Asimit, A. V., Furman, E. & Vernic, R. (2016). Statistical Inference for a New Class of Multivariate Pareto Distributions. Communications in Statistics: Simulation and Computation, 45(2), pp. 456-471. doi: 10.1080/03610918.2013.861627

STATISTICAL INFERENCE FOR A NEW CLASS OF MULTIVARIATE PARETO DISTRIBUTIONS

Alexandru V. Asimit[1]

Cass Business School, City University, London EC1Y 8TZ, United Kingdom. E-mail: asimit@city.ac.uk

Edward Furman

Department of Mathematics and Statistics, York University, Toronto, Ontario M3J 1P3, Canada. E-mail: efurman@mathstat.yorku.ca

Raluca Vernic

Faculty of Mathematics and Informatics, Ovidius University of Constanta, Constanta, and Institute of Mathematical Statistics and Applied Mathematics, Bucharest, Romania. E-mail: rvernic@univ-ovidius.ro

October 17, 2013

Abstract. Various solutions to the parameter estimation problem of the multivariate Pareto distribution of Asimit et al. (2010) are developed and exemplified numerically. Namely, a density of the aforementioned multivariate Pareto distribution with respect to a dominating measure, rather than the corresponding Lebesgue measure, is specified and then employed to investigate the maximum likelihood estimation (MLE) approach. Also, in an attempt to fully enjoy the common shock origins of the multivariate

[1]Corresponding author. Phone: +44(0)2070405282. Fax: +44(0)2070408572.

model of interest, an adapted variant of the expectation maximization (EM) algorithm is formulated and studied. The method of moments is discussed as a convenient way to obtain starting values for the numerical optimization procedures associated with the MLE and EM methods.

Keywords and phrases: Multivariate Pareto distribution, common shock model, maximum likelihood estimation, expectation maximization algorithm, method of moments.

Mathematics Subject Classification: 62F10, 62H12, 60E05.

1. Introduction

Fix a measurable space $(\Omega,\ \mathcal{F})$ and $n \in \mathbf{Z}_+$, and let $\mathbf{Y} = (Y_0, Y_1, \ldots, Y_n)' : \Omega \to \mathbf{R}^{n+1}$ denote an $(n+1)$-dimensional random vector possessing mutually independent Pareto-distributed coordinates $Y_i \backsim Pa(\mu_i, \sigma_i, \alpha_i),\ i = 0,\ 1, \ldots, n$ ('$\backsim$' stands for 'distributed' throughout), such that, for $\sigma_i \in \mathbf{R}_+$ and $\alpha_i \in \mathbf{R}_+$, we have that

$$\overline{F}_{Y_i}(y) = \mathbf{P}[Y_i > y] = \left(1 + \frac{y_i - \mu_i}{\sigma_i}\right)^{-\alpha_i},\ \textit{with}\ y > \mu_i \in \mathbf{R}. \tag{1.1}$$

Then, for $\mu_0 = 0,\ \sigma_0 = 1,\ \alpha_{0j} = \alpha_0 + \alpha_j,\ j = 1, \ldots, n,$ and a map $T : \mathbf{R}^{n+1} \to \mathbf{R}^n$, the random vector $\mathbf{X} = T(\mathbf{Y})$, with the coordinates $X_j = \min(\sigma_j Y_0 + \mu_j, Y_j) \backsim Pa(\mu_j,\ \sigma_j,\ \alpha_{0j})$, is in Asimit et al. (2010) referred to as a multivariate Pareto distribution having arbitrarily parameterized Pareto of the second kind margins (see, Arnold, 1983), and a dependence structure, described by the Marshall and Olkin copula (see, Marshall and Olkin, 1967). Various applications of the just-mentioned multivariate Pareto distribution in e.g. actuarial mathematics and/or operational research stem from its 'common shock' - based formation (see, loc. cit., as well as Asimit et al., 2010).

The multivariate Pareto distribution above, which was in Asimit et al. (2010) denoted by $Pa_n(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$, with the vectors of parameters $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' \in \mathbf{R}^n$, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)' \in \mathbf{R}_+^n$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)' \in \mathbf{R}_+^n$ and a scalar-valued 'dependence' parameter $\alpha_0 > 0$, proved to be quite analytically tractable and thus allowed for a comprehensive study of a number of its properties. More specifically, we derived explicit expressions for, e.g., the decumulative distribution functions (d.d.f.'s), the probability density functions (p.d.f.'s), and the conditional as well as joint moments, proved certain characteristic results, and developed pricing formulas. A discussion of the appropriate inferential statistics techniques for $Pa_n(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$, seems therefore to suggest itself.

Our interest in this paper is therefore to find estimates of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\alpha}$ and α_0. Speaking plainly, the problem is not trivial. Indeed, notice that the maximum likelihood estimation (MLE) seems not at first glance applicable because the d.d.f. of $\mathbf{X} \backsim Pa_n(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$, which is given by (see, loc. cit.)

$$\overline{F}_{\mathbf{X}}(x_1, \ldots, x_n) = \left(1 + \max_{j=1,\ldots,n} \frac{x_j - \mu_j}{\sigma_j}\right)^{-\alpha_0} \prod_{j=1}^{n} \left(1 + \frac{x_j - \mu_j}{\sigma_j}\right)^{-\alpha_j}, \tag{1.2}$$

with $x_j > \mu_j$, $j = 1, \ldots, n$, is not absolutely continuous with respect to the Lebesgue measure on $\mathbf{R}^n$. Furthermore, even the moment-based estimation can become somewhat intricate if, say, the expectation and/or variance are not finite, which can certainly be the case, e.g., we readily have that if $\alpha_{0j} \leq 1$, then $\mathbf{E}[X_j]$ is infinite. It is worthwhile noticing that the aforementioned statistical inconvenience is often an advantage, and it is in fact quite desirable in practical applications for modeling 'particularly heavy' financial risks and/or losses.

In the rest of the paper we attempt to provide possible ways to tackle the parameters estimation issue. To this end, in Section 2.2 we specify a density of $\mathbf{X} \backsim Pa_n(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$

with respect to a dominating measure, instead of the Lebesgue measure, which makes the MLE method feasible, and we discuss the expectation maximization (EM) algorithm to estimate the parameters in Section 2.3. Section 3 reveals a numerical study and concludes the paper. The proofs are relegated to the appendix.

2. Main results

2.1. **Basic properties.** In Asimit et al. (2010) we showed that, for $j = 1, \dots, n$,

- The distribution of X_j is $Pa(\mu_j, \sigma_j, \alpha_{0j})$.
- The mathematical expectation of X_j is, for $\alpha_{0j} > 1$,

$$\mathbf{E}[X_j] = \mu_j + \frac{\sigma_j}{\alpha_{0j} - 1}.$$

- The variance of X_j is, for $\alpha_{0j} > 2$,

$$\mathbf{Var}[X_j] = \frac{\alpha_{0j}\sigma_j^2}{(\alpha_{0j} - 1)^2(\alpha_{0j} - 2)}.$$

- The covariance between X_j and X_k is, for $j \neq k$, $\alpha_{0j} > 1$, $\alpha_{0k} > 1$ and $\alpha_{0jk} = \alpha_0 + \alpha_j + \alpha_k > 2$,

$$\mathbf{Cov}[X_j, X_k] = \frac{\alpha_0 \sigma_j \sigma_k}{(\alpha_{0j} - 1)(\alpha_{0k} - 1)(\alpha_{0jk} - 2)}.$$

- Pearson's correlation coefficient between X_j and X_k is, for $j \neq k$, $\alpha_{0j} > 2$, $\alpha_{0k} > 2$ and $\alpha_{0jk} > 2$,

$$\mathbf{Corr}[X_j, X_k] = \frac{\alpha_0}{\alpha_{0jk} - 2}\sqrt{\frac{(\alpha_{0j} - 2)(\alpha_{0k} - 2)}{\alpha_{0j}\alpha_{0k}}}. \tag{2.1}$$

Furthermore, for $1 \le j \neq k \le n$ and $x_k > \mu_k$, the conditional d.d.f. of $X_j |\, X_k = x_k$, as well as the conditional expectation $\mathbf{E}[X_j |\, X_k = x_k]$ were also derived in Asimit et al. (2010).

As it has been mentioned, the multivariate Pareto distribution of interest in this paper possesses Pareto of the second kind margins and Marshall-Olkin copula-based dependence structure (see, Nelsen, 1999, p. 46). For the sake of the analysis in Section 2.3, we further complement (2.1) with the following two well-known robust measures of association, i.e., Kendall's Tau and Spearman's Rho (see, Nelsen, 1999, p's. 133 and 136; for the proofs).

Lemma 2.1. *Let* $\mathbf{X} \backsim Pa_n(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$ *follow the multivariate Pareto distribution of interest, and let* $1 \leq j \neq k \leq n$*. Then it can be shown that,*

- *Spearman's rank correlation coefficient between* X_j *and* X_k *is*

$$\rho[X_j, X_k] = 3\alpha_0/(2(\alpha_{0j} + \alpha_{0k}) - \alpha_0), \text{ and} \tag{2.2}$$

- *Kendall's tau rank correlation coefficient between* X_j *and* X_k *is*

$$\tau[X_j, X_k] = \alpha_0/(\alpha_{0j} + \alpha_{0k} - \alpha_0). \tag{2.3}$$

2.2. **Density and likelihood.** It is not difficult to see that the p.d.f. does not everywhere exist for d.d.f. (1.2). We skip negligible technical details and only note in passing that the d.d.f. consists of both absolutely continuous and singular components. Thereby, with a fixed index $(n) \in \{1, \ldots, n\}$, such that

$$z_{(n)} = \max_{j=1,\ldots,n} (x_j - \mu_j)/\sigma_j, \tag{2.4}$$

and $\alpha_{(0n)}$ and $\sigma_{(n)}$ denoting the shape and scale parameters corresponding to the just introduced coordinate $z_{(n)}$, we have that the p.d.f. for the absolutely continuous part is

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = \frac{\alpha_{(0n)}}{\sigma_{(n)}} \left(1 + z_{(n)}\right)^{-\left(\alpha_{(0n)}+1\right)} \prod_{j=1,\ j\neq(n)}^{n} \frac{\alpha_j}{\sigma_j} \left(1 + \frac{x_j - \mu_j}{\sigma_j}\right)^{-(\alpha_j+1)}, \tag{2.5}$$

where $x_j > \mu_j$. In addition, for distinct $j_1 \ldots, j_k \in \{1,\ 2, \ldots, n\}$ and a fixed $k \leq n$, the probabilities for the singular component are given by

$$\mathbf{P}\left[\frac{X_{j_1} - \mu_{j_1}}{\sigma_{j_1}} = \cdots = \frac{X_{j_k} - \mu_{j_k}}{\sigma_{j_k}}\right] = \frac{\alpha_0}{\alpha_0 + \sum_{i=1}^{k} \alpha_{j_i}}. \tag{2.6}$$

We further specify a probability density function of the multivariate Pareto distribution with respect to a dominating measure, rather than to the Lebesgue measure on $\mathbf{R}^n$ (see, e.g., Proschan and Sullo, 1976; Hanagal, 1996; for similar approaches). To this end, for $r = 2, \ldots, n$, let $\mathcal{I}_r = \{i_1, \ldots, i_r\} \subseteq \{1, \ldots, n\} = \mathcal{I}_n$, and $\{j_1, \ldots, j_{n-r}\} = \mathcal{I}_n \setminus \mathcal{I}_r$. Also, for $\mathcal{C} \subseteq \mathbf{R}^n$, let us define

$$g_{\mathcal{I}_r}(\mathcal{C}) = \left\{ \left(z_{(n)}, x_{j_1}, ..., x_{j_{n-r}}\right) : \mathbf{x} \in \mathcal{C} \text{ and } \frac{x_{i_1} - \mu_{i_1}}{\sigma_{i_1}} = \cdots = \frac{x_{i_r} - \mu_{i_r}}{\sigma_{i_r}} = z_{(n)} \right\},$$

with $z_{(n)}$ given in (2.4). Then for ν_n denoting the n-dimensional Lebesgue measure, it is possible to introduce another measure $\nu >> \nu_n$ (in words, 'ν dominates ν_n'), such that

$$\nu(\mathcal{C}) = \nu_n(\mathcal{C}) + \sum_{\substack{r=2,\ldots,n \\ \mathcal{I}_r \subseteq \mathcal{I}_n}} \nu_{n-r+1}\left(g_{\mathcal{I}_r}\left(\mathcal{C} \bigcap \{\mathbf{x} \in \mathbf{R}^n : x_i > \mu_i,\ i = 1, \ldots, n\}\right)\right) \tag{2.7}$$

for any $\mathcal{C}$ in the Borel σ-algebra $\mathcal{B}^n$ in $\mathbf{R}^n$.

Furthermore, let $\mathbf{x} = (x_1, \ldots, x_n)' \in \mathbf{R}^n$ be a realization of $\mathbf{X} \backsim Pa_n(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$. We then introduce a number of auxiliary indexes as functions of $\mathbf{x}$, for $j = 1, \ldots, n$, $k = 1, \ldots, n$, and '$\#$' denoting the 'cardinality' of a set, i.e., let

$$v_j = v_j(\mathbf{x}) = \begin{cases} 1, & (x_j - \mu_j)/\sigma_j < z_{(n)} \\ 0, & \text{otherwise} \end{cases},$$

$$s = s(\mathbf{x}) = \begin{cases} 1, & \exists\ j \neq k : (x_j - \mu_j)/\sigma_j = (x_k - \mu_k)/\sigma_k = z_{(n)} \\ 0, & \text{otherwise} \end{cases},$$

and

$$r = r(\mathbf{x}) = \# \left\{ j \in \{1, \ldots, n\} : \frac{x_j - \mu_j}{\sigma_j} = z_{(n)} \right\}.$$

(When no confusion is possible, we omit the argument '$\mathbf{x}$', and we write v_j, s and r instead of $v_j(\mathbf{x}), s(\mathbf{x})$ and $r(\mathbf{x})$, respectively.)

Theorem 2.1. *The density of* $\mathbf{X} \backsim Pa_n(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$ *with respect to* ν *is given by*

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = \alpha_0^s (1 + z_{(n)})^{-\alpha_0 + r - 1} \prod_{j=1}^{n} \left(\frac{\alpha_j}{\sigma_j}\right)^{v_j} \left(\frac{\alpha_{0j}}{\sigma_j}\right)^{(1-s)(1-v_j)} \left(1 + \frac{x_j - \mu_j}{\sigma_j}\right)^{-(\alpha_j+1)},$$

where $x_j > \mu_j \in \mathbf{R},\ j = 1, \ldots, n.$

Before sketching the MLE method, it is worthwhile noticing the particular forms of the bivariate and trivarite p.d.f.

Corollary 2.1. *The p.d.f. of* $\mathbf{X} \backsim Pa_2(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$ *is given by*

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \dfrac{\alpha_j \alpha_{0i}}{\sigma_1 \sigma_2} \left(1 + \dfrac{x_i - \mu_i}{\sigma_i}\right)^{-(\alpha_{0i}+1)} \left(1 + \dfrac{x_j - \mu_j}{\sigma_j}\right)^{-(\alpha_j+1)}, & \dfrac{x_i - \mu_i}{\sigma_i} > \dfrac{x_j - \mu_j}{\sigma_j} > 0 \\ & \text{and } i \neq j \in \{1, 2\} \\ \alpha_0 \left(1 + z_{(2)}\right)^{-(\alpha_0 + \alpha_1 + \alpha_2 + 1)}, & \dfrac{x_1 - \mu_1}{\sigma_1} = \dfrac{x_2 - \mu_2}{\sigma_2} = z_{(2)} > 0 \end{cases}, \tag{2.8}$$

while for $\mathbf{X} \backsim Pa_3(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$ *and* $\{i, j, k\} = \{1, 2, 3\}$ *it holds that*

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \dfrac{\alpha_i \alpha_j \alpha_{0k}}{\sigma_1 \sigma_2 \sigma_3} \left(1 + \dfrac{x_k - \mu_k}{\sigma_k}\right)^{-(\alpha_{0k}+1)} \left(1 + \dfrac{x_i - \mu_i}{\sigma_i}\right)^{-(\alpha_i+1)} \left(1 + \dfrac{x_j - \mu_j}{\sigma_j}\right)^{-(\alpha_j+1)}, \\ \qquad \dfrac{x_k - \mu_k}{\sigma_k} > \max\left\{\dfrac{x_i - \mu_i}{\sigma_i}, \dfrac{x_j - \mu_j}{\sigma_j}\right\} > 0 \\ \dfrac{\alpha_i \alpha_0}{\sigma_i} \left(1 + z_{(3)}\right)^{-(\alpha_0 + \alpha_j + \alpha_k + 1)} \left(1 + \dfrac{x_i - \mu_i}{\sigma_i}\right)^{-(\alpha_i+1)}, \\ \qquad \dfrac{x_k - \mu_k}{\sigma_k} = \dfrac{x_j - \mu_j}{\sigma_j} = z_{(3)} > \dfrac{x_i - \mu_i}{\sigma_i} > 0 \\ \alpha_0 \left(1 + z_{(3)}\right)^{-(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + 1)}, \quad \dfrac{x_1 - \mu_1}{\sigma_1} = \dfrac{x_2 - \mu_2}{\sigma_2} = \dfrac{x_3 - \mu_3}{\sigma_3} = z_{(3)} > 0 \end{cases}.$$

Theorem 2.1 establishes an absolutely continuous p.d.f., which can be used to develop the MLE for the multivariate Pareto distribution of interest, as it is shown in the sequel.

Let $(\mathbf{X}_i)_{i=1}^m$ be m independent copies of $\mathbf{X} \backsim Pa_n(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$, with the realization of, say, $\mathbf{X}_i$, being denoted by $\mathbf{x}_i = (x_{1,i}, \ldots, x_{n,i})'$, and let, for $j = 1, \ldots, n$,

$$u_0 = \sum_{i=1}^{m} s(\mathbf{x}_i), \quad u_j = \sum_{i=1}^{m} v_j(\mathbf{x}_i), \quad w_j = \sum_{i=1}^{m} (1 - s(\mathbf{x}_i))(1 - v_j(\mathbf{x}_i)).$$

Let $z_{(n)i} = \max_{j=1,\ldots,n}(x_{j,i} - \mu_j)/\sigma_j$, for a fixed $i = 1, \ldots, m$. The following statement is a clear consequence of Theorem 2.1 and is therefore given without proof.

Corollary 2.2. *The log-likelihood function for the* $Pa_n(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$ *sample* $(\mathbf{X}_i)_{i=1}^m$ *is*

$$\ln \mathcal{L}\left(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0; \mathbf{x}_1, \ldots, \mathbf{x}_m\right) = u_0 \ln \alpha_0 + \sum_{i=1}^{m} \left(r\left(\mathbf{x}_i\right) - \alpha_0 - 1\right) \ln(1 + z_{(n)i}) \tag{2.9}$$

$$+ \sum_{j=1}^{n} \left(u_j \ln \alpha_j + w_j \ln \alpha_{0j} - (u_j + w_j) \ln \sigma_j - (\alpha_j + 1) \sum_{i=1}^{m} \ln \left(1 + \frac{x_{j,i} - \mu_j}{\sigma_j} \right) \right),$$

which simplifies to the findings of Hanagal (1996) for $\mu_j = \sigma_j \equiv 1$.

At this point, the ideal solution is of course to estimate all $(3n + 1)$ parameters of $Pa_n(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$ applying the just derived log-likelihood. However, this can become rather cumbersome (if not impossible) if, say, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are unknown, since in such a case, we have that, e.g., s and v_j are unknown as well. Remarkably, it is possible to tackle the aforementioned complication by following an alternative route to estimating the parameters of interest, that we in fact do in Subsection 2.3.

To complement the current discussion, we further outline a number of seemingly useful observations, which can be of importance to practitioners under specific constraints. To start off, we note that the obvious estimates for μ_j are $\hat{\mu}_j = \min_{i=1,\ldots,m} x_{j,i}$, where $j = 1, \ldots, n$, and we thus have to actually estimate $(2n + 1)$ parameters, only, with further simplifications sometimes possible.

Indeed, an interesting special case in this respect is the one when the multivariate Pareto distribution of interest possesses identically distributed margins, or, more generally, when $\sigma_j \equiv \sigma$. Then we readily observe that the v, s and r functions do not depend on the values of σ, and therefore a system of $(n + 2)$ non-linear equations must be solved to obtain the estimates of σ, α_0 and $\boldsymbol{\alpha}$. The system is given below, and it is not solvable analytically

$$\begin{cases} \sum_{i=1}^{m}\left(\sum_{j=1}^{n}(\alpha_j+1)\dfrac{x_{j,i}-\hat{\mu}_j}{\sigma+x_{j,i}-\hat{\mu}_j}-(r(\mathbf{x}_i)-\alpha_0-1)\dfrac{z_{(n)i}}{1+z_{(n)i}}\right)-\sum_{j=1}^{n}(u_j+w_j)=0, \\ \dfrac{u_0}{\alpha_0}-\sum_{i=1}^{m}\ln\left(1+z_{(n)i}\right)+\sum_{j=1}^{n}\dfrac{w_j}{\alpha_{0j}}=0, \text{ and} \\ \dfrac{u_j}{\alpha_j}+\dfrac{w_j}{\alpha_{0j}}-\sum_{i=1}^{m}\ln\left(1+\dfrac{x_{j,i}-\hat{\mu}_j}{\sigma}\right)=0, \quad \text{for each } j=1,\ldots,n \end{cases}.$$

In a variety of practical applications, it may be convenient to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ using the marginal (univariate) estimation discussed in, e.g., Arnold (1983), and to utilize the log-likelihood function obtained in Corollary 2.2 to find the estimates of α_0 and $\boldsymbol{\alpha}$. Thereby, assuming that we have the estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\sigma}}$, we readily end up with the system

$$\begin{cases} \dfrac{u_0}{\alpha_0}-\sum_{i=1}^{m}\ln\left(1+z_{(n)i}\right)+\sum_{j=1}^{n}\dfrac{w_j}{\alpha_{0j}}=0, \text{ and} \\ \dfrac{u_j}{\alpha_j}+\dfrac{w_j}{\alpha_{0j}}-\sum_{i=1}^{m}\ln\left(1+\dfrac{x_{j,i}-\hat{\mu}_j}{\hat{\sigma}_j}\right)=0, \quad \text{for each } j=1,\ldots,n \end{cases}, \tag{2.10}$$

which can be solved numerically for α_0 and $\boldsymbol{\alpha}$ in order to obtain $\hat{\alpha}_0$ and $\hat{\boldsymbol{\alpha}}$.

Furthermore, we may want to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, as well as $(\alpha_{0j})_{j=1}^{n}$ using marginal (univariate) estimation techniques, and to estimate α_0 with the help of (2.9) thereafter. In such a case, with $(\hat{\alpha}_{0j})_{j=1}^{n}$ denoting the estimates of $(\alpha_{0j})_{j=1}^{n}$, the only equation to solve is

$$\frac{u_0}{\alpha_0}-\sum_{i=1}^{m}\ln\left(1+z_{(n)i}\right)+\sum_{j=1}^{n}\left(\sum_{i=1}^{m}\ln\left(1+\frac{x_{j,i}-\hat{\mu}_j}{\hat{\sigma}_j}\right)-\frac{u_j}{\hat{\alpha}_{0j}-\alpha_0}\right)=0. \tag{2.11}$$

In the next subsection we discuss an alternative method, which allows to estimate the parameters of interest simultaneously. To this end, we note in passing that in the context of the map $\mathbf{X}=T(\mathbf{Y})$, the random vector $\mathbf{Y}\in\mathbf{R}^{n+1}$ is a latent variable, and only $\mathbf{X}\in\mathbf{R}^n$ is practically observed. This interpretation, as well as the unimodal nature of the multivariate Pareto distribution considered herein, strongly hint at the appropriateness of the method. In the rest of the paper, we keep our discussion restricted to the bivariate and trivariate cases to make the exposition simple and to circumvent notational inconveniences inevitably arising when the general case is considered.

2.3. **Expectation maximization algorithm.** The general version of the EM algorithm was described and analyzed by Dempster et al. (1977) (see, also Wu, 1983) as an alternative to solving complex MLE problems. Karlis (2003) realized the method in the context of the multivariate exponential distribution of Marshall and Olkin, which is an example of a multivariate common shock model with exponential margins. As the multivariate Pareto distribution of interest clearly allows for a missing data interpretation, the utilization of the EM algorithm is quite natural.

To start off, we readily note that in the bivariate case, the missing data is represented by the latent random vector $\mathbf{Y} = (Y_0,\ Y_1,\ Y_2)'$, whereas the random vector $\mathbf{X} = T(\mathbf{Y}) = (X_1,\ X_2)' \backsim Pa_2(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$ denotes the practically observable variables of interest. Consequently, the complete data is in that case a five dimensional random vector possessing the p.d.f. $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y};\ \boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)' \in \boldsymbol{\Theta} \subset \mathbf{R}^7$. The EM algorithm then iteratively improves the initial estimate $\boldsymbol{\theta}^{(0)}$ by constructing new estimates $\boldsymbol{\theta}^{(k+1)},\ k \in \mathbf{N}$, that do not decrease the complete data expected analogue of (2.9).

More specifically, denoting, as before, by $(\mathbf{X}_i)_{i=1}^m$ and $\mathbf{x}_i = (x_{1,i}, x_{2,i})'$, the m independent copies of $\mathbf{X} \backsim Pa_2(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$ and the realization of say $\mathbf{X}_i$, respectively, the complete data expected log-likelihood is naturally formulated as

$$\begin{aligned} Q\left(\boldsymbol{\theta}; \mathbf{x}_1, \ldots, \mathbf{x}_m, \boldsymbol{\theta}^{(k)}\right) &= \mathbf{E}\left[\ln \prod_{i=1}^{m} f_{\mathbf{X}_i,\mathbf{Y}}(\mathbf{X}_i,\ \mathbf{Y};\ \boldsymbol{\theta}) \mid \mathbf{X}_i = \mathbf{x}_i,\ \boldsymbol{\theta}^{(k)}\right] \\ &= \int_{\mathbf{R}^3} \sum_{i=1}^{m} \ln f_{\mathbf{X}_i,\mathbf{Y}}(\mathbf{x}_i, \mathbf{y}; \boldsymbol{\theta})\, f_{\mathbf{Y}|\mathbf{X}_i=\mathbf{x}_i,\ \boldsymbol{\theta}^{(k)}}(\mathbf{y})\, d\mathbf{y}, \qquad (2.12) \end{aligned}$$

and it must be calculated for every $k = 0,\ 1, \ldots$. The estimate $\boldsymbol{\theta}^{(k+1)}$ is then determined as the expected log-likelihood maximizer, i.e., given $\boldsymbol{\theta}^{(k)}$, we have that

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Q\left(\boldsymbol{\theta}; \mathbf{x}_1, \ldots, \mathbf{x}_m, \boldsymbol{\theta}^{(k)}\right). \tag{2.13}$$

We note in passing that it is well-accepted to refer to (2.12) and (2.13) as the 'E' and 'M' steps, respectively. The two steps are repeated until a convergence criterion has been achieved.

In view of the recurrent nature of the EM algorithm, the $k = 0$ case, that corresponds to the initial expected log-likelihood $Q\left(\boldsymbol{\theta}; \mathbf{x}_1, \ldots, \mathbf{x}_m, \boldsymbol{\theta}^{(0)}\right)$, requires a somewhat special treatment. Speaking literally, the problem boils down to determining the starting value $\boldsymbol{\theta}^{(0)}$ to then allow for the evaluation of the consequent estimates. In this respect we suggest to utilize the empirical variates of the appropriate mean, variance, Pearson's correlation, as well as of Spearman's and/or Kendall's coefficients of association to trigger the method of moments (MM) estimation technique. We can thereby obtain, e.g., the estimates of α_0,

$$\begin{aligned}
\hat{\alpha}_0 &= \frac{c\left(\hat{\alpha}_{01} + \hat{\alpha}_{02} - 2\right)}{1+c},\ \text{with } c = \widehat{\mathbf{Corr}}[X_1, X_2]\sqrt{\frac{\hat{\alpha}_{01}\hat{\alpha}_{02}}{\left(\hat{\alpha}_{01} - 2\right)\left(\hat{\alpha}_{02} - 2\right)}},\\
\hat{\alpha}_{0\rho} &= \frac{2\left(\hat{\alpha}_{01} + \hat{\alpha}_{02}\right)\hat{\rho}[X_1, X_2]}{3 + \hat{\rho}[X_1, X_2]},\ \text{and } \hat{\alpha}_{0\tau} = \frac{\left(\hat{\alpha}_{01} + \hat{\alpha}_{02}\right)\hat{\tau}[X_1, X_2]}{1 + \hat{\tau}[X_1, X_2]},
\end{aligned}$$

using the empirical values of Pearson's correlation, Spearman's ρ and Kendall's τ, respectively. In the similar fashion, the entire $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)'$ can be found using appropriate MM equations, and it is thereafter used as $\boldsymbol{\theta}^{(0)}$ to start with the EM algorithm.

Explicit expressions for (2.12) are generally rarely derivable. In the context of the multivariate Pareto distribution of interest, the derivation is however possible with an effort.

Lemma 2.2. *Let* $\mathbf{X} \backsim Pa_2(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$ *be the observable random vector, and let* $Y_0 \backsim Pa(0, 1, \alpha_0)$ *and* $Y_j \backsim Pa(\mu_j, \sigma_j, \alpha_j),\ j = 1,\ 2$ *denote the latent variables. In addition, let*

$z(x_j)=(x_j-\mu_j)/\sigma_j$. The conditional p.d.f. of $\mathbf{Y}=(Y_0,Y_1,Y_2)'$ on $\mathbf{X}=(X_1,X_2)'$ is

$$f_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{y}\,|\mathbf{x}\right)=\begin{cases}\dfrac{\alpha_0\alpha_i}{\alpha_{0i}}\left(1+y_0\right)^{-\alpha_0-1}\left(1+z(x_i)\right)^{\alpha_0}, & y_j=x_j, y_i=x_i \text{ and } y_0>z(x_i)\\ \dfrac{\alpha_0\alpha_i}{\sigma_i\alpha_{0i}}\left(1+y_0\right)^{\alpha_i}\left(1+\dfrac{y_i-\mu_i}{\sigma_i}\right)^{-\alpha_i-1}, & y_j=x_j, y_i>x_i \text{ and } y_0=z(x_i)\end{cases},$$

for $z(x_i)>z(x_j)$, $i\neq j\in\{1,2\}$, and

$$f_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{y}\,|\mathbf{x}\right)=\frac{\alpha_1\alpha_2}{\sigma_1\sigma_2}\left(1+y_0\right)^{\alpha_1+\alpha_2}\left(1+\frac{y_1-\mu_1}{\sigma_1}\right)^{-\alpha_1-1}\left(1+\frac{y_2-\mu_2}{\sigma_2}\right)^{-\alpha_2-1},$$

for $z(x_1)=z(x_2)=y_0<\min\left\{(y_1-\mu_1)/\sigma_1,(y_2-\mu_2)/\sigma_2\right\}$.

We further employ Lemma 2.2 to estimate the parameters of the multivariate Pareto distribution of interest. To this end, for $i=1,\ldots,m$, let us redenote by $\mathbf{x}_i=(x_{1,i},x_{2,i})'$ a realization of the bivariate Pareto of the second kind random vector $\mathbf{X}_i$, which is an independent copy of $\mathbf{X}\backsim Pa_2(\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha},\alpha_0)$, and let $(\mathbf{X}_i)_{i=1}^m$ be a sequences of such copies. As we have already noted, the estimation of the parameters of Pa_2 is not indeed trivial. The EM algorithm with its time consuming M-step does not contribute to the tractability, either. Therefore, we suggest an adapted variant of the algorithm to estimate the vector $\boldsymbol{\theta}^*=(\boldsymbol{\alpha},\alpha_0)'$. We note in passing that the obvious estimate for $\boldsymbol{\mu}=(\mu_1,\mu_2)'$ is $\hat{\boldsymbol{\mu}}=(\hat{\mu}_1,\hat{\mu}_2)'$, such that $\hat{\mu}_j=\min_{i=1,\ldots,m}x_{j,i},\ j=1,\ 2$, and we estimate the vector $\boldsymbol{\sigma}=(\sigma_1,\sigma_2)'$ separately employing marginal (univariate) estimation techniques.

Namely, the $(k+1)$-th step of the adapted EM algorithm utilized in the sequel is

E step - Evaluate $Q\left(\boldsymbol{\alpha},\alpha_0;\mathbf{x}_1,\ldots,\mathbf{x}_m,\boldsymbol{\alpha}^{(k)},\alpha_0^{(k)}\right)$ using identity (2.12) and Lemma 2.2.

M1 step - Obtain the maximum likelihood estimates $\boldsymbol{\alpha}^{(k+1)},\alpha_0^{(k+1)}$ of $\boldsymbol{\alpha}$ and α_0 à la (2.13).

M2 step - Use the estimates from the $M1$-step above to update the marginal maximum likelihood estimate $\boldsymbol{\sigma}^{(k+1)}=(\sigma_1^{(k+1)},\sigma_2^{(k+1)})'$ of $\boldsymbol{\sigma}=(\sigma_1,\sigma_2)'$.

The aforementioned three steps are repeated until a convergence criterion has been reached.

Recall that, for $j = 1,\ 2$ and $i = 1, \ldots, m$, we have that $X_{j,i} = \min\left(\sigma_j Y_{0,i} + \mu_j, Y_{j,i}\right)$. To facilitate the exposition of the main result herein, let $z_{j,i}^{(k)} = \left(x_{j,i} - \hat{\mu}_j\right)/\sigma_j^{(k)}$ and also $z_{(2)i}^{(k)} = \max_j z_{j,i}^{(k)}$. In addition, denote by $w_0^{(k)},\ w_1^{(k)}$ and $w_2^{(k)}$ the cardinalities of the sets $S_0^{(k)} = \left\{i : z_{1,i}^{(k)} = z_{2,i}^{(k)}\right\}$, $S_1^{(k)} = \left\{i : z_{1,i}^{(k)} > z_{2,i}^{(k)}\right\}$ and $S_2^{(k)} = \left\{i : z_{1,i}^{(k)} < z_{2,i}^{(k)}\right\}$, respectively.

Theorem 2.2. *The expected log-likelihood for the* $(k+1)$*-th,* $k \in \mathbf{N}$*, step is*

$$
\begin{aligned}
& Q(\boldsymbol{\alpha}, \alpha_0; \mathbf{x}_1, \ldots, \mathbf{x}_m, \boldsymbol{\alpha}^{(k)}, \alpha_0^{(k)}) \\
\propto\ & m \ln\left(\alpha_0 \alpha_1 \alpha_2\right) - \alpha_0 \left(\sum_{i=1}^{m} \ln\left(1 + z_{(2)i}^{(k)}\right) + \frac{\alpha_2^{(k)} w_2^{(k)}}{\alpha_0^{(k)} \alpha_{02}^{(k)}} + \frac{\alpha_1^{(k)} w_1^{(k)}}{\alpha_0^{(k)} \alpha_{01}^{(k)}} \right) \\
-\ & \sum_{j=1}^{2} \alpha_j \left(\sum_{i=1}^{m} \ln\left(1 + z_{j,i}^{(k)}\right) + \frac{\alpha_0^{(k)} w_j^{(k)}}{\alpha_j^{(k)} \alpha_{0j}^{(k)}} + \frac{w_0^{(k)}}{\alpha_j^{(k)}} \right). \qquad (2.14)
\end{aligned}
$$

Theorem 2.2 clearly establishes the E step of the adapted EM algorithm. Thereby, the next statement follows straightforwardly, and it establishes the consequent $M1$ step.

Corollary 2.3. *The* $(k+1)$*-th,* $k \in \mathbf{N}$*, step estimates of the coordinates of* $\boldsymbol{\theta}^* = (\boldsymbol{\alpha}, \alpha_0)'$ *are conveniently obtained as*

$$
\begin{aligned}
\alpha_0^{(k+1)} &= m \left(\sum_{i=1}^{m} \ln\left(1 + z_{(2)i}^{(k)}\right) + \frac{\alpha_1^{(k)} w_1^{(k)}}{\alpha_0^{(k)} \alpha_{01}^{(k)}} + \frac{\alpha_2^{(k)} w_2^{(k)}}{\alpha_0^{(k)} \alpha_{02}^{(k)}} \right)^{-1}, \\
\alpha_1^{(k+1)} &= m \left(\sum_{i=1}^{m} \ln\left(1 + z_{1,i}^{(k)}\right) + \frac{\alpha_0^{(k)} w_1^{(k)}}{\alpha_1^{(k)} \alpha_{01}^{(k)}} + \frac{w_0^{(k)}}{\alpha_1^{(k)}} \right)^{-1}, \ \textit{and} \\
\alpha_2^{(k+1)} &= m \left(\sum_{i=1}^{m} \ln\left(1 + z_{2,i}^{(k)}\right) + \frac{\alpha_0^{(k)} w_2^{(k)}}{\alpha_2^{(k)} \alpha_{02}^{(k)}} + \frac{w_0^{(k)}}{\alpha_2^{(k)}} \right)^{-1}.
\end{aligned}
$$

At last, to establish the $M2$ step of the adapted EM algorithm, the system

$$
\sigma_j^{(k+1)} \sum_{i=1}^{m} \left(\sigma_j^{(k+1)} + x_{j,i} - \hat{\mu}_j\right)^{-1} = \frac{m \alpha_{0j}^{(k+1)}}{\alpha_{0j}^{(k+1)} + 1}, \ j = 1, 2, \qquad (2.15)
$$

is solved numerically for $\sigma_j^{(k+1)}$ employing $\alpha_{0j}^{(k+1)},\ j = 1,\ 2$, evaluated at the $M1$ step.

Note 2.1. *Noticeably, Corollary 2.3 extends to the trivariate case in a similar fashion. Namely, let* $\{j,l,q\} = \{1,2,3\}$ *and denote by* $w_j^{(k)}, w_{jl}^{(k)}$ *and* $w_0^{(k)}$ *the cardinalities of the sets* $S_j^{(k)} = \left\{i : z_{j,i}^{(k)} > \max\{z_{l,i}^{(k)}, z_{q,i}^{(k)}\}\right\}, S_{jl}^{(k)} = \left\{i : z_{j,i}^{(k)} = z_{l,i}^{(k)} > z_{q,i}^{(k)}\right\}$ *and* $S_0^{(k)} = \left\{i : z_{1,i}^{(k)} = z_{2,i}^{(k)} = z_{3,i}^{(k)}\right\}$, *respectively. Then the* $(k+1)$*-th step estimates of the coordinates of* $\boldsymbol{\theta}^* = (\boldsymbol{\alpha}, \alpha_0)'$ *arise as*

$$
\begin{aligned}
\alpha_0^{(k+1)} &= m\left(\sum_{i=1}^{m} \ln\left(1 + z_{(3)i}^{(k)}\right) + \frac{1}{\alpha_0^{(k)}} \sum_{j=1}^{3} \frac{\alpha_j^{(k)} w_j^{(k)}}{\alpha_{0j}^{(k)}}\right)^{-1}, \\
\alpha_j^{(k+1)} &= m\left(\sum_{i=1}^{m} \ln\left(1 + z_{j,i}^{(k)}\right) + \frac{1}{\alpha_j^{(k)}} \left(\frac{\alpha_0^{(k)} w_j^{(k)}}{\alpha_{0j}^{(k)}} + w_{jl}^{(k)} + w_{jq}^{(k)} + w_0^{(k)}\right)\right)^{-1},
\end{aligned}
$$

where $j = 1,2,3$.

After our derivations herein had been accomplished, we found a work by Kundu and Dey (2009), in which yet somewhat different approach to the EM algorithm was followed in the context of a bivariate Weibull distribution. More specifically, in the aforementioned paper, the orderings among the coordinates of $\mathbf{Y} = (Y_0, Y_1, Y_2)'$, rather than the coordinates themselves, were considered a missing information. The two results are however isomorphic.

3. A numerical illustration

3.1. **Bivariate case.** To exemplify and compare the various estimation methods presented above, we have generated bivariate Pareto random vectors $\mathbf{X}_i \backsim Pa_2(\boldsymbol{\mu} = (1,\ 2)', \boldsymbol{\sigma} = (2,\ 3)', \boldsymbol{\alpha} = (2,\ 2)', \alpha_0 = 2)$, $i = 1, \ldots, m = 5000$, and we have then applied the multivariate MLE (see, Subsection 2.2), as well as the MM and the adapted EM (see, Subsection 2.3) methods to estimate $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\alpha}$ and α_0. To rank the various estimation techniques, we have used the well-known Pearson's χ^2 test (see, Greenwood and Nikulin, 1996; for details).

As expected in our case, the χ^2 test at a significance level of 0.01 has rejected the Pearson's correlation-based MM method, only. This is not surprising, bearing in mind the expression for the conditional expectation, derived in Corollary 3.1 of Asimit et al. (2010), i.e., the linear correlation is not a good measure of dependence for the multivariate Pareto distribution of interest. To rank among other estimation techniques, we have employed the χ^2 values (see, Table 1; the lower - the better). Other entries of the table are briefly explained below.

The obvious estimates of μ_1 and μ_2 have been obtained as $\hat{\mu}_1 = 1.0000029$ and $\hat{\mu}_2 = 2.0002214$, and these values have been used in all three estimation techniques. The estimates of σ_1 and σ_2 have been developed as solutions of the marginal MLE system of equations – in the context of the MM and the multivariate MLE methods, and as iteratively updated solutions of the marginal MLE system – in the context of the adapted EM algorithm. The advantage of the latter technique is reflected in the corresponding χ^2 values. Similarly, the adapted EM algorithm seems to have outperformed the MM and the multivariate MLE when estimating $\alpha_0,\ \alpha_1$ and α_2.

Next in order to verify the perfomance of the proposed EM moethod as opposed to sample size, we have kept $\mu_1 = \mu_2 = 1,\ \sigma_1 = \sigma_2 = 1$ and $\alpha_1 = \alpha_2 = \alpha_0 = 2$ fixed and let the sample size vary. Also, as the starting values did not influence the Average Estimates (AEs), we have further fixed $\sigma_1 = \sigma_2 = \alpha_1 = \alpha_2 = \alpha_0 = 1.5$ as initial values. The stopping criterion hinged on the difference between parameters' consequetive values (stop if less than 10^{-6} in absolute value). Based on 100 replications, we thereby obtained the AE and the mean squared error (MSE) for each parameter, as well as the average number of iterations (AI), required. The results are depicted in Table 2, and we notice that the estimation naturaly improves with sample size.

Parameters	MM based on			Multivariate MLE	Adapted EM
	Pearson's **Corr**	Spearman's ρ	Kendall's τ		
$\hat{\alpha}_0$	2.18480	2.09250	2.08020	2.03579	2.04602
$\hat{\alpha}_1$	1.80185	1.89415	1.90645	1.95086	1.95178
$\hat{\alpha}_2$	1.99119	2.08349	2.09579	2.14020	2.04601
$\hat{\sigma}_1$	2.01025				2.01722
$\hat{\sigma}_2$	3.12215				3.04672
$\hat{\mu}_1$	1.0000029				
$\hat{\mu}_2$	2.0002214				
χ^2	27.9646	17.6281	17.1714	17.2520	14.2909

Table 1. Estimated parameters for the simulated bivariate Pareto $\mathbf{X} \backsim Pa_2(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$, with $\boldsymbol{\mu} = (1,\ 2)',\ \boldsymbol{\sigma} = (2,\ 3)',\ \boldsymbol{\alpha} = (2,\ 2)',\ \alpha_0 = 2.$

To elucidate the influence of the dependence on the outputs of the adapted EM algorithm, we let the values of α_0 vary, and set $\mu_1 = \mu_2 = 1,\ \sigma_1 = \sigma_2 = 1$ and $\alpha_1 = \alpha_2 = 2.$ In this respect, Table 3 seems to imply that the weaker the dependence is, the more effective the EM algorithm becomes.

3.2. **Trivariate case.** To conclude, we have also applied the EM method in the trivariate case with varying sample sizes and fixed $\mu_1 = \mu_2 = \mu_3 = 1,\ \sigma_1 = \sigma_2 = \sigma_3 = 1$ and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_0 = 2.$ The outcomes are depicted in Table 4, and they are comparable with these in Table 2.

	$m=1000$,AI$=1697$		$m=2000$,AI$=1651$		$m=3000$,AI$=1568$		$m=4000$,AI$=1515$	
	AE	MSE	AE	MSE	AE	MSE	AE	MSE
μ_1	1.0001	6×10^{-8}	1.0001	2×10^{-8}	1.00006	10^{-8}	1.00005	10^{-9}
μ_2	1.0002	9×10^{-8}	1.0001	2×10^{-8}	1.00006	10^{-8}	1.00005	10^{-8}
σ_1	1.0010	0.0190	1.0119	0.0166	1.0089	0.0120	1.0017	0.0047
σ_2	1.0141	0.0230	1.0144	0.0214	1.0171	0.0095	1.0014	0.0069
α_1	2.0178	0.1060	2.1578	0.1643	2.0870	0.0772	2.0578	0.0536
α_2	2.2152	0.3095	2.0327	0.1836	2.1368	0.0831	2.0439	0.0420
α_0	1.9557	0.0836	1.9424	0.0373	1.9682	0.0239	1.9411	0.0205

Table 2. The AE, MSE and AI indexes for the adapted EM method with varying sample size and $\mathbf{X} \backsim Pa_2(\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha},\alpha_0)$, where $\boldsymbol{\mu}=(1,\ 1)'$, $\boldsymbol{\sigma}=(1,\ 1)'$, $\boldsymbol{\alpha}=(2,\ 2)'$, $\alpha_0=2$.

.

	$\alpha_0 = 1$,AI= 686		$\alpha_0 = 1.5$,AI= 1047		$\alpha_0 = 2$,AI= 1515		$\alpha_0 = 2.5$,AI= 2122	
	AE	MSE	AE	MSE	AE	MSE	AE	MSE
μ_1	1.00008	1×10^{-8}	1.0001	2×10^{-8}	1.00006	10^{-8}	1.00004	10^{-9}
μ_2	1.00007	1×10^{-8}	1.0001	2×10^{-8}	1.00006	10^{-8}	1.00006	10^{-8}
σ_1	1.0039	0.0046	0.9916	0.0052	1.0017	0.0047	1.0219	0.0098
σ_2	0.9994	0.0063	1.0195	0.0101	1.0014	0.0069	1.0029	0.0130
α_1	2.0342	0.0247	2.0083	0.0276	2.0578	0.0536	2.0228	0.0543
α_2	2.0329	0.0340	2.0119	0.0499	2.0912	0.0420	2.0912	0.0987
α_0	0.9803	0.0041	1.4533	0.0093	1.9411	0.0205	2.4223	0.0404

TABLE 3. The AE, MSE and AI indexes for the adapted EM method with varying α_0 and $\mathbf{X} \backsim Pa_2(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$, where $\boldsymbol{\mu} = (1,\ 1)'$, $\boldsymbol{\sigma} = (1,\ 1)'$, $\boldsymbol{\alpha} = (2,\ 2)'$ and $m = 4000$.

.

	$m = 1000$,AI= 1789		$m = 2000$,AI= 1642		$m = 3000$,AI= 1618		$m = 3500$,AI= 1541	
	AE	MSE	AE	MSE	AE	MSE	AE	MSE
μ_1	1.0002	7×10^{-8}	1.0001	10^{-8}	1.00005	10^{-8}	1.00007	10^{-8}
μ_2	1.0002	9×10^{-8}	1.0001	2×10^{-8}	1.00007	10^{-8}	1.00007	10^{-8}
μ_3	1.0002	8×10^{-8}	1.0001	3×10^{-8}	1.00006	10^{-8}	1.00004	10^{-8}
σ_1	0.9749	0.0298	1.0019	0.0212	1.0174	0.0072	1.0023	0.0040
σ_2	1.0205	0.0389	0.9941	0.0181	1.0070	0.0136	0.9944	0.0139
σ_3	1.0057	0.0339	1.0122	0.0149	1.0132	0.0113	0.9989	0.0043
α_1	2.0531	0.2025	2.1319	0.1773	2.0590	0.0517	2.0284	0.0501
α_2	2.2112	0.3228	2.1021	0.1415	2.1211	0.1074	1.9915	0.0600
α_3	2.1902	0.2850	1.9654	0.1055	2.1401	0.1006	2.0716	0.0625
α_0	1.8670	0.0611	1.8740	0.0367	1.9148	0.0284	1.8910	0.0280

Table 4. The AE, MSE and AI indexes for the adapted EM method with varying sample sizes and $\mathbf{X} \backsim Pa_3(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \alpha_0)$, where $\boldsymbol{\mu} = (1,\ 1,\ 1)'$, $\boldsymbol{\sigma} = (1,\ 1,\ 1)'$, $\boldsymbol{\alpha} = (2,\ 2,\ 2)'$ and $\alpha_0 = 2$

.

Acknowledgment

The authors gratefully acknowledge the two anonymous referees for a number of insightful questions and suggestions. Ed Furman appreciates the financial support of his works by the Natural Science and Engineering Research Council of Canada (NSERC).

References

Arnold, B. C., 1983. *Pareto distributions.* International Cooperative Publ. House, Fairland, MD.

Asimit, A. V., Furman, E. and Vernic, R., 2010. On a multivariate Pareto distribution. *Insurance: Mathematics and Economics* **46** (2), 308 – 316.

Dempster, A. P., Laird, M. and Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B* **39**(1), 1 – 38.

Greenwood, P. E. and Nikulin, M. S., 1996. *A guide to chi-squared testing.* Wiley, New York.

Hanagal, D. D., 1996. A multivariate Pareto distribution. *Communications in Statistics - Theory and Methods* **25**, 1471 – 1488.

Karlis, D., 2003. ML estimation for multivariate shock models via an EM algorithm. *Ann. Inst. Statist. Math.* **55**(4), 817 – 830.

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

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

Nelsen, R. B., 1999. *An Introduction to Copulas.* Springer-Verlag, New York.

Proschan, F. and Sullo, P., 1976. Estimating the parameters of a multivariate exponential distribution. *Journal of the American Statistical Association* **71**(354), 465 – 472.

Wu, C. F. Jeff, 1983. On the convergence properties of the EM algorithm. *Annals of Statistics* **11**(1), 95 – 103.

4. Appendix

Proof of Theorem 2.1. According to (2.7), we readily have that, for any $\mathcal{C}$ in the Borel σ-algebra $\mathcal{B}^n$ in $\mathbf{R}^n$,

$$\mathbf{P}[\{\mathbf{X} \in \mathcal{C}\}] = \int_{\mathcal{C}} f_{\mathbf{X}} d\nu = \int_{\mathcal{C}} f_{\mathbf{X}} d\nu_n + \sum_{r=2}^{n} \int_{\mathcal{C}} f_{\mathbf{X}} d\nu_{n-r+1}.$$

Therefore, to treat the right most side, and for each $r = 2, \ldots, n$ and $\mathcal{T}_r = \mathcal{I}_n \backslash \mathcal{I}_r$, we start with the probability

$$\begin{aligned}
& \mathbf{P}\left[\{\mathbf{X} > \mathbf{x}\} \bigcap \left\{\frac{X_{i_1} - \mu_{i_1}}{\sigma_{i_1}} = \cdots = \frac{X_{i_r} - \mu_{i_r}}{\sigma_{i_r}}\right\}\right] \\
\stackrel{1}{=} \; & \mathbf{P}\left[\bigcap_{l \in \mathcal{T}_r} \{Y_l > x_l\} \bigcap_{j=1}^{r} \left\{Y_0 \le \frac{Y_{i_j} - \mu_{i_j}}{\sigma_{i_j}}\right\} \bigcap \{Y_0 > z_{(n)}\}\right] \\
\stackrel{2}{=} \; & \left(\prod_{l \in \mathcal{T}_r} \overline{F}_{Y_l}(x_l)\right) \int_{z_{(n)}}^{\infty} f_{Y_0}(y_0) \left(\prod_{j=1}^{r} \int_{\sigma_{i_j} y_0 + \mu_{i_j}}^{\infty} f_{Y_{i_j}}(y_j)\, dy_j\right) dy_0 \\
= \; & \left(\prod_{l \in \mathcal{T}_r} \overline{F}_{Y_l}(x_l)\right) \int_{z_{(n)}}^{\infty} \alpha_0 (1 + y_0)^{-\left(\alpha_0 + \sum_{j=1}^{r} \alpha_{i_j} + 1\right)} dy_0 \\
= \; & \frac{\alpha_0}{\alpha_0 + \sum_{j=1}^{r} \alpha_{i_j}} \left(1 + z_{(n)}\right)^{-\left(\alpha_0 + \sum_{j=1}^{r} \alpha_{i_j}\right)} \prod_{l \in \mathcal{T}_r} \left(1 + \frac{x_l - \mu_l}{\sigma_l}\right)^{-\alpha_l},
\end{aligned}$$

where $\stackrel{1}{=}$ is because, for $\sigma_j \in \mathbf{R}_+$, $\mu_j \in \mathbf{R}$ and $j = 1, \ldots, n$ we have that

$$\frac{\min(\sigma_j Y_0 + \mu_j, Y_j) - \mu_j}{\sigma_j} = \min\left(Y_0, \frac{Y_j - \mu_j}{\sigma_j}\right),$$

and $\stackrel{2}{=}$ holds by independence. The corresponding density is then obtained (recall the dimension of the measure ν_{n-r+1}) by differentiating with respect to $z_{(n)}$, as well as with respect to each x_l with $l \in \mathcal{T}_r$, as

$$f(x_1,\ldots,x_n)=\alpha_0\left(1+z_{(n)}\right)^{-\left(\alpha_0+\sum_{j=1}^{r}\alpha_{i_j}+1\right)}\prod_{l\in\mathcal{T}_r}\frac{\alpha_l}{\sigma_l}\left(1+\frac{x_l-\mu_l}{\sigma_l}\right)^{-(\alpha_l+1)},$$

which along with (2.5) and keeping in mind (2.7) and the Radon-Nikodym theorem completes the proof. □

Proof of Lemma 2.2. It is clear that the d.d.f. of the complete data random vector is conveniently written as

$$\begin{aligned}&\mathbf{P}[\{\mathbf{Y}>\mathbf{y}\}\cap\{\mathbf{X}>\mathbf{x}\}]\\ =\ &\mathbf{P}[Y_0>\max\left(z(x_1),z(x_2),y_0\right)]\mathbf{P}[Y_1>\max(x_1,y_1)]\mathbf{P}[Y_2>\max(x_2,y_2)],\end{aligned}$$

where $y_j>\mu_j$ and $x_j>\mu_j,\ j=1,\ 2$, as well as $y_0>0$. Consequently various orderings of $z(x_1),\ z(x_2)$ and y_0 must be treated separately. More specifically, utilizing (2.8), we readily have that, e.g., for $z(x_1)=z(x_2)=y_0$ and thus $y_0<\min\left(\frac{y_1-\mu_1}{\sigma_1},\frac{y_2-\mu_2}{\sigma_2}\right)$, the conditional p.d.f. of interest reduces to

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\ \mathbf{x})=f_{Y_0}(y_0)f_{Y_1}(y_1)f_{Y_2}(y_2)(f_{\mathbf{X}}(\mathbf{x}))^{-1},$$

as required.

A somewhat more tedious case is the one where $z(x_1)\neq z(x_2)$ and $y_0>\max(z(x_1),z(x_2))$, which implies that $x_j=y_j,\ j=1,\ 2$. Then the d.d.f. to be considered is

$$\begin{aligned}&\mathbf{P}\left[\{\mathbf{Y}>\mathbf{y}\}\cap\{\mathbf{X}>\mathbf{x}\}\cap\{X_1=Y_1\}\cap\{X_2=Y_2\}\cap\{Y_0>\max\{z(X_1),z(X_2)\}\}\right]\\ =\ &\mathbf{P}\left[\{Y_0>y_0\}\cap\{\sigma_1Y_0+\mu_1>Y_1>x_1\}\cap\{\sigma_2Y_0+\mu_2>Y_2>x_2\}\right]\\ =\ &\int_{y_0}^{\infty}f_{Y_0}(u_0)\left(\prod_{j=1}^{2}\int_{x_j}^{\sigma_ju_0+\mu_j}f_{Y_j}(u_j)\,du_j\right)du_0=\int_{y_0}^{\infty}f_{Y_0}(u_0)\prod_{j=1}^{2}\left(\overline{F}_{Y_j}(x_j)-\overline{F}_{Y_j}(\sigma_ju_0+\mu_j)\right)du_0\\ =\ &\overline{F}_{Y_0}(y_0)\,\overline{F}_{Y_1}(x_1)\,\overline{F}_{Y_2}(x_2)+H_0(y_0)+H_1(y_0,x_1)+H_2(y_0,x_2)\,,\end{aligned}$$

which thus yields

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\ \mathbf{x}) = f_{Y_0}(y_0) f_{Y_1}(x_1) f_{Y_2}(x_2) (f_{\mathbf{X}}(\mathbf{x}))^{-1},$$

and it in turn disintegrates as expected keeping in mind (2.8).

The proof of the remaining two expressions is knocked out in an entirely similar fashion, and it is thus left to the reader. This completes the proof of the lemma. □

Proof of Theorem 2.2. By definition and utilizing Lemma 2.2, we have that

$$\begin{aligned}
Q(\boldsymbol{\alpha}, \alpha_0; \mathbf{x}_1, \ldots, \mathbf{x}_m, \boldsymbol{\alpha}^{(k)}, \alpha_0^{(k)}) &= \mathbf{E}\left[\sum_{i=1}^{m} \ln f_{\mathbf{X}_i,\mathbf{Y}}\left(\mathbf{X}_i,\ \mathbf{Y};\ \boldsymbol{\sigma}^{(k)}, \boldsymbol{\alpha}, \alpha_0\right) \Big|\ \mathbf{X}_i = \mathbf{x}_i,\ \boldsymbol{\theta}^{(k)}\right] \\
= \ & \mathbf{E}\left[\left(\sum_{i=1}^{m} \ln\left(f_{Y_0}\left(Y_0\right) f_{Y_1}\left(X_{1,i}\right) f_{Y_2}\left(X_{2,i}\right)\right)\right) \mathbf{1}\left\{ {z_{1,i}^{(k)} \neq z_{2,i}^{(k)},\ Y_j = X_{j,i}, j=1,2 \atop Y_0 > \max\left\{z_{1,i}^{(k)}, z_{2,i}^{(k)}\right\}} \right\} \Big|\ \mathbf{X}_i = \mathbf{x}_i,\ \ \boldsymbol{\theta}^{(k)}\right] \\
+ \ & \mathbf{E}\left[\left(\sum_{i=1}^{m} \ln\left(f_{Y_0}\left(Y_0\right) \frac{f_{Y_1}\left(X_{1,i}\right)}{\sigma_2^{(k)}} f_{Y_2}\left(Y_2\right)\right)\right) \mathbf{1}\left\{ {Y_0 = z_{2,i}^{(k)} > z_{1,i}^{(k)} \atop Y_2 > X_{2,i},\ Y_1 = X_{1,i}} \right\} \Big|\ \mathbf{X}_i = \mathbf{x}_i,\ \ \boldsymbol{\theta}^{(k)}\right] \\
+ \ & \mathbf{E}\left[\left(\sum_{i=1}^{m} \ln\left(f_{Y_0}\left(Y_0\right) \frac{f_{Y_2}\left(X_{2,i}\right)}{\sigma_1^{(k)}} f_{Y_1}\left(Y_1\right)\right)\right) \mathbf{1}\left\{ {Y_0 = z_{1,i}^{(k)} > z_{2,i}^{(k)} \atop Y_1 > X_{1,i},\ Y_2 = X_{2,i}} \right\} \Big|\ \mathbf{X}_i = \mathbf{x}_i,\ \ \boldsymbol{\theta}^{(k)}\right] \\
+ \ & \mathbf{E}\left[\left(\sum_{i=1}^{m} \ln\left(f_{Y_0}\left(Y_0\right) f_{Y_1}\left(Y_1\right) f_{Y_2}\left(Y_2\right)\right)\right) \mathbf{1}\left\{ {Y_0 = z_{1,i}^{(k)} = z_{2,i}^{(k)} \atop Y_0 < \max\left\{\frac{Y_1 - \hat{\mu}_1}{\sigma_1^{(k)}}, \frac{Y_2 - \hat{\mu}_2}{\sigma_2^{(k)}}\right\}} \right\} \Big|\ \mathbf{X}_i = \mathbf{x}_i,\ \ \boldsymbol{\theta}^{(k)}\right].
\end{aligned}$$

Further, according to Lemma 2.2, three distinct cases must be considered. Namely, for, e.g., $z_{2,i}^{(k)} > z_{1,i}^{(k)}$, $i = 1, \dots, m$, we have that

$$
\begin{aligned}
& Q_1(\boldsymbol{\alpha}, \alpha_0; \mathbf{x}_1, \dots, \mathbf{x}_m, \boldsymbol{\alpha}^{(k)}, \alpha_0^{(k)}) \\
= \; & \mathbf{E}\left[\left(\sum_{i=1}^{m} \ln\left(f_{Y_0}\left(Y_0\right) f_{Y_1}\left(X_{1,i}\right) f_{Y_2}\left(X_{2,i}\right)\right)\right) \mathbf{1}\left\{\substack{Y_0 > z_{2,i}^{(k)} > z_{1,i}^{(k)}, \\ X_{j,i} = Y_j, j=1,\ 2}\right\} \,\Big|\, \mathbf{X}_i = \mathbf{x}_i,\ \boldsymbol{\theta}^{(k)}\right] \\
+ \; & \mathbf{E}\left[\left(\sum_{i=1}^{m} \ln\left(f_{Y_0}\left(Y_0\right) \frac{f_{Y_1}\left(X_{1,i}\right)}{\sigma_2^{(k)}} f_{Y_2}\left(Y_2\right)\right)\right) \mathbf{1}\left\{\substack{Y_0 = z_{2,i}^{(k)} > z_{1,i}^{(k)}, \\ Y_2 > X_{2,i}, X_{1,i} = Y_1}\right\} \,\Big|\, \mathbf{X}_i = \mathbf{x}_i,\ \boldsymbol{\theta}^{(k)}\right] \\
= \; & \sum_{\substack{i=1,\dots,m \\ z_{2,i}^{(k)} > z_{1,i}^{(k)}}} \int \ln\left(\alpha_0\left(1+y_0\right)^{-\alpha_0-1} \frac{\alpha_1\alpha_2}{\sigma_1^{(k)}\sigma_2^{(k)}}\left(1+z_{1,i}^{(k)}\right)^{-\alpha_1-1}\left(1+z_{2,i}^{(k)}\right)^{-\alpha_2-1}\right) \\
\times \; & \frac{\alpha_0^{(k)}\alpha_2^{(k)}}{\alpha_{02}^{(k)}}\left(1+z_{2,i}^{(k)}\right)^{\alpha_0^{(k)}}\left(1+y_0\right)^{-\alpha_0^{(k)}-1} \mathbf{1}\{y_0 > z_{2,i}^{(k)}\} dy_0 \\
+ \; & \sum_{\substack{i=1,\dots,m \\ z_{2,i}^{(k)} > z_{1,i}^{(k)}}} \int \ln\left(\frac{\alpha_0}{\sigma_2^{(k)}}\left(1+z_{2,i}^{(k)}\right)^{-\alpha_0-1} \frac{\alpha_1\alpha_2}{\sigma_1^{(k)}\sigma_2^{(k)}}\left(1+z_{1,i}^{(k)}\right)^{-\alpha_1-1}\left(1+\frac{y_2-\hat{\mu}_2}{\sigma_2^{(k)}}\right)^{-\alpha_2-1}\right) \\
\times \; & \frac{\alpha_0^{(k)}\alpha_2^{(k)}}{\sigma_2^{(k)}\alpha_{02}^{(k)}}\left(1+z_{2,i}^{(k)}\right)^{\alpha_2^{(k)}}\left(1+\frac{y_2-\hat{\mu}_2}{\sigma_2^{(k)}}\right)^{-\alpha_2^{(k)}-1} \mathbf{1}\{y_2 > x_{2,i}\} dy_2 \\
\propto \; & \sum_{\substack{i=1,\dots,m \\ z_{2,i}^{(k)} > z_{1,i}^{(k)}}} \left(\ln\left(\alpha_0\alpha_1\alpha_2\right) - \left(\alpha_0+\alpha_2\right)\ln\left(1+z_{2,i}^{(k)}\right) - \alpha_1\ln\left(1+z_{1,i}^{(k)}\right) - \frac{\alpha_0\alpha_2^{(k)}}{\alpha_0^{(k)}\alpha_{02}^{(k)}} - \frac{\alpha_2\alpha_0^{(k)}}{\alpha_2^{(k)}\alpha_{02}^{(k)}}\right),
\end{aligned}
$$

neglecting the terms that do not depend on α_0 and/or $\boldsymbol{\alpha}$.

A similar expression follows, e.g., by symmetry, for the opposite case, where $z_{2,i}^{(k)} < z_{1,i}^{(k)}$, i.e., we have that

$$
\begin{aligned}
& Q_2(\boldsymbol{\alpha}, \alpha_0; \mathbf{x}_1, \dots, \mathbf{x}_m, \boldsymbol{\alpha}^{(k)}, \alpha_0^{(k)}) \\
= \; & \sum_{\substack{i=1,\dots,m \\ z_{2,i}^{(k)} < z_{1,i}^{(k)}}} \left(\ln\left(\alpha_0\alpha_1\alpha_2\right) - \left(\alpha_0+\alpha_1\right)\ln\left(1+z_{1,i}^{(k)}\right) - \alpha_2\ln\left(1+z_{2,i}^{(k)}\right) - \frac{\alpha_0\alpha_1^{(k)}}{\alpha_0^{(k)}\alpha_{01}^{(k)}} - \frac{\alpha_1\alpha_0^{(k)}}{\alpha_1^{(k)}\alpha_{01}^{(k)}}\right).
\end{aligned}
$$

Finally, for $z_{2,i}^{(k)} = z_{1,i}^{(k)}$, we readily obtain that

$$\begin{aligned}
& Q_3(\boldsymbol{\alpha}, \alpha_0; \mathbf{x}_1, \ldots, \mathbf{x}_m, \boldsymbol{\alpha}^{(k)}, \alpha_0^{(k)}) \\
= \; & \mathbf{E}\left[\left(\sum_{i=1}^{m} \ln\left(f_{Y_0}\left(Y_0\right) f_{Y_1}\left(Y_1\right) f_{Y_2}\left(Y_2\right)\right) \right) \mathbf{1}\left\{ \substack{Y_0 = z_{1,i}^{(k)} = z_{2,i}^{(k)} \\ Y_0 < \max\left\{ \frac{Y_1 - \hat{\mu}_1}{\sigma_1^{(k)}}, \frac{Y_2 - \hat{\mu}_2}{\sigma_2^{(k)}} \right\}} \right\} \,\Big|\, \mathbf{X}_i = \mathbf{x}_i, \;\; \boldsymbol{\theta}^{(k)} \right] \\
\propto \; & \sum_{\substack{i=1,\ldots,m \\ z_{2,i}^{(k)} = z_{1,i}^{(k)}}} \left(\ln\left(\alpha_0 \alpha_1 \alpha_2\right) - \left(\alpha_0 + \alpha_1 + \alpha_2\right) \ln\left(1 + z_{(2)i}^{(k)}\right) - \frac{\alpha_1}{\alpha_1^{(k)}} - \frac{\alpha_2}{\alpha_2^{(k)}} \right).
\end{aligned}$$

The expected log - likelihood of interest then follows as

$$Q(\boldsymbol{\alpha}, \alpha_0; \mathbf{x}_1, \ldots, \mathbf{x}_m, \boldsymbol{\alpha}^{(k)}, \alpha_0^{(k)}) = \sum_{i=1}^{3} Q_i(\boldsymbol{\alpha}, \alpha_0; \mathbf{x}_1, \ldots, \mathbf{x}_m, \boldsymbol{\alpha}^{(k)}, \alpha_0^{(k)}),$$

which reduces to (2.14) as required and hence completes the proof. □