariate distributions—it gained interest later on. Nowadays, there are various methods of constructing multivariate distributions, see e.g., the review [1]. Some of these methods follow lines from the univariate distributions. In this sense, in this paper, we propose a bivariate composite distribution built on the same idea as the univariate composite (or two-spliced) distribution (see [2] for the splicing method in the univariate case).

Two-component spliced distributions are usually encountered in univariate extreme value theory, where a classical heavy-tailed distribution (such as the generalized Pareto) is used to model the tail, in combination with a less heavy-tailed distribution used for the so-called *bulk* model; see, e.g., the review [3]. More precisely, such a distribution is defined from two different distributions on distinct intervals, with the aim to better capture tails of distributions such as the loss ones. A two-component spliced distribution was called *composite* in [4], where a particular form of such distribution, namely the lognormal–Pareto composite distribution, was studied in connection with skewed and heavy-tailed loss data.

Therefore, the bivariate distribution we propose equals a certain bivariate distribution on one domain and another bivariate distribution on another domain. More precisely, we aim at using a more heavy-tailed bivariate distribution beyond some thresholds, such as the Pareto one. As in the univariate case, the motivation of such a model is to better capture the behavior of dependent random data that present many small and medium pairs of values but also some very large ones; we note that this could be the case with, e.g., insurance or financial data arising from two dependent lines of business. In this sense, we recall the discussions in [5,6], where it was noticed that for the particular bivariate insurance data set under study (consisting of auto claims, property damage costs, and medical expenses), the best globally fitted distribution does not provide the best model for tail risk measures because heavier-tailed distribution is needed.

Thus, in this paper, we consider the bivariate Pareto distribution of the first kind for the tail (i.e., from some thresholds on) and the bivariate Gumbel exponential distribution for the remaining domain. In Section 2, we define some notation and recall the just-mentioned bivariate distributions. In Section 3, we define the general bivariate composite distribution, while in Section 4, we introduce the particular composite Gumbel–Pareto distribution and study some continuity conditions, marginal distributions, and moments. Further, we discuss simulation from this particular bivariate distribution, and in order to reduce the computing time, we propose two procedures for parameter estimation: the first one is based on marginal estimation and completed by a limited full Maximum Likelihood Estimation (MLE), and the second one is based on conditional MLE. The estimation procedures are illustrated on simulated data in Section 5.1 and on a real auto insurance data set in Section 5.2, followed by a conclusions section. The paper ends with Appendix A containing the proofs.

2. Preliminaries

2.1. Notation

We shall use the incomplete gamma function (or generalized plica function) defined by

$$\Gamma(\alpha, z_0, z_1) = \int_{z_0}^{z_1} x^{\alpha-1} e^{-x} dx, z_1 > z_0 \geq 0.$$

We also introduce the notation

$$\Gamma(\alpha, z_0, z_1; k) = \int_{z_0}^{z_1} x^{\alpha-1} e^{-kx} dx, z_1 > z_0 \geq 0, k > 0,$$

and note that

$$\Gamma(\alpha, z_0, z_1; k) = \frac{1}{k^{\alpha}} \Gamma(\alpha, kz_0, kz_1).$$

We recall the exponential integral notation

$$E_1(z) = \int_z^{\infty} \frac{e^{-t}}{t} dt.$$

The following result holds (its proof is given in Appendix A).

Lemma 1. *With the above notation, with* $0 \leq z_0 < z_1$,

$$\begin{aligned}
(i)\ \Gamma(-1, z_0, z_1) &= \frac{e^{-z_0}}{z_0} - \frac{e^{-z_1}}{z_1} - \Gamma(0, z_0, z_1), z_0 > 0. \\
(ii)\ \Gamma(1, z_0, z_1; k) &= \frac{1}{k}\left(e^{-kz_0} - e^{-kz_1}\right). \\
(iii)\ \Gamma(2, z_0, z_1; k) &= \frac{1}{k^2}\left[e^{-kz_0}(kz_0 + 1) - e^{-kz_1}(kz_1 + 1)\right].
\end{aligned}$$

In particular,

$$\begin{aligned}
(iii.1)\ \Gamma(2, 0, \theta; k) &= \frac{1}{k^2}\left[1 - (1 + k\theta)e^{-k\theta}\right], \\
(iii.2)\ \Gamma(2, \theta, \infty; k) &= \frac{e^{-k\theta}}{k^2}(1 + k\theta).
\end{aligned}$$

$$(iv)\ \Gamma(3, z_0, z_1; k) = \frac{1}{k^3}\left[e^{-kz_0}\left(k^2 z_0^2 + 2kz_0 + 2\right) - e^{-kz_1}\left(k^2 z_1^2 + 2kz_1 + 2\right)\right].$$

In particular,

$$\begin{aligned}(iv.1)\ \Gamma(3,0,\theta;k) &= \frac{1}{k^3}\left[2-\left(2+2k\theta+k^2\theta^2\right)e^{-k\theta}\right],\\ (iv.2)\ \Gamma(3,\theta,\infty;k) &= \frac{1}{k^3}e^{-k\theta}\left(k^2\theta^2+2k\theta+2\right)=\frac{e^{-k\theta}}{k^3}\left[(k\theta+1)^2+1\right].\end{aligned}$$

$$\begin{aligned}(v)\ \int_c^\infty \frac{e^{-ky}}{y}dy &= E_1(ck), k>0,\\ (vi)\ \int_c^\infty \frac{e^{-ky}}{y^2}dy &= \frac{e^{-ck}}{c}-kE_1(ck).\end{aligned}$$

For $\theta_1>0,\theta_2>0$, we also define the following domains

$$\begin{aligned}D_{11} &= \{(x_1,x_2)|0<x_1\le\theta_1, 0<x_2\le\theta_2\},\\ D_{12} &= \{(x_1,x_2)|0<x_1\le\theta_1, x_2>\theta_2\},\\ D_{21} &= \{(x_1,x_2)|x_1>\theta_1, 0<x_2\le\theta_2\},\\ D_{22} &= \{(x_1,x_2)|x_1>\theta_1, x_2>\theta_2\}\\ D &= D_{11}\cup D_{12}\cup D_{21}.\end{aligned}$$

2.2. Bivariate Classical Distributions

The following two bivariate continuous distributions are used in the bivariate composite model.

2.2.1. Gumbel's Bivariate Exponential Distribution, Gu_2

Gumbel's [7] bivariate exponential distribution has pdf (see also [8])

$$e^{-(x_1+x_2+\beta x_1x_2)}[(1+\beta x_1)(1+\beta x_2)-\beta], x_1>0, x_2>0,$$

with standard exponential marginal distributions. We shall, however, consider a more general bivariate pdf, having general exponential distributions (see e.g., [9]). Therefore, let $\mathbf{Y}=(Y_1,Y_2)$ follow Gumbel's bivariate exponential distribution, $\mathbf{Y}\sim Gu_2(\lambda_1,\lambda_2,\beta),\lambda_1,\lambda_2>0, 0\le\beta\le 1$, defined by the joint pdf

$$g_{\mathbf{Y}}(x_1,x_2)=\lambda_1\lambda_2e^{-(\lambda_1x_1+\lambda_2x_2+\beta\lambda_1\lambda_2x_1x_2)}[(1+\beta\lambda_1x_1)(1+\beta\lambda_2x_2)-\beta], x_1>0, x_2>0. \quad (1)$$

Its cdf is given by

$$F_{\mathbf{Y}}(x_1,x_2)=\Pr(Y_1\le x_1, Y_2\le x_2)=1-e^{-\lambda_1x_1}-e^{-\lambda_2x_2}+e^{-(\lambda_1x_1+\lambda_2x_2+\beta\lambda_1\lambda_2x_1x_2)}; \quad (2)$$

its joint survival function is

$$\bar{F}_{\mathbf{Y}}(x_1,x_2)=\Pr(Y_1>x_1,Y_2>x_2)=e^{-(\lambda_1x_1+\lambda_2x_2+\beta\lambda_1\lambda_2x_1x_2)};$$

while the marginal distributions are exponentials with pdf $g_{Y_i}(x_i)=\lambda_ie^{-\lambda_ix_i}, x_i>0$, cdf $G_{Y_i}(x_i)=1-e^{-\lambda_ix_i}$, and expected value $\lambda_i^{-1}, i=1,2$.

In view of the bivariate composite model defined in the next section, an easy calculation yields the following lemma.

Lemma 2. *Let $\mathbf{Y}\sim Gu_2(\lambda_1,\lambda_2,\beta)$ and $\theta_1>0,\theta_2>0$. Then, with the above notation, it holds that*

$$P_D=\Pr(\mathbf{Y}\in D)=1-e^{-(\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2)}. \quad (3)$$

The next lemmas are also needed.

Lemma 3. *If* $\mathbf{Y} \sim Gu_2(\lambda_1, \lambda_2, \beta)$ *and* $\theta_1 > 0, \theta_2 > 0$, *then*

$$\begin{aligned} L_1(x_1; \theta_2) &= \int_0^{\theta_2} g_{\mathbf{Y}}(x_1, x_2) dx_2 = \lambda_1 e^{-\lambda_1 x_1} \Big[1 - (1 + \beta\lambda_2\theta_2) e^{-\lambda_2\theta_2(1+\beta\lambda_1 x_1)} \Big], x_1 > 0, \\ L_2(x_2; \theta_1) &= \int_0^{\theta_1} g_{\mathbf{Y}}(x_1, x_2) dx_1 = \lambda_2 e^{-\lambda_2 x_2} \Big[1 - (1 + \beta\lambda_1\theta_1) e^{-\lambda_1\theta_1(1+\beta\lambda_2 x_2)} \Big], x_2 > 0. \end{aligned}$$

Lemma 4. *If* $\mathbf{Y} \sim Gu_2(\lambda_1, \lambda_2, \beta)$ *and* $\theta_1 > 0, \theta_2 > 0$, *then*

$$\begin{aligned} I(\theta_1, \theta_2) &= \int_{\theta_1}^{\infty} \int_{\theta_2}^{\infty} x_1 x_2 g_{\mathbf{Y}}(x_1, x_2) dx_1 dx_2 \\ &= \frac{1}{\beta\lambda_1\lambda_2} \Bigg[\left(2 - \frac{1}{1+\beta\lambda_1\theta_1} - \frac{1}{1+\beta\lambda_2\theta_2} + \beta\lambda_1\lambda_2\theta_1\theta_2 \right) e^{-(\lambda_1\theta_1 + \lambda_2\theta_2 + \beta\lambda_1\lambda_2\theta_1\theta_2)} \\ &\quad + E_1\left(\frac{(1+\beta\lambda_1\theta_1)(1+\beta\lambda_2\theta_2)}{\beta} \right) e^{\frac{1}{\beta}} \Bigg]. \end{aligned}$$

Lemma 5. *Given that* $Y_1 = y_1$, *the conditional cdf of the marginal* Y_2 *of* $\mathbf{Y} \sim Gu_2(\lambda_1, \lambda_2, \beta)$ *is*

$$F_{Y_2|Y_1=y_1}(y_2) = 1 - (1 + \beta\lambda_2 y_2) e^{-\lambda_2 y_2 (1+\beta\lambda_1 y_1)}.$$

2.2.2. Bivariate Pareto Distribution of the First Kind, PaI_2

Let $\mathbf{Z} = (Z_1, Z_2)$ follow the bivariate Pareto of the first kind distribution, $\mathbf{Z} \sim PaI_2(a, \theta_1, \theta_2), a > 0, \theta_1 > 0, \theta_2 > 0$. Its pdf is (see [10])

$$f_{\mathbf{Z}}(x_1, x_2) = a(a+1) \frac{(\theta_1\theta_2)^{a+1}}{(\theta_2 x_1 + \theta_1 x_2 - \theta_1\theta_2)^{a+2}}, x_1 > \theta_1, x_2 > \theta_2. \quad (4)$$

Its marginal distributions are univariate Pareto of the first kind, having pdf and cdf, respectively,

$$f_{Z_i}(x_i) = a \frac{\theta_i^a}{x_i^{a+1}}, F_{Z_i}(x_i) = 1 - \left(\frac{\theta_i}{x_i} \right)^a, x_i > \theta_i, i = 1, 2.$$

Moreover, we recall the formulas of the expected values and variances

$$\mathbb{E}Z_i = \frac{a\theta_i}{a-1}, a > 1,\ Var(Z_i) = \frac{a\theta_i^2}{(a-1)^2(a-2)}, a > 2, i = 1, 2,$$

while the formula of the covariance is

$$cov(Z_1, Z_2) = \frac{\theta_1\theta_2}{(a-1)^2(a-2)}.$$

From here, it is easy to see that

$$\mathbb{E}[Z_1 Z_2] = \theta_1\theta_2 \frac{a^2 - a - 1}{(a-1)(a-2)}. \quad (5)$$

3. A Bivariate Composite Model

We shall now define the bivariate composite model. Let $\mathbf{X} = (X_1, X_2)$ be a bivariate random vector, and let $\theta_1, \theta_2 \in \mathbb{R}$. We say that $\mathbf{X}$ follows a bivariate composite distribution if its pdf is defined as

$$\begin{aligned} f(x_1, x_2) &= \begin{cases} r f_1(x_1, x_2), \{x_1 \le \theta_1, x_2 \le \theta_2\} \cup \{x_1 \le \theta_1, x_2 > \theta_2\} \cup \{x_1 > \theta_1, x_2 \le \theta_2\} \\ (1-r) f_2(x_1, x_2), x_1 > \theta_1, x_2 > \theta_2 \end{cases} \\ &= \begin{cases} r f_1(x_1, x_2), (x_1, x_2) \in D \\ (1-r) f_2(x_1, x_2), (x_1, x_2) \in D_{22} \end{cases}, \end{aligned} \quad (6)$$

where $0 \leq r \leq 1$ is a normalizing constant. We note that, in general, f_1 and f_2 are pdfs of distributions truncated on the domains D and D_{22}, respectively. Therefore, we can rewrite this composite distribution as a two-component mixture model with mixing weights r and $1 - r$, i.e.,

$$f(x_1, x_2) = r f_1(x_1, x_2) + (1 - r) f_2(x_1, x_2). \quad (7)$$

This form can be used for random number generation.

We would like our pdf to be at least continuous. However, in this case, the bivariate density changes shape on the line segments $\{x_1 = \theta_1, x_2 > \theta_2\}$ and $\{x_1 > \theta_1, x_2 = \theta_2\}$, which generally restricts the continuity condition; more precisely, imposing continuity on, e.g., the first segment, results in

$$\frac{r}{P_D} f_1(\theta_1, x_2) = (1 - r) f_2(\theta_1, x_2), \quad (8)$$

which, in general, cannot be satisfied for all $x_2 > \theta_2$. We can impose a continuity condition at (θ_1, θ_2) and obtain the restriction for r

$$r = \left(1 + \frac{f_1(\theta_1, \theta_2)}{f_2(\theta_1, \theta_2)}\right)^{-1}. \quad (9)$$

We can also impose continuity conditions to the marginal pdfs, since each one is two-spliced as we see in next section.

4. Particular Case: Bivariate Composite Gumbel–Pareto Distribution

In particular, we shall assume that f_1 is the pdf of a Gumbel bivariate exponential distribution truncated on the domain D, and that f_2 is a bivariate Pareto pdf defined on D_{22}, which is left truncated by its nature. Therefore, let $\theta_1 > 0, \theta_2 > 0$, and let $\mathbf{Y} = (Y_1, Y_2)$ follow Gumbel's bivariate distribution (1) truncated on the domain D, with parameters $\lambda_1, \lambda_2 > 0, \beta \in [0, 1]$, and having pdf

$$f_1(x_1, x_2) = \frac{g_{\mathbf{Y}}(x_1, x_2)}{P_D}, (x_1, x_2) \in D.$$

Additionally, let $\mathbf{Z} = (Z_1, Z_2) \sim PaI_2(a, \theta_1, \theta_2), a > 0$. Then, using P_D from (3), the pdf (6) of $\mathbf{X}$ becomes

$$f(x_1, x_2) = \begin{cases} r \frac{\lambda_1 \lambda_2 e^{-(\lambda_1 x_1 + \lambda_2 x_2 + \beta \lambda_1 \lambda_2 x_1 x_2)}}{1 - e^{-(\lambda_1 \theta_1 + \lambda_2 \theta_2 + \beta \lambda_1 \lambda_2 \theta_1 \theta_2)}} [(1 + \beta \lambda_1 x_1)(1 + \beta \lambda_2 x_2) - \beta], (x_1, x_2) \in D \\ (1 - r) a (a + 1) \frac{(\theta_1 \theta_2)^{a+1}}{(\theta_2 x_1 + \theta_1 x_2 - \theta_1 \theta_2)^{a+2}}, (x_1, x_2) \in D_{22} \end{cases}. \quad (10)$$

Note that by taking $r = 0$, we obtain the bivariate Pareto pdf; with $r = 1$, we obtain the bivariate Gumbel truncated on the domain D; if we take $r = 1$ and $\theta_1 = \theta_2 = 0$, (10) reduces to the usual Gumbel pdf. If $\beta = 0$, the Gumbel component becomes the bivariate exponential with independent marginals.

If we impose the continuity condition at (θ_1, θ_2), we obtain the following formula of r

$$r = \left(1 + \frac{g_{\mathbf{Y}}(\theta_1, \theta_2)}{f_2(\theta_1, \theta_2) P_D}\right)^{-1} = \left(1 + \frac{(1 + \beta \lambda_1 \theta_1)(1 + \beta \lambda_2 \theta_2) - \beta}{\left(e^{\lambda_1 \theta_1 + \lambda_2 \theta_2 + \beta \lambda_1 \lambda_2 \theta_1 \theta_2} - 1\right)(a + 1) a} \lambda_1 \lambda_2 \theta_1 \theta_2\right)^{-1}. \quad (11)$$

In the left side of Figure 1, we plotted a composite Gumbel–Pareto pdf satisfying marginal continuity conditions and the continuity condition at (θ_1, θ_2); see (*iii*) in Proposition 2. However, as discussed above, this pdf is not continuous everywhere; e.g., the continuity condition (8) becomes, in this case, $\frac{r}{P_D} g_{\mathbf{Y}}(\theta_1, x_2) = (1 - r) f_{\mathbf{Z}}(\theta_1, x_2)$, which, given the pdfs $g_{\mathbf{Y}}$ and $f_{\mathbf{Z}}$, cannot be satisfied for all $x_2 > \theta_2$. This can be seen from the right plot of the same figure, where we focused better on the threshold lines $\{x_1 = \theta_1, x_2 > \theta_2\}$ and $\{x_1 > \theta_1, x_2 = \theta_2\}$.

In Figure 2, we plotted another composite Gumbel–Pareto pdf with different parameters and all continuity conditions, having a more heavy-tailed Pareto component ($a < 1$). A certain flexibility of the pdf's shape can be noticed from the two plots. However, in both pdf plots, note the areas of strong decrease for small values of x_1 and x_2 due to the exponential characteristic of the Gumbel distribution.

We also plotted in Figure 3 the marginal pdfs of the two composite Gumbel–Pareto distributions considered in Figures 1 and 2, and we note their continuity and exponential type shapes for small values of x.

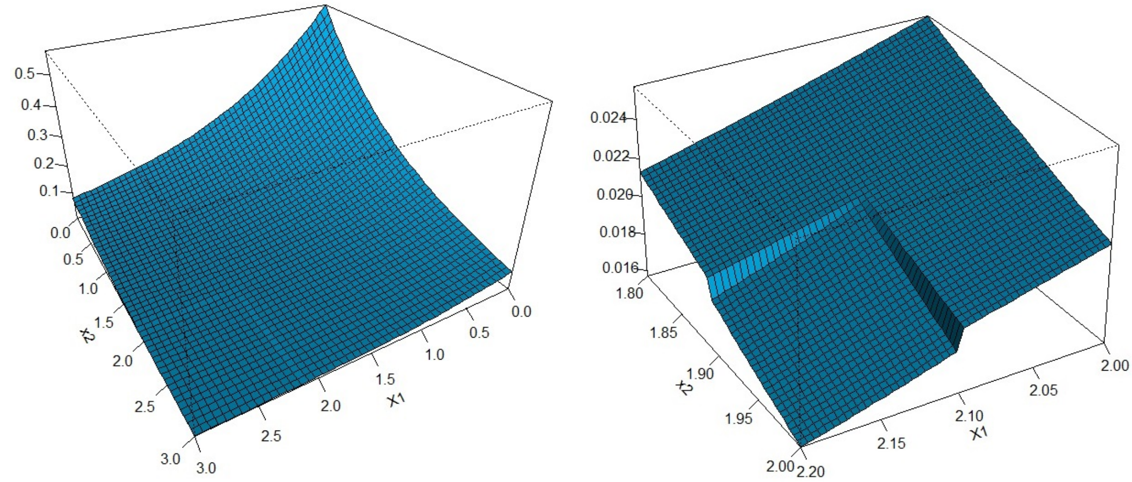

Figure 1. Left: composite Gumbel–Pareto pdf with continuous marginals and continuity at (θ_1, θ_2); **Right**: zoom of the same pdf (parameters: $\lambda_1 = 0.81, \lambda_2 = 0.9, \beta = 0.2, a = 1.0258, \theta_1 = 2.1$, $\theta_2 = 1.89$).

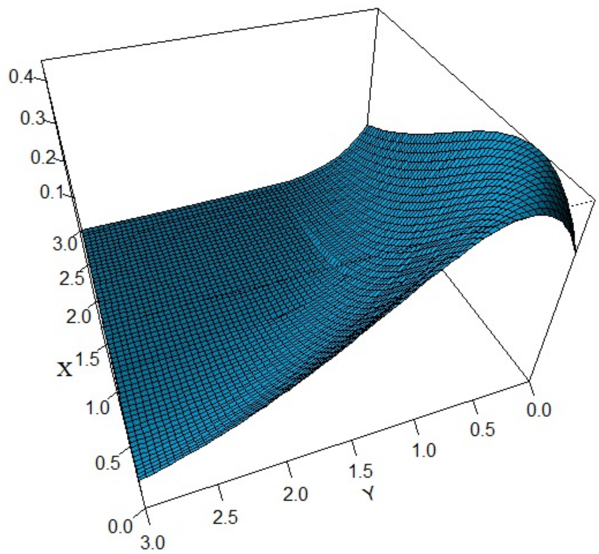

Figure 2. Composite Gumbel–Pareto pdf with continuous marginals and continuity at (θ_1, θ_2) (parameters: $\lambda_1 = 1, \lambda_2 = 1.2, \beta = 0.7, a = 0.7515, \theta_1 = 1.2, \theta_2 = 1$).

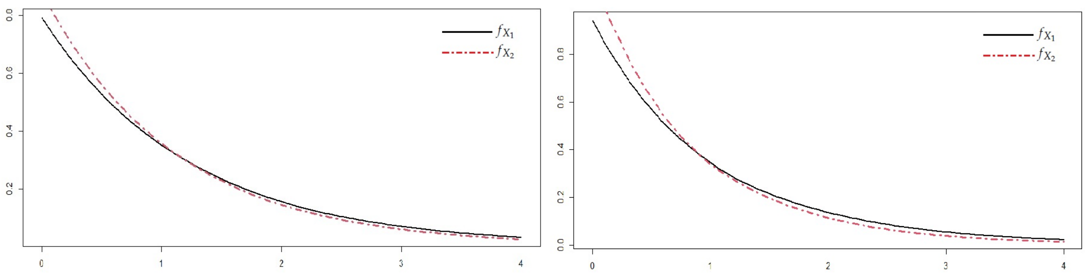

Figure 3. Marginal pdfs of composite Gumbel–Pareto distribution: **left**—with parameters from Figure 1; **right**—with parameters from Figure 2 (f_{X_1} solid line, f_{X_2} dashed line).

4.1. Some Properties

The marginal distributions of $\mathbf{X}$ are both of univariate composite type, having a standard exponential pdf up to the threshold.

Proposition 1. *(i) For the composite Gumbel–Pareto distribution, the marginal pdfs of X_1 and X_2 are given by*

$$f_{X_1}(x_1) = \begin{cases} \frac{r}{P_D}\lambda_1 e^{-\lambda_1 x_1}, 0 < x_1 \le \theta_1 \\ \frac{r}{P_D}\lambda_1 e^{-\lambda_1 x_1}\left[1-(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2(1+\beta\lambda_1 x_1)}\right] + (1-r)a\frac{\theta_1^a}{x_1^{a+1}}, x_1 > \theta_1 \end{cases},$$

$$f_{X_2}(x_2) = \begin{cases} \frac{r}{P_D}\lambda_2 e^{-\lambda_2 x_2}, 0 < x_2 \le \theta_2 \\ \frac{r}{P_D}\lambda_2 e^{-\lambda_2 x_2}\left[1-(1+\beta\lambda_1\theta_1)e^{-\lambda_1\theta_1(1+\beta\lambda_2 x_2)}\right] + (1-r)a\frac{\theta_2^a}{x_2^{a+1}}, x_2 > \theta_2 \end{cases}.$$

(ii) Further, the cdfs of X_1 and X_2 are

$$F_{X_1}(x_1) = \begin{cases} \frac{r}{P_D}\left(1-e^{-\lambda_1 x_1}\right), 0 < x_1 \le \theta_1 \\ 1 + \frac{r}{P_D}e^{-\lambda_1 x_1}\left(e^{-\lambda_2\theta_2(1+\beta\lambda_1 x_1)} - 1\right) - (1-r)\left(\frac{\theta_1}{x_1}\right)^a, x_1 > \theta_1 \end{cases},$$

$$F_{X_2}(x_2) = \begin{cases} \frac{r}{P_D}\left(1-e^{-\lambda_2 x_2}\right), 0 < x_2 \le \theta_2 \\ 1 + \frac{r}{P_D}e^{-\lambda_2 x_2}\left(e^{-\lambda_1\theta_1(1+\beta\lambda_2 x_2)} - 1\right) - (1-r)\left(\frac{\theta_2}{x_2}\right)^a, x_2 > \theta_2 \end{cases}.$$

We can impose marginal continuity conditions and combine them with the continuity condition at (θ_1, θ_2). The following restrictions result.

Proposition 2. *Let $\mathbf{X}$ follow the bivariate composite Gumbel–Pareto distribution. Then:*
(i) By imposing the continuity condition to the marginal X_1, we obtain

$$r_1 = \left(1 + \frac{\lambda_1\theta_1}{a}\frac{1+\beta\lambda_2\theta_2}{e^{\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2} - 1}\right)^{-1}.$$

(ii) By imposing the continuity condition to the marginal X_2, we obtain

$$r_2 = \left(1 + \frac{\lambda_2\theta_2}{a}\frac{1+\beta\lambda_1\theta_1}{e^{\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2} - 1}\right)^{-1}.$$

(iii) By simultaneously imposing continuity conditions to the marginals X_1 and X_2, we obtain

$$\lambda_1\theta_1 = \lambda_2\theta_2.$$

If, moreover, we also impose the continuity condition at (θ_1, θ_2), the following restriction must be fulfilled

$$a = \lambda_1\theta_1\left(1 + \beta\lambda_1\theta_1 - \frac{\beta}{1+\beta\lambda_1\theta_1}\right) - 1.$$

Proposition 3. *(i) The expected values of the marginals are given for $a > 1$ by*

$$\begin{aligned}
\mathbb{E}X_1 &= \frac{r}{\lambda_1 P_D}\left[1 - e^{-(\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2)}\left(\frac{1}{1+\beta\lambda_2\theta_2} + \lambda_1\theta_1\right)\right] + (1-r)\frac{a\theta_1}{a-1}, \\
\mathbb{E}X_2 &= \frac{r}{\lambda_2 P_D}\left[1 - e^{-(\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2)}\left(\frac{1}{1+\beta\lambda_1\theta_1} + \lambda_2\theta_2\right)\right] + (1-r)\frac{a\theta_2}{a-1}.
\end{aligned}$$

(ii) The second-order moments of the marginals are given for $a > 2$ by

$$\begin{aligned}
\mathbb{E}\left[X_1^2\right] &= \frac{r}{\lambda_1^2 P_D}\left[2 - e^{-(\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2)}\frac{1 + (\lambda_1\theta_1(1+\beta\lambda_2\theta_2)+1)^2}{(1+\beta\lambda_2\theta_2)^2}\right] + (1-r)\frac{a\theta_1^2}{a-2}, \\
\mathbb{E}\left[X_2^2\right] &= \frac{r}{\lambda_2^2 P_D}\left[2 - e^{-(\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2)}\frac{1 + (\lambda_2\theta_2(1+\beta\lambda_1\theta_1)+1)^2}{(1+\beta\lambda_1\theta_1)^2}\right] + (1-r)\frac{a\theta_2^2}{a-2}.
\end{aligned}$$

Proposition 4. *The expected value of the product X_1X_2 is*

$$\begin{aligned}
\mathbb{E}[X_1X_2] &= \frac{r}{P_D\beta\lambda_1\lambda_2}\left[\left(E_1\left(\frac{1}{\beta}\right) - E_1\left(\frac{(1+\beta\lambda_1\theta_1)(1+\beta\lambda_2\theta_2)}{\beta}\right)\right)e^{\frac{1}{\beta}}\right. \\
&\quad \left. -\left(2 - \frac{1}{1+\beta\lambda_1\theta_1} - \frac{1}{1+\beta\lambda_2\theta_2} + \beta\lambda_1\lambda_2\theta_1\theta_2\right)e^{-(\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2)}\right] \\
&\quad +(1-r)\theta_1\theta_2\frac{a^2-a-1}{(a-1)(a-2)}.
\end{aligned}$$

In view of the random generation procedure, we also need the following result on the conditional distribution of a marginal.

Proposition 5. *The conditional cdf of the marginal X_2 given $X_1 = x_1$ is*

$$F_{X_2|X_1=x_1}(x_2) = \begin{cases} 1 - e^{-\lambda_2 x_2(1+\beta\lambda_1 x_1)}(1+\beta\lambda_2 x_2), x_1 \le \theta_1, x_2 > 0 \\ \frac{r}{P_D}\lambda_1 e^{-\lambda_1 x_1}\frac{1-e^{-\lambda_2 x_2(1+\beta\lambda_1 x_1)}(1+\beta\lambda_2 x_2)}{\frac{r}{P_D}\lambda_1 e^{-\lambda_1 x_1}\left[1-(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2(1+\beta\lambda_1 x_1)}\right]+(1-r)a\frac{\theta_1^a}{x_1^{a+1}}}, x_1 > \theta_1, x_2 \le \theta_2 \\ 1-(1-r)\frac{a\theta_1^a\theta_2^{a+1}(\theta_1 x_2+\theta_2 x_1-\theta_1\theta_2)^{-(a+1)}}{\frac{r}{P_D}\lambda_1 e^{-\lambda_1 x_1}\left[1-(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2(1+\beta\lambda_1 x_1)}\right]+(1-r)a\frac{\theta_1^a}{x_1^{a+1}}}, x_1 > \theta_1, x_2 > \theta_2 \end{cases}.$$

4.2. Simulation

We propose two methods for generating random values from the bivariate composite Gumbel–Pareto distribution. The first one is the inversion method, while the second one is based on the representation in expression (7).

Method I: In the bivariate case, the inversion method consists of two steps:

1. Generate a value x_1 from the marginal distribution of X_1 by inverting its cdf given in Proposition 1;
2. Generate a value x_2 from the conditional distribution of X_2 given $X_1 = x_1$ by inverting the conditional cdf given in Proposition 5. Thus, the resulting pair (x_1, x_2) is simulated from (10).

Method II: Starting from the two-component mixture representation (7) with mixing weights r and $1-r$, we propose the following algorithm:

1. Generate a value b from the Bernoulli distribution with parameter r;
2. If $b = 1$, then generate the pair (x_1, x_2) from the Gumbel distribution truncated on D;
3. If $b = 0$, then generate the pair (x_1, x_2) from the bivariate Pareto distribution (4).

Now the problem is to generate values from the two bivariate distributions: Gumbel and Pareto. Bivariate Pareto values can be generated without difficulty by the inversion method as described in Method I. Concerning the Gumbel distribution truncated on D, the following cdfs (obtained similarly to the ones in Propositions 1 and 5) can be used for inversion:
The cdf of the truncated Gumbel marginal, Y_1^D:

$$F_{Y_1^D}(x_1) = \begin{cases} \frac{1}{P_D}\left(1 - e^{-\lambda_1 x_1}\right), 0 < x_1 \le \theta_1 \\ 1 + \frac{1}{P_D} e^{-\lambda_1 x_1}\left(e^{-\lambda_2\theta_2(1+\beta\lambda_1 x_1)} - 1\right), x_1 > \theta_1 \end{cases};$$

The conditional cdf of the marginal Y_2^D given $Y_1^D = x_1$ of the truncated Gumbel distribution:

$$F_{Y_2^D|Y_1^D=x_1}(x_2) = \begin{cases} 1 - (1+\beta\lambda_2 x_2)e^{-\lambda_2 x_2(1+\beta\lambda_1 x_1)}, x_1 \le \theta_1 \\ \frac{1-(1+\beta\lambda_2 x_2)e^{-\lambda_2 x_2(1+\beta\lambda_1 x_1)}}{1-(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2(1+\beta\lambda_1 x_1)}}, x_1 > \theta_1 \end{cases}, x_2 > 0.$$

4.3. Parameter Estimation

For a univariate composite distribution, estimating the parameters is already a difficult problem because the threshold where the distribution changes shape is itself a parameter. Therefore, the usual approach in the univariate case consists of sorting the data, assuming that the threshold lies between each two consecutive data points, and finding the corresponding MLE solution; then, the best MLE solution is selected from among the available ones. Alternatively, a set of possible thresholds can be defined, and for each such value, the resulting likelihood is maximized; see also the review [11] for threshold estimation approaches.

In the bivariate case, the estimation problem becomes even more difficult because there are two unknown thresholds θ_1, θ_2 to estimate. Let $\mathbf{x} = (x_{1i}, x_{2i})_{i=1}^n$ be a bivariate data sample of size n, let λ_1, λ_2, β, a, r denote the rest of the parameters of the bivariate density defined in (10) (note that r might be obtained from a continuity condition such as (11) or the ones in Proposition 2, if imposed), and let L denote the likelihood function

$$L(\mathbf{x}; \lambda_1, \lambda_2, \beta, a, \theta_1, \theta_2, r) = \prod_{\{(x_{1i}, x_{2i}) \in D\}} r f_1(x_1, x_2) \prod_{\{(x_{1i}, x_{2i}) | x_{1i} > \theta_1, x_{2i} > \theta_2\}} (1-r) f_2(x_1, x_2). \quad (12)$$

The log-likelihood function defined from (12) is the weighted sum of the two partial log-likelihood functions associated with the two distributions of the composite model: the Gumbel and the Pareto. Since the MLE exists for both distributions (see [10] for the bivariate Pareto distribution), then for a known r, we can easily find the MLE of our composite model. The aim of the proposed MLE procedures is to find the best value of r.

In the following, we propose two alternative methods to estimate the parameters.

Method 1: An approach similar to the one described in the univariate case would be to sort the marginal data, obtaining $\left(x_{1(i)}\right)_{i=1}^n$ and $\left(x_{2(i)}\right)_{i=1}^n$, assume that each threshold lies, correspondingly, between each two consecutive marginal data points, find the MLEs, and choose the best one. However, this procedure is very time-consuming in the bivariate case, so we propose to combine it with marginal estimation in a two-part method as follows:

I. Perform marginal estimation for both marginals; since the marginals are univariate composite distributions, the approach described above for the univariate case can be used. This would give starting values for the marginal parameters and the approximate location of the marginally estimated thresholds $\tilde{\theta}_1, \tilde{\theta}_2$.

II. Let $\left(x_{1(i)}\right)_{i=1}^{n}$ and $\left(x_{2(i)}\right)_{i=1}^{n}$ denote the (increasing) sorted marginal data and assume that the marginally estimated thresholds $\tilde{\theta}_j \in m_{j(k_j)}$, $j = 1,2$, where $m_{j(k_j)} = \left(x_{j(k_j)}, x_{j(k_j+1)}\right)$. Now consider $\left(m_{j(k_j-h)}\right)_{h=1}^{l}$ the l intervals preceding and $\left(m_{j(k_j+h)}\right)_{h=1}^{l}$ the l intervals following the interval $m_{j(k_j)}$ that covers $\tilde{\theta}_j$, $j = 1,2$, as long as they exist; for each combination of such intervals, perform full MLE and keep the best solution. The resulting algorithm is:

Step 1. For $m_{j(k_j-l)}$ to $m_{j(k_j+l)}$, $j = 1,2$,
evaluate λ_1, λ_2, β, a, θ_1, θ_2, r as solutions of the optimization problem:

$$\max \log L(\mathbf{x}; \lambda_1, \lambda_2, \beta, a, \theta_1, \theta_2, r),$$

under the constraints θ_1 and θ_2 in the corresponding intervals, and continuity conditions, if imposed.

Step 2. Among the solutions obtained from Step 1, choose the one that maximizes the log-likelihood function.

Note that in this way, for reasonable choices of $m_{1(k_1)}, m_{2(k_2)}$ and l, the computing time is significantly reduced.

Method 2: The second method is a more analytical procedure for a specific sample; it takes into account that the parameter β of the bivariate Gumbel–Pareto density (10) is restricted to the $[0,1]$ interval. This allows us to define a grid for it and to optimize the rest of the parameters for each value in this grid. The following procedure is designed, assuming the continuity conditions given in (*i*–*iii*) of Proposition 2 and the conditional likelihood defined by:

$$L^c(\mathbf{x}; \lambda_1, \theta_1, \theta_2 | \beta) = \prod_{\{(x_{1i}, x_{2i}) \in D\}} r f_1(x_1, x_2) \prod_{\{(x_{1i}, x_{2i}) | x_{1i} > \theta_1, x_{2i} > \theta_2\}} (1-r) f_2(x_1, x_2),$$

with the continuity conditions (constraints)

$$\begin{aligned} \lambda_2 &= \frac{\lambda_1 \theta_1}{\theta_2}, \\ a &= \lambda_1 \theta_1 \left(1 + \beta \lambda_1 \theta_1 - \frac{\beta}{1 + \beta \lambda_1 \theta_1}\right) - 1, \\ \beta &\in [0,1). \end{aligned}$$

The conditional likelihood $L^c(\mathbf{x}; \lambda_2, \theta_1, \theta_2 | \beta)$ is defined similarly. The procedure for maximizing $\log L^c(\mathbf{x}; \lambda_1, \theta_1, \theta_2 | \beta)$ is described below:

Step 1. Obtain initial values for the parameters θ_1, θ_2, and λ_1 as follows:

- The initially estimated thresholds are $\tilde{\theta}_1 = x_{1([np_1])}$ and $\tilde{\theta}_2 = x_{2([np_2])}$, where p_j, $j = 1,2$, are two given large proportions, and $[\cdot]$ denotes the integer part. An initial value for each proportion can be deduced from the Hill plot or by doing MLE of the univariate Pareto for the tail.
- The initially estimated value of the exponential parameter $\tilde{\lambda}_1$ is obtained by MLE of the univariate truncated exponential distribution with density function:

$$f(x_1 | x_1 \le \tilde{\theta}_1) = \frac{f_{X_1}(x_1)}{F_{X_1}(\tilde{\theta}_1)} = \frac{\lambda_1 e^{-\lambda_1 x_1}}{1 - e^{-\lambda_1 \tilde{\theta}_1}}.$$

Step 2. Define a grid for $\beta \in [0,1)$, i.e., $(\beta_g)_{g=1}^{G}$. For each β_g, the estimated parameters $\hat{\theta}_1^g$, $\hat{\theta}_2^g$, and $\hat{\lambda}_1^g$ are obtained by maximizing the conditional log-likelihood function $\log L_g^c(\mathbf{x}; \lambda_1, \theta_1, \theta_2 | \beta_g)$. The `optim()` function of `R` software with the "Nelder–Mead"

method can be used; this works reasonably well for non-differentiable functions. The parameters $\hat{\lambda}_2^g$ and $\hat{a}^g$ are estimated using the continuity conditions.
Step 3. Let $(\log \tilde{L}_g^c)_{g=1}^G$ be the optimal values of the log-likelihood obtained at Step 2, and let $\left(\hat{\lambda}_1^g, \hat{\theta}_1^g, \hat{\theta}_2^g, \beta_g\right)$ be the corresponding parameters. The final estimated parameters are:

$$\left(\hat{\lambda}_1, \hat{\theta}_1, \hat{\theta}_2, \hat{\beta}\right) = \arg \max_{g=1,\ldots,G} \log \tilde{L}_g^c$$

with

$$\hat{\lambda}_2 = \frac{\hat{\lambda}_1 \hat{\theta}_1}{\hat{\theta}_2},$$

$$\hat{a} = \hat{\lambda}_1 \hat{\theta}_1 \left(1 + \hat{\beta}\hat{\lambda}_1\hat{\theta}_1 - \frac{\hat{\beta}}{1 + \hat{\beta}\hat{\lambda}_1\hat{\theta}_1}\right) - 1.$$

5. Numerical Illustration

In this section, we present two numerical illustrations: the first one is on simulated data, and the second one is on a real data set.

5.1. Numerical Illustration Using Simulated Data

In this section, we used simulated data to check the performance of the first estimation procedure (Method 1) proposed in Section 4.3. The true values of the parameters were selected such that they satisfied all the continuity conditions given in Proposition 2: Gumbel: $\lambda_1 = 1, \lambda_2 = 1.2, \beta = 0.7$; Pareto: $a = 0.7515, \theta_1 = 1.2, \theta_2 = 1$, while $r = 0.9086$, $P_D = 0.9669$. Note that due to the heavy-tailedness of the Pareto distribution ($a < 1$), there is no expected value for this particular distribution (its pdf is plotted in Figure 2).

With the aim of studying the properties of Method 1, using the two simulation methods described in Section 4.2, we generated 100 samples of size $n = 200$ and $n = 1000$, respectively, for the two methods. For each such sample, in the first step, we performed marginal estimation by imposing the continuity condition for each marginal (which restricts the parameters r, as stated in Proposition 2). As a consequence, β and a are estimated twice (for each marginal), and because of the differences in these estimations, we cannot rely only on marginal estimation. However, marginal estimation provides starting values for performing full MLE, and even better, gives an idea of where to look for the thresholds. More precisely, we restricted the search to about 40 intervals for each θ_j, i.e., we took $l = 20$. Thus, the computing time was significantly reduced compared to the threshold search through all data.

Finally, we estimated the Mean Square Error $MSE = \frac{1}{100}\sum_{i=1}^{100}\left(\theta - \hat{\theta}\right)^2$ and the Mean Absolute Error $MAE = \frac{1}{100}\sum_{i=1}^{100}\left|\theta - \hat{\theta}\right|$, where θ and $\hat{\theta}$ represent the true and estimated parameters, respectively.

With the estimated parameters obtained from the 100 replicas generated with each simulation method, we obtained the MSE and the MAE that are shown in Table 1. The results indicate that both error criteria decrease when the sample size increases. Some differences between the two simulation methods can be observed (e.g., the MSE of β is larger for simulation Method II than for simulation Method I, while the MSE of a is smaller for simulation Method II than for simulation Method I), but we believe that these differences are due to the randomness of the results, where some samples fall more in the Pareto part or in the exponential part; further simulation investigation is worthwhile, assuming that the estimation method can be modified to reduce the computing time.

Concerning Method 2, as already noticed, it is a more analytical procedure for a specific sample, and therefore, it cannot be standardized and we cannot perform several iterations to calculate MSE and MAE.

All the computations were preformed in `R` software using an optimization function with constraints to implement the continuity restrictions. The code is available upon request from the authors.

Table 1. Simulation results with 100 replicas for MSE and MAE with sample sizes $n = 200$ and $n = 1000$.

	Simulation Method I											
	λ_1		λ_2		β		a		θ_1		θ_2	
n	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
200	0.0037	0.0538	0.0060	0.0717	0.0124	0.1077	0.3855	0.6152	0.0530	0.2226	0.0144	0.1069
1000	0.0034	0.0511	0.0048	0.0610	0.0118	0.1044	0.0275	0.1625	0.0450	0.2109	0.0049	0.0619
	Simulation Method II											
	λ_1		λ_2		β		a		θ_1		θ_2	
n	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
200	0.0054	0.0721	0.0041	0.0447	0.1530	0.3900	0.0640	0.2493	0.0694	0.2619	0.0985	0.3010
1000	0.0048	0.0610	0.0002	0.0131	0.1234	0.3367	0.0174	0.1267	0.0638	0.2564	0.0567	0.2229

5.2. Numerical Illustration with Real Data

In this section, we fit our proposed bivariate Gumbel–Pareto distribution to a random sample of $n = 518$ motor insurance claims that include bodily injury. For these claims, we separately know the cost of property damage including third-part liability (variable X_1) and the cost of exceptional medical expenses not covered by public social security (variable X_2). The data were provided by a major insurer in Spain in the year 2002 and correspond to claims that occurred in the year 2000. These data were studied in previous works (see [5,6,12]).

In Table 2, we display the descriptive statistics of the original data divided by 1000; this change of scale is convenient, and it facilitates the MLE of the parameters. These descriptive statistics show that both variables have a strong right skewness. Furthermore, the left plot in Figure 4 shows the scatterplot of both cost variables in the original scale divided by 1000, where the existence of extreme values in both variables can be noticed. When we have right-skewed variables with extreme values, the MLE of a simple distribution as, e.g., the exponential, the Weibull, or the log-normal, tends to underestimate the probability on the right tail. Figure 5 displays the univariate exponential pdf fitted by MLE to each marginal variable; with these densities, we also plotted the observed costs: on top the costs of property damage, including third-part liability, and on bottom the costs of exceptional medical expenses not covered by public social security. For better visibility, the domains of the cost variables were divided in two parts, resulting in two plots for each marginal. Figure 5 shows how the density reaches zero in the part of the domain where there are still sample observations; so clearly, this model assumes a zero probability where it should not. Similar results are obtained using univariate Weibull and log-normal densities.

Therefore, the composite model with a Pareto right tail is a good way to improve the MLE fit for both univariate and bivariate data. Moreover, graphical analysis (e.g., the Hill plot) indicates that both variables have a Pareto tail with a shape parameter very close to 1, i.e., we have heavy-tailed marginal distributions. Thus, we can conclude that their distributions have only the first-order moment finite, or they do not have finite moments at all. In the left scatterplot of Figure 4, we can note that the sample information on extreme values is scarce; this is a difficulty in samples from heavy-tailed or Pareto distributions.

Table 2. Descriptive statistics of property damage and third-party liability costs (X_1) and exceptional medical expenses (X_2).

	Mean	**STD**	**Min**	**Q25**	**Median**	**Q75**	**Max**	**Kurtosis**	**Skewness**
X1	1.83	6.87	0.01	0.26	0.68	1.39	137.94	15.70	301.30
X2	0.28	0.86	0.00	0.02	0.09	0.20	11.86	8.06	85.35

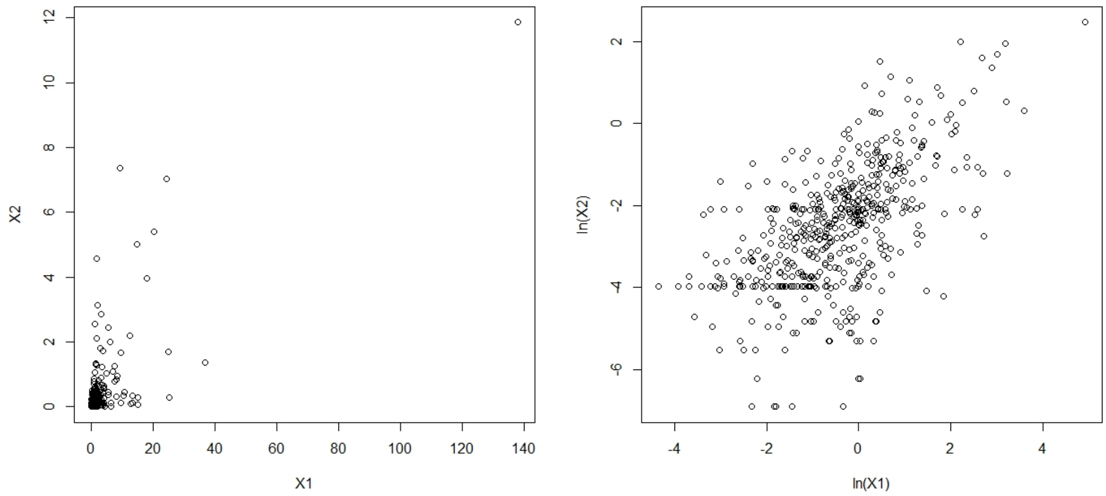

Figure 4. Scatterplots of X_1 vs. X_2 in original (**left**) and natural (**right**) logarithm scales.

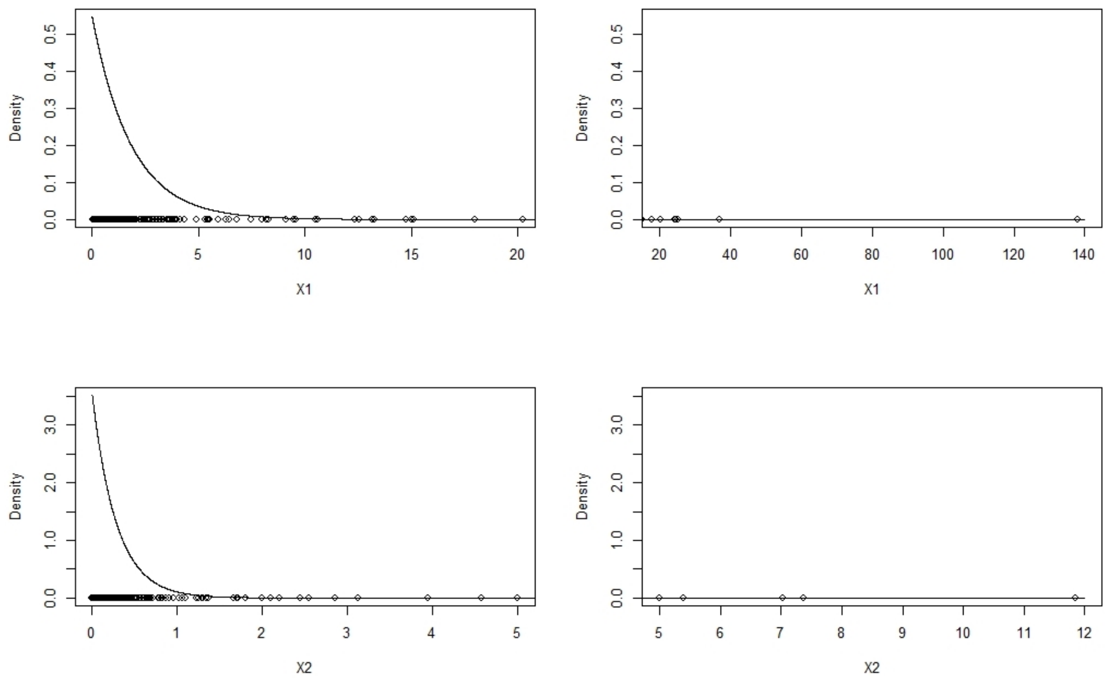

Figure 5. Exponential pdf fitted by MLE and sample data shown as points on the horizontal axis for both marginals.

To asses the joint behavior of X_1 and X_2, we calculated the Pearson linear correlation and the Kendall and the Spearman rank correlation coefficients, displayed in Table 3. These results show a strong dependence between the two cost variables. However, as can be seen from Figure 4, which presents the data scatterplot in both original and natural logarithm scales, the dependence is not linear. As shown in [12], these data exhibit extreme value dependence, i.e., the higher the costs, the stronger the dependency. This behavior can also be observed in Figure 4. Furthermore, [10] shows that when the bivariate Pareto parameter a is $a \leq 2$, as is the case with our cost data, the theoretical variance and covariance do not exist or cannot be calculated. Therefore, the Pearson linear correlation cannot be interpreted.

Further, from the right plot in Figure 4, it can be observed that for small values of both variables, the shape of the point cloud is spherical, i.e., the dependence is almost zero; however, for larger values, the shape indicates positive dependence between both variables. Clearly, this denotes a change of the joint distribution between the smaller and the larger costs.

Table 3. Sample linear and rank correlation coefficients.

	Pearson	**Kendall**	**Spearman**
Correlation	0.7288	0.4252	0.5903

In Table 4, we present the MLE parameters for Gumbel's bivariate exponential distribution described in Section 2.2.1 and for the Gumbel–Pareto distribution from Section 4. The estimated parameters of the latter were obtained with Method 2 described in Section 4.3, imposing all continuity conditions (Method 1 yielded similar results). The initial values of the thresholds were taken from the Hill plots, and in this case, $p_1 = p_2 = 0.102$, resulting in $[np] = [518 \times 0.103] = 52$, i.e., $\tilde{\theta}_1 = 3.1$ and $\tilde{\theta}_2 = 0.5$; also, $\tilde{\lambda}_1 = 1.4175$. Comparing the AICs, BICs, and CAICs given in Table 4 indicates that the bivariate Gumbel–Pareto clearly outperforms Gumbel's bivariate exponential distribution. Moreover, from MLE, the dependence parameter of Gumbel's bivariate exponential distribution, β, is zero, and it is close to zero for the Gumbel–Pareto distribution, which is coherent with the scatterplot in Figure 4.

In Figure 6, we also plotted a partial histogram of the data alongside the corresponding Gumbel–Pareto pdf with the estimated parameters, while in Figure 7, we plotted the marginal histograms with the fitted pdfs.

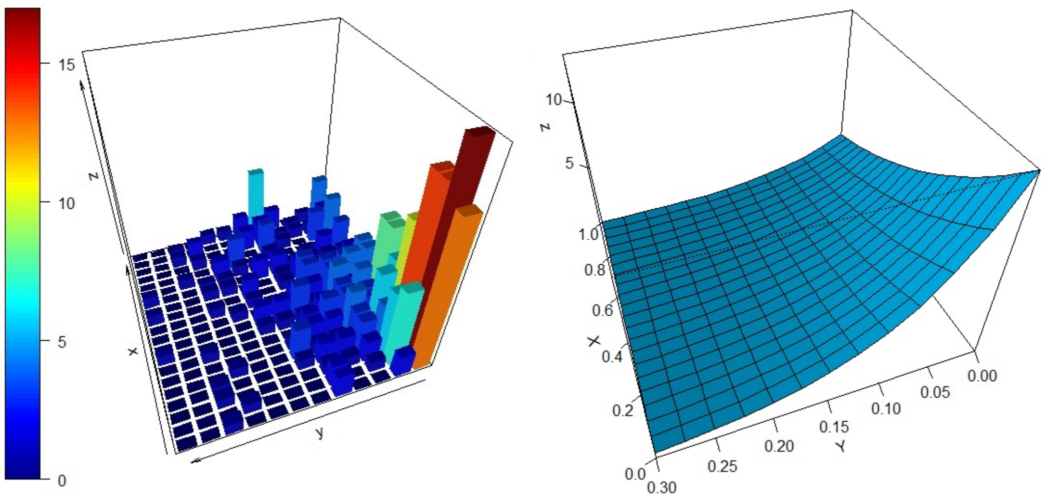

Figure 6. Histogram of real data (**left**) and Gumbel–Pareto pdf with the estimated parameters (**right**).

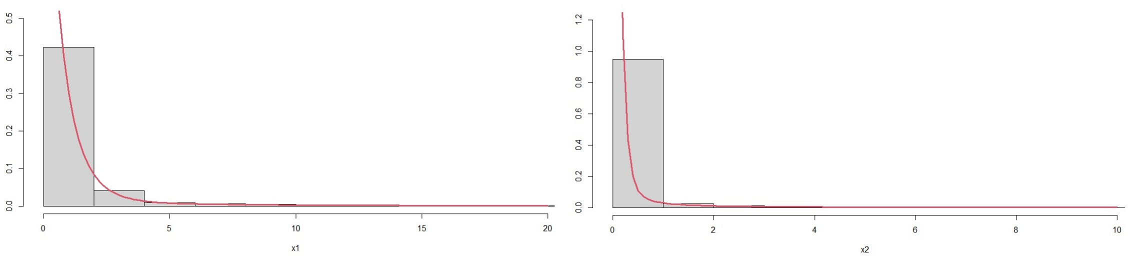

Figure 7. Histogram of real data marginals with fitted pdfs: **left**, X_1; **right**, X_2.

Table 4. MLE of bivariate distributions with standard errors in parentheses.

	Gumbel	**Gumbel–Pareto**
$\hat{\lambda}_1$	0.5472 (0.0240)	1.4184 (0.0328)
$\hat{\lambda}_2$	3.5221 (0.1548)	11.1996
$\hat{\theta}_1$	-	0.9870 (0.0040)
$\hat{\theta}_2$	-	0.1250 (0.0003)
$\hat{\beta}$	0.0000	0.0455 (0.0465)
a	-	0.4292
r	-	0.8303
$\log L$	−696.1630	−272.5549
AIC	1398.3261	557.1097
BIC	1411.0759	582.6096
CAIC	1414.0759	588.6096

Finally, as a risk management application, we estimated the total risk of loss for the aggregate cost random variable $S = X_1 + X_2$ using Monte Carlo simulation, and based on it, we calculated the Value-at-Risk (VaR) measure. VaR is equivalent to an extreme quantile of the distribution, i.e., $VaR_\alpha(S) = \inf\{s \in \mathbb{R} | \Pr(S \leq s) \geq \alpha\}$, where α is close to 1. In Table 5, we present the VaR results with $\alpha = 0.95, 0.99, 0.995$ for: the empirical distribution of the original data, the distribution of S simulated from Gumbel's bivariate exponential distribution, and the distribution of S simulated from the Gumbel–Pareto distribution. Furthermore, we added the VaR obtained for the bivariate log-normal distribution fitted to the data; note that this distribution underestimates the risk in a way similar to that of Gumbel's bivariate exponential.

Table 5. Value-at-Risk for the empirical distribution and alternative distributions, obtained using Monte Carlo simulation.

	95%	**99%**	**99.50%**
Empirical	7.926	25.409	31.216
Gumbel	6.312	9.700	11.178
Gumbel–Pareto	6.361	114.067	410.897
Log-normal	6.529	15.122	20.787

When data follow a heavy-tailed distribution, the empirical VaR depends on the maximum data observed, and it is not an efficient estimator. The Gumbel–Pareto distribution provides an estimation that extrapolates beyond the observed maximum cost and takes into account the long and heavy bivariate tail with dependent marginal distributions.

6. Conclusions

To model bivariate dependent data that exhibit many small/medium values but also some very large values (i.e., extreme values), in this paper, we proposed a bivariate two-component spliced distribution. This distribution assumes a bivariate Pareto distribution on the domain consisting of values larger than some thresholds, and a bivariate Gumbel distribution on the complementary domain. We discussed some properties of the new distribution and focused on parameter estimation, proposing two alternative procedures. Because performing full MLE for this distribution may become time-prohibitive for larger data sets, as further work, we plan to investigate alternative methods that could reduce the computing time. Additionally, starting from the mixture formula (7), we plan to address the problem of parameter identifiability (see, e.g., [13] or [14]). Goodness-of-fit tests are envisaged for a future study.

Moreover, we also plan to study other such distributions by replacing the bivariate Gumbel with alternative distributions.

Author Contributions: Conceptualization, C.B. and R.V.; methodology, C.B. and R.V.; software, A.B., C.B. and R.V.; formal analysis, A.B., C.B. and R.V.; writing—original draft preparation, A.B. and R.V.; writing—review and editing, C.B. and R.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available from the authors.

Acknowledgments: The authors are very grateful to the three referees for their valuable comments that helped to significantly improve the paper. Catalina Bolancé acknowledges the Spanish Ministry of Science, Innovation and Universities, grant PID2019-105986GB-C21.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

Appendix A. Proofs

***Proof of Lemma* 1.** Using integration by parts, it is easy to prove (i)–(iv); (v) results by changing variable $t = ky$, while (vi) is obtained by parts and by using (v). □

***Proof of Lemma* 3.** Without loss of generality, we prove the formula of L_2; proof of L_1 results in a similar way.

$$\begin{aligned}
L_2(x_2;\theta_1) &= \lambda_2 e^{-\lambda_2 x_2}\int_0^{\theta_1}\lambda_1 e^{-x_1(\lambda_1+\beta\lambda_1\lambda_2x_2)}[\beta\lambda_1x_1(1+\beta\lambda_2x_2)+1+\beta\lambda_2x_2-\beta]dx_1\\
&= \lambda_2 e^{-\lambda_2 x_2}\Big[\beta\lambda_1^2(1+\beta\lambda_2x_2)\int_0^{\theta_1}x_1e^{-x_1(\lambda_1+\beta\lambda_1\lambda_2x_2)}dx_1\\
&\quad +\lambda_1(1+\beta\lambda_2x_2-\beta)\int_0^{\theta_1}e^{-x_1(\lambda_1+\beta\lambda_1\lambda_2x_2)}dx_1\Big]\\
&= \lambda_2 e^{-\lambda_2 x_2}\Big[\beta\lambda_1^2(1+\beta\lambda_2x_2)\Gamma(2,0,\theta_1;\lambda_1+\beta\lambda_1\lambda_2x_2)\\
&\quad +\lambda_1(1+\beta\lambda_2x_2-\beta)\Gamma(1,0,\theta_1;\lambda_1+\beta\lambda_1\lambda_2x_2)\Big].
\end{aligned}$$

Using formulas (ii) and (iii.1) from Lemma 1, we obtain, with some calculation,

$$\begin{aligned}
L_2(x_2;\theta_1) &= \lambda_2 e^{-\lambda_2x_2}\left(\frac{\beta\lambda_1^2(1+\beta\lambda_2x_2)}{\lambda_1^2(1+\beta\lambda_2x_2)^2}\Big[1-(1+\theta_1\lambda_1(1+\beta\lambda_2x_2))e^{-\theta_1\lambda_1(1+\beta\lambda_2x_2)}\Big]\right.\\
&\quad \left.+\lambda_1(1+\beta\lambda_2x_2-\beta)\frac{1-e^{-\theta_1\lambda_1(1+\beta\lambda_2x_2)}}{\lambda_1(1+\beta\lambda_2x_2)}\right)\\
&= \frac{\lambda_2e^{-\lambda_2x_2}}{1+\beta\lambda_2x_2}\Big[\beta-e^{-\theta_1\lambda_1(1+\beta\lambda_2x_2)}(\beta+\beta\theta_1\lambda_1(1+\beta\lambda_2x_2)+(1+\beta\lambda_2x_2-\beta))\\
&\quad +1+\beta\lambda_2x_2-\beta\Big]\\
&= \lambda_2e^{-\lambda_2x_2}\Big[-e^{-\theta_1\lambda_1(1+\beta\lambda_2x_2)}(1+\beta\theta_1\lambda_1)+1\Big].
\end{aligned}$$

□

Proof of Lemma 4. We write

$$I(\theta_1,\theta_2) = \int_{\theta_1}^{\infty} \lambda_1 x_1 e^{-\lambda_1 x_1} J(\theta_2) dx_1,$$

where

$$\begin{aligned} J(\theta_2) &= \int_{\theta_2}^{\infty} \lambda_2 x_2 e^{-\lambda_2 x_2(1+\beta\lambda_1 x_1)} [\beta\lambda_2 x_2(1+\beta\lambda_1 x_1) + (1+\beta\lambda_1 x_1) - \beta] dx_2 \\ &= \beta\lambda_2^2(1+\beta\lambda_1 x_1) \int_{\theta_2}^{\infty} x_2^2 e^{-\lambda_2 x_2(1+\beta\lambda_1 x_1)} dx_2 \\ &\quad +\lambda_2(1+\beta\lambda_1 x_1 - \beta) \int_{\theta_2}^{\infty} x_2 e^{-\lambda_2 x_2(1+\beta\lambda_1 x_1)} dx_2 \\ &= \beta\lambda_2^2(1+\beta\lambda_1 x_1)\Gamma(3,\theta_2,\infty;\lambda_2(1+\beta\lambda_1 x_1)) \\ &\quad +\lambda_2(1+\beta\lambda_1 x_1 - \beta)\Gamma(2,\theta_2,\infty;\lambda_2(1+\beta\lambda_1 x_1)), \end{aligned}$$

and using the corresponding formulas (*iv*.2) and (*iii*.2) from Lemma 1, we obtain

$$\begin{aligned} J(\theta_2) &= \beta\lambda_2^2(1+\beta\lambda_1 x_1) \frac{e^{-\theta_2\lambda_2(1+\beta\lambda_1 x_1)}}{(\lambda_2(1+\beta\lambda_1 x_1))^3} \Big[(\theta_2\lambda_2(1+\beta\lambda_1 x_1)+1)^2 + 1 \Big] \\ &\quad +\lambda_2(1+\beta\lambda_1 x_1 - \beta) \frac{e^{-\theta_2\lambda_2(1+\beta\lambda_1 x_1)}}{(\lambda_2(1+\beta\lambda_1 x_1))^2} (\theta_2\lambda_2(1+\beta\lambda_1 x_1)+1) \\ &= \frac{e^{-\theta_2\lambda_2(1+\beta\lambda_1 x_1)}}{\lambda_2(1+\beta\lambda_1 x_1)^2} \Big[\beta\Big(\theta_2^2\lambda_2^2(1+\beta\lambda_1 x_1)^2 + 2\theta_2\lambda_2(1+\beta\lambda_1 x_1) + 2\Big) \\ &\quad +\theta_2\lambda_2(1+\beta\lambda_1 x_1)^2 + (1+\beta\lambda_1 x_1) - \beta\theta_2\lambda_2(1+\beta\lambda_1 x_1) - \beta \Big] \\ &= \frac{e^{-\theta_2\lambda_2(1+\beta\lambda_1 x_1)}}{\lambda_2(1+\beta\lambda_1 x_1)^2} \Big[\theta_2\lambda_2(1+\beta\lambda_1 x_1)^2(1+\beta\lambda_2\theta_2) + (1+\beta\lambda_1 x_1)(1+\beta\lambda_2\theta_2) + \beta \Big]. \end{aligned}$$

Inserting this result into the equation of $I(\theta_1,\theta_2)$ yields

$$\begin{aligned} I(\theta_1,\theta_2) &= \frac{\lambda_1 e^{-\lambda_2\theta_2}}{\lambda_2} \int_{\theta_1}^{\infty} \frac{x_1 e^{-\lambda_1 x_1(1+\beta\lambda_2\theta_2)}}{(1+\beta\lambda_1 x_1)^2} \Big[\theta_2\lambda_2(1+\beta\lambda_1 x_1)^2(1+\beta\lambda_2\theta_2) \\ &\quad + (1+\beta\lambda_1 x_1)(1+\beta\lambda_2\theta_2) + \beta \Big] dx_1, \end{aligned}$$

and by changing variable $\beta y = (1+\beta\lambda_1 x_1)$ and letting $c = \frac{1+\beta\lambda_1\theta_1}{\beta}$, we obtain

$$\begin{aligned} I(\theta_1,\theta_2) &= \frac{\lambda_1 e^{-\lambda_2\theta_2}}{\lambda_2} \int_{c}^{\infty} \frac{\beta y - 1}{\beta\lambda_1(\beta y)^2} e^{-\frac{\beta y-1}{\beta}(1+\beta\lambda_2\theta_2)} \Big[\theta_2\lambda_2(\beta y)^2(1+\beta\lambda_2\theta_2) \\ &\quad + (\beta y)(1+\beta\lambda_2\theta_2) + \beta \Big] \frac{dy}{\lambda_1} \\ &= \frac{e^{\frac{1}{\beta}}}{\beta\lambda_1\lambda_2} \int_{c}^{\infty} e^{-y(1+\beta\lambda_2\theta_2)} (\beta y - 1) \left[\theta_2\lambda_2(1+\beta\lambda_2\theta_2) + \frac{1+\beta\lambda_2\theta_2}{\beta y} + \frac{1}{\beta y^2} \right] dy. \end{aligned}$$

For simplicity, we denote $u_2 = 1 + \beta\lambda_2\theta_2$; hence

$$
\begin{aligned}
I(\theta_1,\theta_2) &= \frac{e^{\frac{1}{\beta}}}{\beta\lambda_1\lambda_2}\int_c^\infty e^{-yu_2}\left(\lambda_2\theta_2 u_2\beta y - \theta_2\lambda_2 u_2 + u_2 - \frac{u_2}{\beta y} + \frac{1}{y} - \frac{1}{\beta y^2}\right)dy \\
&= \frac{e^{\frac{1}{\beta}}}{\beta\lambda_1\lambda_2}\left[\beta\lambda_2\theta_2 u_2\int_c^\infty ye^{-yu_2}dy + u_2(1-\lambda_2\theta_2)\int_c^\infty e^{-yu_2}dy\right. \\
&\quad \left. + \left(1-\frac{u_2}{\beta}\right)\int_c^\infty \frac{e^{-yu_2}}{y}dy - \frac{1}{\beta}\int_c^\infty \frac{e^{-yu_2}}{y^2}dy\right].
\end{aligned}
$$

Using (*iii*.2), (*v*), and (*vi*) from Lemma 1, we evaluate

$$
\begin{aligned}
I(\theta_1,\theta_2) &= \frac{e^{\frac{1}{\beta}}}{\beta\lambda_1\lambda_2}[\beta\lambda_2\theta_2 u_2\Gamma(2,c,\infty;u_2) + u_2(1-\lambda_2\theta_2)\Gamma(1,c,\infty;u_2) \\
&\quad + \left(1-\frac{u_2}{\beta}\right)E_1(cu_2) - \frac{1}{\beta}\left(\frac{e^{-cu_2}}{c} - u_2E_1(cu_2)\right)\Bigg] \\
&= \frac{e^{\frac{1}{\beta}}}{\beta\lambda_1\lambda_2}\left[\beta\lambda_2\theta_2 u_2\frac{e^{-cu_2}}{u_2^2}(1+cu_2) + u_2(1-\lambda_2\theta_2)\frac{e^{-cu_2}}{u_2} - \frac{e^{-cu_2}}{\beta c} + E_1(cu_2)\right] \\
&= \frac{1}{\beta\lambda_1\lambda_2}\left[\left(\beta\lambda_2\theta_2\left(\frac{1}{u_2}+c\right) + 1 - \lambda_2\theta_2 - \frac{1}{\beta c}\right)e^{\frac{1}{\beta}-cu_2} + e^{\frac{1}{\beta}}E_1(cu_2)\right].
\end{aligned}
$$

We now insert the formulas of c and u_2; note that

$$
cu_2 = \frac{(1+\beta\lambda_1\theta_1)(1+\beta\lambda_2\theta_2)}{\beta}, \frac{1}{\beta} - cu_2 = -(\lambda_1\theta_1 + \lambda_2\theta_2 + \beta\lambda_1\lambda_2\theta_1\theta_2),
$$

and with some calculation, we obtain the stated formula of $I(\theta_1,\theta_2)$. □

***Proof of Lemma* 5.** We note that

$$
\begin{aligned}
F_{Y_2|Y_1=y_1}(y_2) &= \int_0^{y_2}\frac{g_{\mathbf{Y}}(y_1,x)}{g_{Y_1}(y_1)}dx = \frac{1}{g_{Y_1}(y_1)}L_1(y_1;y_2) \\
&= \frac{\lambda_1 e^{-\lambda_1 y_1}\left[1-(1+\beta\lambda_2 y_2)e^{-\lambda_2 y_2(1+\beta\lambda_1 y_1)}\right]}{\lambda_1 e^{-\lambda_1 y_1}},
\end{aligned}
$$

where we used the formula of L_1 from Lemma 3. This easily yields the stated result. □

***Proof of Proposition* 1.** We prove the formulas for X_1, with the formulas for X_2 resulting in a similar manner.
(*i*) Since $f_{X_1}(x_1) = \int_0^\infty f(x_1,x_2)dx_2$, we have two cases:
Case $x_1 \le \theta_1$: it is easy to see that

$$
f_{X_1}(x_1) = \frac{r}{P_D}\int_0^\infty g_{\mathbf{Y}}(x_1,x_2)dx_2 = \frac{r}{P_D}\lambda_1 e^{-\lambda_1 x_1}.
$$

Case $x_1 > \theta_1$: in this case,

$$
\begin{aligned}
f_{X_1}(x_1) &= \frac{r}{P_D}\int_0^{\theta_2} g_{\mathbf{Y}}(x_1,x_2)dx_2 + (1-r)\int_{\theta_2}^\infty f_{\mathbf{Z}}(x_1,x_2)dx_2 \\
&= \frac{r}{P_D}L_1(x_1;\theta_2) + (1-r)a\frac{\theta_1^a}{x_1^{a+1}}.
\end{aligned}
$$

We insert the formula of L_1 from Lemma 3 and obtain the stated formula of f_{X_1}.
(*ii*) Based on the formula of f_{X_1}, we again have two cases:
Case $x_1 \leq \theta_1$: clearly, here we obtain the cdf of the exponential distribution of Y_1.
Case $x_1 > \theta_1$: in this case,

$$\begin{aligned} F_{X_1}(x_1) &= \int_0^{\theta_1} \frac{r}{P_D}\lambda_1 e^{-\lambda_1 x}dx + \int_{\theta_1}^{x_1} \frac{r}{P_D}\lambda_1 e^{-\lambda_1 x}dx \\ &\quad - \int_{\theta_1}^{x_1} \frac{r}{P_D}\lambda_1 e^{-\lambda_1 x}(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2(1+\beta\lambda_1 x)}dx + \int_{\theta_1}^{x_1}(1-r)a\frac{\theta_1^a}{x_1^{a+1}}dx. \end{aligned}$$

The first two integrals add to the cdf of the exponential distribution of Y_1 in x_1, while the last integral yields the cdf of the Pareto distribution of Z_1. Therefore,

$$\begin{aligned} F_{X_1}(x_1) &= \frac{r}{P_D}\left(1-e^{-\lambda_1 x_1}\right) - \frac{r}{P_D}\lambda_1(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2}\int_{\theta_1}^{x_1} e^{-\lambda_1 x(1+\beta\lambda_2\theta_2)}dx \\ &\quad +(1-r)\left(1-\left(\frac{\theta_1}{x_1}\right)^a\right) \\ &= \frac{r}{P_D}\left(1-e^{-\lambda_1 x_1}\right) + \frac{r}{P_D}e^{-\lambda_2\theta_2}e^{-\lambda_1 x(1+\beta\lambda_2\theta_2)}\Big|_{\theta_1}^{x_1} + (1-r)\left(1-\left(\frac{\theta_1}{x_1}\right)^a\right) \\ &= \frac{r}{P_D}\left(1-e^{-\lambda_1\theta_1-\lambda_2\theta_2-\beta\lambda_1\lambda_2\theta_1\theta_2} + e^{-\lambda_2\theta_2-\lambda_1 x_1(1+\beta\lambda_2\theta_2)} - e^{-\lambda_1 x_1}\right) \\ &\quad +(1-r)\left(1-\left(\frac{\theta_1}{x_1}\right)^a\right) \\ &= r + \frac{r}{P_D}\left(e^{-\lambda_2\theta_2-\lambda_1 x_1(1+\beta\lambda_2\theta_2)} - e^{-\lambda_1 x_1}\right) + (1-r) - (1-r)\left(\frac{\theta_1}{x_1}\right)^a, \end{aligned}$$

where for the last equality, we used formula (3) of P_D. From here, the formula of F_{X_1} is immediate. □

***Proof of Proposition* 2.** (*i*) The continuity condition $f_{X_1}(\theta_1-) = f_{X_1}(\theta_1+)$ yields

$$\begin{aligned} \frac{r}{P_D}\lambda_1 e^{-\lambda_1\theta_1} &= \frac{r}{P_D}\lambda_1 e^{-\lambda_1\theta_1}\left[1-(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2(1+\beta\lambda_1\theta_1)}\right] + (1-r)a\frac{\theta_1^a}{\theta_1^{a+1}} \\ &\Leftrightarrow\ -\frac{r\lambda_1(1+\beta\lambda_2\theta_2)e^{-(\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2)}}{1-e^{-(\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2)}} + \frac{a}{\theta_1} - \frac{ra}{\theta_1} = 0 \\ &\Leftrightarrow\ \frac{a}{\theta_1} = r\frac{a}{\theta_1}\left(1+\frac{\theta_1}{a}\frac{\lambda_1(1+\beta\lambda_2\theta_2)}{e^{\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2}-1}\right), \end{aligned}$$

which yields Formula (*i*). The proof of Formula (*ii*) is similar.
(*iii*) We equate $r_1 = r_2$ from (*i*) and (*ii*) and obtain

$$\begin{aligned} & 1+\frac{\lambda_1\theta_1}{a}\frac{1+\beta\lambda_2\theta_2}{e^{\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2}-1} = 1+\frac{\lambda_2\theta_2}{a}\frac{1+\beta\lambda_1\theta_1}{e^{\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2}-1} \\ \Leftrightarrow\ & \lambda_1\theta_1(1+\beta\lambda_2\theta_2) = \lambda_2\theta_2(1+\beta\lambda_1\theta_1) \\ \Leftrightarrow\ & \lambda_1\theta_1 = \lambda_2\theta_2. \end{aligned}$$

Moreover, the continuity condition at (θ_1, θ_2) means $r = r_1 = r_2$; hence, using (11) and $\lambda_1\theta_1 = \lambda_2\theta_2$, we obtain

$$\begin{aligned} & 1+\frac{(1+\beta\lambda_1\theta_1)(1+\beta\lambda_2\theta_2)-\beta}{\left(e^{\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2}-1\right)(a+1)a}\lambda_1\lambda_2\theta_1\theta_2 = 1+\frac{\lambda_1\theta_1}{a}\frac{1+\beta\lambda_2\theta_2}{e^{\lambda_1\theta_1+\lambda_2\theta_2+\beta\lambda_1\lambda_2\theta_1\theta_2}-1} \\ \Leftrightarrow\ & \left((1+\beta\lambda_1\theta_1)^2-\beta\right)\lambda_1\theta_1 = (a+1)(1+\beta\lambda_1\theta_1), \end{aligned}$$

from which results the stated formula of a. □

***Proof of Proposition* 3.** We calculate the expected value and the second-order moment for X_1 (those of X_2 result in a similar way). Using the expected value of the exponential and Pareto distributions, we have

$$
\begin{aligned}
\mathbb{E}X_1 &= \int_0^\infty x_1 f_{X_1}(x_1)dx_1 = \frac{r}{P_D}\int_0^{\theta_1} \lambda_1 x_1 e^{-\lambda_1 x_1} dx_1 \\
&+ \frac{r}{P_D}\int_{\theta_1}^\infty \lambda_1 x_1 e^{-\lambda_1 x_1}\left[1-(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2(1+\beta\lambda_1 x_1)}\right]dx_1 + (1-r)\int_{\theta_1}^\infty x_1 \frac{a\theta_1^a}{x_1^{a+1}}dx_1 \\
&= \frac{r}{P_D}\left[\int_0^\infty \lambda_1 x_1 e^{-\lambda_1 x_1}dx_1 - \lambda_1(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2}\int_{\theta_1}^\infty x_1 e^{-\lambda_1 x_1(1+\beta\lambda_2\theta_2)}dx_1\right] \\
&+ (1-r)\frac{a\theta_1}{a-1} \\
&= \frac{r}{P_D}\left[\frac{1}{\lambda_1} - \lambda_1(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2}\Gamma(2,\theta_1,\infty;\lambda_1(1+\beta\lambda_2\theta_2))\right] + (1-r)\frac{a\theta_1}{a-1}.
\end{aligned}
$$

Inserting (*iii*.2) from Lemma 1 yields

$$
\mathbb{E}X_1 = \frac{r}{P_D}\left[\frac{1}{\lambda_1} - \lambda_1(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2}\frac{e^{-\lambda_1\theta_1(1+\beta\lambda_2\theta_2)}}{\lambda_1^2(1+\beta\lambda_2\theta_2)^2}(1+\lambda_1\theta_1(1+\beta\lambda_2\theta_2))\right] + (1-r)\frac{a\theta_1}{a-1},
$$

from which the expected value formula is immediate. The moment of second order is

$$
\begin{aligned}
\mathbb{E}X_1^2 &= \int_0^\infty x_1^2 f_{X_1}(x_1)dx_1 = \frac{r}{P_D}\left[\lambda_1\int_0^\infty x_1^2 e^{-\lambda_1 x_1}dx_1\right. \\
&\left. -\lambda_1(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2}\int_{\theta_1}^\infty x_1^2 e^{-\lambda_1 x_1(1+\beta\lambda_2\theta_2)}dx_1\right] + (1-r)\int_{\theta_1}^\infty x_1^2\frac{a\theta_1^a}{x_1^{a+1}}dx_1 \\
&= \frac{r\lambda_1}{P_D}\left[\Gamma(3,0,\infty;\lambda_1) - (1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2}\Gamma(3,\theta_1,\infty;\lambda_1(1+\beta\lambda_2\theta_2))\right] + (1-r)\frac{a\theta_1^2}{a-2}.
\end{aligned}
$$

Based on (*iv*.2) from Lemma 1, we obtain

$$
\begin{aligned}
\mathbb{E}X_1^2 &= \frac{r\lambda_1}{P_D}\left[\frac{2}{\lambda_1^3} - (1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2}\frac{e^{-\lambda_1\theta_1(1+\beta\lambda_2\theta_2)}}{\lambda_1^3(1+\beta\lambda_2\theta_2)^3}\left(1+(\lambda_1\theta_1(1+\beta\lambda_2\theta_2)+1)^2\right)\right] \\
&+ (1-r)\frac{a\theta_1^2}{a-2}.
\end{aligned}
$$

The stated formula of $\mathbb{E}X_1^2$ easily results from here, which completes the proof. □

***Proof of Proposition* 4.** We write

$$
\begin{aligned}
\mathbb{E}[X_1X_2] &= \int_0^\infty\int_0^\infty x_1x_2 f(x_1,x_2)dx_1dx_2 \\
&= \int\int_D \frac{r\lambda_1\lambda_2}{P_D}x_1x_2e^{-(\lambda_1x_1+\lambda_2x_2+\beta\lambda_1\lambda_2x_1x_2)}[(1+\beta\lambda_1x_1)(1+\beta\lambda_2x_2)-\beta]dx_1dx_2 \\
&+ \int\int_{D_{22}}(1-r)a(a+1)\frac{(\theta_1\theta_2)^{a+1}x_1x_2}{(\theta_2x_1+\theta_1x_2-\theta_1\theta_2)^{a+2}}dx_1dx_2 \\
&= \frac{r}{P_D}I_1 + (1-r)I_2.
\end{aligned}
$$

We separately calculate the two integrals. We start with the second one, which from Formula (5) is given by

$$I_2 = \int_{\theta_1}^{\infty}\int_{\theta_1}^{\infty} a(a+1)\frac{x_1x_2(\theta_1\theta_2)^{a+1}}{(\theta_2x_1+\theta_1x_2-\theta_1\theta_2)^{a+2}}dx_1dx_2 = \theta_1\theta_2\frac{a^2-a-1}{(a-1)(a-2)}.$$

In what concerns I_1, we note that given the definition of the domain D with the notation from Lemma 4, we have

$$\begin{aligned} I_1 &= \int\int_D \lambda_1\lambda_2x_1x_2e^{-(\lambda_1x_1+\lambda_2x_2+\beta\lambda_1\lambda_2x_1x_2)}[(1+\beta\lambda_1x_1)(1+\beta\lambda_2x_2)-\beta]dx_1dx_2 \\ &= I(0,0)-I(\theta_1,\theta_2). \end{aligned}$$

Now using the formula in Lemma 4, we note that

$$I(0,0) = \frac{1}{\beta\lambda_1\lambda_2}E_1\left(\frac{1}{\beta}\right)e^{\frac{1}{\beta}},$$

and the stated formula of $\mathbb{E}[X_1X_2]$ results immediately. □

***Proof of Proposition* 5.** We recall that

$$f_{X_2|X_1=x_1}(x_2) = \frac{f(x_1,x_2)}{f_{X_1}(x_1)},$$

and according to Proposition 1, we note that we must consider three different cases: $(x_1 \le \theta_1, x_2 > 0)$; $(x_1 > \theta_1, x_2 \le \theta_2)$ and $(x_1 > \theta_1, x_2 > \theta_2)$.
Case I: $x_1 \le \theta_1$ and $x_2 > 0$. In this case,

$$F_{X_2|X_1=x_1}(x_2) = \int_0^{x_2}\frac{\frac{r}{P_D}g_{\mathbf{Y}}(x_1,x)}{\frac{r}{P_D}\lambda_1e^{-\lambda_1x_1}}dx,$$

and using Lemma 5, we obtain the first formula of $F_{X_2|X_1=x_1}$.
Case II: $x_1 > \theta_1$ and $x_2 \le \theta_2$. Now, we have

$$F_{X_2|X_1=x_1}(x_2) = \int_0^{x_2}\frac{\frac{r}{P_D}g_{\mathbf{Y}}(x_1,x)}{\frac{r}{P_D}\lambda_1e^{-\lambda_1x_1}\left[1-(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2(1+\beta\lambda_1x_1)}\right]+(1-r)a\frac{\theta_1^a}{x_1^{a+1}}}dx,$$

and, as in Case I, we easily get the second formula of $F_{X_2|X_1=x_1}(x_2)$.
Case III: $x_1 > \theta_1$ and $x_2 > \theta_2$. In this case,

$$\begin{aligned} F_{X_2|X_1=x_1}(x_2) &= \int_0^{\theta_2}\frac{\frac{r}{P_D}g_{\mathbf{Y}}(x_1,x)}{\frac{r}{P_D}\lambda_1e^{-\lambda_1x_1}\left[1-(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2(1+\beta\lambda_1x_1)}\right]+(1-r)a\frac{\theta_1^a}{x_1^{a+1}}}dx \\ &+ \int_{\theta_2}^{x_2}\frac{(1-r)(a+1)a\frac{(\theta_1\theta_2)^{a+1}}{(\theta_2x_1+\theta_1x-\theta_1\theta_2)^{a+2}}}{\frac{r}{P_D}\lambda_1e^{-\lambda_1x_1}\left[1-(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2(1+\beta\lambda_1x_1)}\right]+(1-r)a\frac{\theta_1^a}{x_1^{a+1}}}dx. \end{aligned}$$

The first integral equals the formula obtained in Case II by taking $x_2 = \theta_2$, while for the second integral, we evaluate

$$\begin{aligned} J &= (a+1)a\int_{\theta_2}^{x_2}(\theta_1\theta_2)^{a+1}(\theta_2 x_1+\theta_1 x-\theta_1\theta_2)^{-a-2}dx \\ &= (a+1)a(\theta_1\theta_2)^{a+1}\frac{1}{\theta_1}\frac{(\theta_2 x_1+\theta_1 x-\theta_1\theta_2)^{-a-1}}{-(a+1)}\Big|_{\theta_2}^{x_2} \\ &= -a\theta_1^a\theta_2^{a+1}\Big[(\theta_2 x_1+\theta_1 x_2-\theta_1\theta_2)^{-a-1}-(\theta_2 x_1+\theta_1\theta_2-\theta_1\theta_2)^{-a-1}\Big] \\ &= -a\frac{\theta_1^a\theta_2^{a+1}}{(\theta_2 x_1+\theta_1 x_2-\theta_1\theta_2)^{a+1}}+a\frac{\theta_1^a\theta_2^{a+1}}{\theta_2^{a+1}x_1^{a+1}}. \end{aligned}$$

Therefore,

$$\begin{aligned} F_{X_2|X_1=x_1}(x_2) &= \frac{r}{P_D}\lambda_1 e^{-\lambda_1 x_1}\frac{1-e^{-\lambda_2\theta_2(1+\beta\lambda_1 x_1)}(1+\beta\lambda_2\theta_2)}{\frac{r}{P_D}\lambda_1 e^{-\lambda_1 x_1}\left[1-(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2(1+\beta\lambda_1 x_1)}\right]+(1-r)a\frac{\theta_1^a}{x_1^{a+1}}} \\ &\quad+(1-r)\frac{\frac{a\theta_1^a}{x_1^{a+1}}-a\frac{\theta_1^a\theta_2^{a+1}}{(\theta_2 x_1+\theta_1 x_2-\theta_1\theta_2)^{a+1}}}{\frac{r}{P_D}\lambda_1 e^{-\lambda_1 x_1}\left[1-(1+\beta\lambda_2\theta_2)e^{-\lambda_2\theta_2(1+\beta\lambda_1 x_1)}\right]+(1-r)a\frac{\theta_1^a}{x_1^{a+1}}}, \end{aligned}$$

from which the last formula of $F_{X_2|X_1=x_1}$ is immediate. □

References

1. Sarabia, J. M.; Gómez-Déniz, E. Construction of multivariate distributions: A review of some recent results. *SORT* **2008**, *32*, 3–36.
2. Klugman, S.A.; Panjer, H.H.; Willmot, G.E. *Loss Models: From Data to Decisions*; John Wiley & Sons: New York, NY, USA, 1998.
3. Scarrott, C. Univariate extreme value mixture modeling. In *Extreme Value Modeling and Risk Analysis: Methods and Applications*; Dipak, K.D., Jun, Y., Eds.; CRC Press: Boca Raton, FL, USA, 2016; pp. 41–67.
4. Cooray, K.; Ananda, M.M. Modeling actuarial data with a composite lognormal-Pareto model. *Scand. Actuar. J.* **2005**, *5*, 321–334. [CrossRef]
5. Bolancé, C.; Guillen, M.; Pelican, E.; Vernic, R. Skewed bivariate models and nonparametric estimation for the CTE risk measure. *Insur. Math. Econ.* **2008**, *43*, 386–393. [CrossRef]
6. Bahraoui, Z.; Bolancé, C.; Pelican, E.; Vernic, R. On the bivariate Sarmanov distribution and copula. An application on insurance data using truncated marginal distributions. *SORT* **2015**, *39*, 209–230.
7. Gumbel, E.J. Bivariate exponential distributions. *J. Am. Stat. Assoc.* **1960**, *55*, 698–707. [CrossRef]
8. Kotz, S.; Balakrishnan, N.; Johnson, N.L. *Continuous Multivariate Distributions, Volume 1: Models and Applications*; John Wiley & Sons: New York, NY, USA, 2004.
9. Castillo, E.; Sarabia, J.M.; Hadi, A.S. Fitting continuous bivariate distributions to data. *Statistician* **1997**, *46*, 355–369. [CrossRef]
10. Mardia, K.V. Multivariate Pareto distributions. *Ann. Math. Stat.* **1962**, *33*, 1008–1015. [CrossRef]
11. Scarrott, C.J.; MacDonald, A. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT Stat. J.* **2012**, *10*, 33–60.
12. Bahraoui, Z.; Bolancé, C.; Pérez-Marín, A.M. Testing extreme value copulas to estimate the quantile. *SORT* **2014**, *38*, 89–102.
13. Chen, J. Optimal rate of convergence for finite mixture models. *Ann. Stat.* **1995**, *23*, 221–233. [CrossRef]
14. Kim, D.; Lindsay, B.G. Empirical identifiability in finite mixture models. *Ann. Inst. Stat. Math.* **2015**, *67*, 745–772. [CrossRef]