

Kent Academic Repository

Full text document (pdf)

Citation for published version

Alai, Daniel and Landsman, Zinoviy (2017) Lifetime dependence models generated by multiply monotone functions. *Scandinavian Actuarial Journal* . ISSN 0346-1238.

DOI

<https://doi.org/10.1080/03461238.2017.1405840>

Link to record in KAR

<http://kar.kent.ac.uk/64924/>

Document Version

Author's Accepted Manuscript

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

Enquiries

For any further enquiries regarding the licence status of this document, please contact:

researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

Lifetime Dependence Models Generated by Multiply Monotone Functions

Daniel H. Alai¹ Zinoviy Landsman²

School of Mathematics, Statistics and Actuarial Science
University of Kent, Canterbury, Kent CT2 7NZ, UK

Department of Statistics, Actuarial Research Center
University of Haifa, Mount Carmel, Haifa 31905, Israel

Abstract

We study a family of distributions generated from multiply monotone functions that includes a multivariate Pareto and, previously unidentified, exponential-Pareto distribution. We utilize an established link with Archimedean survival copulas to provide further examples, including a multivariate Weibull distribution, that may be used to fit light, or heavy-tailed phenomena, and which exhibit various forms of dependence, ranging from positive to negative. Because the model is intended for the study of joint lifetimes, we consider the effect of truncation and formulate properties required for a number of parameter estimation procedures based on moments and quantiles. For the quantile-based estimation procedure applied to the multivariate Weibull distribution, we also address the problem of optimal quantile selection.

Keywords: Multiply monotone functions; Archimedean survival copulas; Pareto distribution; Weibull distribution; Multivariate truncation

¹d.h.alai@kent.ac.uk

²landsman@stat.haifa.ac.il

1 Introduction

We consider constructions of multivariate distributions, introduced in Hendriks and Landsman (2017), generated from multiply monotone functions; see Williamson (1956). This family includes a multivariate Pareto distribution as a special case. The Pareto is an important and well-known distribution due to its ability to capture heavy-tailed phenomena; see e.g. Mardia (1962), Arnold (1985) and Embrechts *et al.* (1997). Furthermore, truncation is an inherent feature of the *peaks over threshold* method used to study extreme value distributions and relevant to statistical theory in general. More recently, the multivariate Pareto distribution has received considerable attention in finance and insurance; see e.g. Chiragiev and Landsman (2007), Asimit *et al.* (2010), Vernic (2011) and Alai *et al.* (2016). In the latter, the nature of the application required the multivariate distribution to be truncated, yielding some interesting theoretical results as well as estimation procedures.

Applications in actuarial science, for both life and general insurance, often require truncated distributions. It is well-known that for general insurance, claims are often subject to a deductible. This is an important feature of the data, which needs to be addressed appropriately by the claims severity distribution. For joint lifetime modelling, as in Alai *et al.* (2016), relevant data is conditional upon individuals surviving to a certain age; for example, the pension age. Public pension sustainability, the shift in employer-sponsored pension plans from defined benefit to defined contribution, and the buying-out of existing pensions via the purchase of bulk annuities are all very important consequences of changes in longevity that can be captured via an appropriate dependence structure between lives. With this application in mind, we establish tools that allow for a richer variety of dependence between lifetimes, whilst addressing the inherent truncation in the data. This is instrumental in the pricing and risk management of life-benefit products such as bulk annuities.

In this paper, we study the effect of both univariate and multivariate truncation on distributions generated from multiply monotone functions. We consider the impact of truncation on parameter estimation techniques that rely on moments and quantiles. Following Hendriks and Landsman (2017), let $\mathbf{X} = (X_1, \dots, X_n)$ be a multivariate random vector with strictly positive components $X_i > 0$ such that its joint survival function is given by

$$P(X_1 > x_1, \dots, X_n > x_n) = h\left(\sum_{i=1}^n \lambda_i x_i\right), \quad x_i \geq 0, \quad (1)$$

for $\lambda_i > 0, \forall i$, where h is d -times monotone, $d \geq n$. That is, for $k \in \{1, \dots, d\}$,

$$(-1)^k h^{(k)}(x) \geq 0, \quad x > 0. \quad (2)$$

In the special case that $\lambda_i = 1, \forall i$, we retrieve the Schur-constant survival function; see e.g. Caramellino and Spizzichino (1994) and Nelsen (2005). To ensure we have a survival function with a well-defined inverse, we assume $h(x)$ is strictly decreasing for $x \geq 0$. The joint density of any subset of \mathbf{X} exists and can be found by taking the derivative of the function h an appropriate number of times. For example, the density of \mathbf{X} is given by

$$f_{\mathbf{X}}(x_1, \dots, x_n) = (-1)^n \lambda_1 \cdots \lambda_n h^{(n)}\left(\sum_{i=1}^n \lambda_i x_i\right) \geq 0, \quad x_i > 0,$$

and the density of X_i is given by

$$f_i(x_i) = (-1)\lambda_i h^{(1)}(\lambda_i x_i) \geq 0, \quad x_i > 0.$$

The marginal survival function of X_i could be determined from the density above, or directly from the joint survival function as follows:

$$\begin{aligned} P(X_i > x_i) &= P(X_1 > 0, \dots, X_{i-1} > 0, X_i > x_i, X_{i+1} > 0, \dots, X_n > 0) \\ &= h(\lambda_i x_i), \quad x_i \geq 0. \end{aligned}$$

It is clear that the function h always behaves as a survival function, whether for the n -variate joint distribution, a univariate marginal distribution, or anything in between. As such, we require

$$h(0) = 1, \quad \lim_{x \rightarrow \infty} h(x) = 0. \quad (3)$$

Relationship to Archimedean Survival Copulas

The requirements on h are identical to those of Archimedean copula generators, and hence, the multivariate distribution we consider has intimate links with this family of copulas. In general, the literature on copulas is vast; they represent a very important tool in the dependence modelling arsenal; see e.g. Frees and Valdez (1998), Nelsen (1999), McNeil *et al.* (2005) and McNeil and Nešlehová (2009) as well as Genest and Rivest (1993), Genest *et al.* (1995) and Hofert *et al.* (2012) for more on estimation procedures.

In order to showcase the various forms of dependence that can be attained via the construction of Hendriks and Landsman (2017), which includes both positive and negative dependence, we demonstrate the effects of selecting well-known Archimedean copula generators to form distributions as given by Equation (1). In each case, we demonstrate the necessary steps to address truncation and provide explicit expressions to obtain parameter estimates.

In some detail, provided that h satisfies the conditions in Equations (2) and (3), it is the generator of an Archimedean survival copula. In Hendriks and Landsman (2017), it is shown that a multivariate distribution is identically specified either through the Archimedean survival copula with generator h whose marginal survival distributions also follow the functional form of h , or through the joint survival function as given by Equation (1).

Let the marginal distributions of the X_i have survival functions that are given by $P(X_i > x_i) = h(\lambda_i x_i) = p_i$ and apply Archimedean survival copula \widehat{C} with generator h ,

$$\widehat{C}(p_1, \dots, p_n) = h(h^{-1}(p_1) + \dots + h^{-1}(p_n)),$$

where h^{-1} is the inverse of h such that $h(h^{-1}(p)) = p$ and h is d -times monotone, $d \geq n$; see e.g. Nelsen (1999) and McNeil and Nešlehová (2009). Consequently, we have that

$$\begin{aligned} P(X_1 > x_1, \dots, X_n > x_n) &= \widehat{C}(P(X_1 > x_1), \dots, P(X_n > x_n)) \\ &= \widehat{C}(p_1, \dots, p_n) \\ &= h(h^{-1}(p_1) + \dots + h^{-1}(p_n)) \\ &= h(h^{-1}(h(\lambda_1 x_1)) + \dots + h^{-1}(h(\lambda_n x_n))) \\ &= h(\lambda_1 x_1 + \dots + \lambda_n x_n). \end{aligned}$$

The other direction is shown similarly, and hence the two constructions are equivalent.

Organization of the paper: In Section 2 we consider the truncated moments and quantiles of the univariate distribution generated by h . We address the effect of multivariate truncation on the moments of the marginal distributions as well the distribution of the minimum and maximum in Section 3. We consider some examples of multivariate distributions in Section 4, providing all the necessary details for moment and quantile estimation procedures. In Section 5 we consider the quantile-based estimation procedure and address the problem of selecting optimal quantile levels. Section 6 concludes the paper.

2 Univariate Truncation

In the following two sections, we derive some relevant properties of multivariate distributions specified via the joint survival function. We address, first, univariate truncation of the marginal distribution, followed by multivariate truncation.

2.1 Multiply Monotone Generator h

For $\mathbf{X} = (X_1, \dots, X_n)$ with strictly positive components $X_i > 0$, we have

$$P(X_1 > x_1, \dots, X_n > x_n) = h\left(\sum_{i=1}^n \lambda_i x_i\right), \quad x_i \geq 0,$$

$\lambda_i > 0, \forall i$, where h is d -times monotone, $d \geq n$. We refer to h as the multiply monotone generator of the multivariate distribution of \mathbf{X} . We presently introduce the antiderivatives of h , of which we make ample use below.

$$\begin{aligned} h^{(-1)}(x) &= -\int_x^\infty h(y)dy, \\ h^{(-2)}(x) &= -\int_x^\infty \left\{ -\int_y^\infty h(z)dz \right\} dy, \end{aligned}$$

and so on. Recursively, we have that

$$h^{(-k)}(x) = -\int_x^\infty h^{(-(k-1))}(y)dy.$$

We generally apply the antiderivatives with argument $\lambda x + c$ for constants λ and c ; substitution yields

$$h^{(-k)}(\lambda x + c) = \int_x^\infty (-1)\lambda h^{(-(k-1))}(\lambda y + c)dy.$$

2.2 Moments of the Truncated Marginal Distribution

Consider the truncation of X_i with truncation point τ , ${}_\tau X_i = X_i | X_i > \tau$. We refer to this as univariate truncation since nothing is claimed about the level of the remaining marginal distributions $X_j, j \neq i$; we address multivariate truncation in Section 3. We provide a useful result for the raw moments of X_i and ${}_\tau X_i$.

Theorem 1 Consider $\mathbf{X} = (X_1, \dots, X_n)$ with strictly positive components and marginal survival functions $h_i(\lambda_i x_i)$, $\lambda_i > 0, \forall i$, for d -times monotone h_i , $d \geq n$. If finite, the k^{th} raw moment, which we denote α_k , for $k \in \{1, 2, \dots, d\}$, of X_i and ${}_{\tau}X_i = X_i | X_i > \tau$, $\tau \geq 0$, are given by

$$\begin{aligned}\alpha_k(X_i) &= (-1)^k k! \frac{h_i^{(-k)}(0)}{\lambda_i^k}, \\ \alpha_k({}_{\tau}X_i) &= h_i(\lambda_i \tau)^{-1} \sum_{j=0}^k (-1)^j \tau^{k-j} \frac{k!}{(k-j)!} \frac{h_i^{(-j)}(\lambda_i \tau)}{\lambda_i^j},\end{aligned}$$

where $h_i^{(-k)}(x) = -\int_x^{\infty} h_i^{(-(k-1))}(y) dy$ and $h_i^{(0)}(x) = h_i(x)$.

Proof. We prove the result directly. In general, we assume integrals corresponding to finite moments are well-defined; for specific h_i , one can be more precise and relate the existence of moments to values of the canonical parameter θ .

$$\alpha_k({}_{\tau}X_i) = h_i(\lambda_i \tau)^{-1} \int_{\tau}^{\infty} x^k (-1) \lambda_i h_i^{(1)}(\lambda_i x) dx.$$

Apply integration by parts with $u = x^k$ and $dv = (-1) \lambda_i h_i^{(1)}(\lambda_i x) dx$; consequently, $du = kx^{k-1} dx$ and $v = -h_i^{(0)}(\lambda_i x)$.

$$\alpha_k({}_{\tau}X_i) = h_i(\lambda_i \tau)^{-1} \left\{ -x^k h_i^{(0)}(\lambda_i x) \Big|_{x=\tau}^{\infty} + \int_{\tau}^{\infty} kx^{k-1} h_i^{(0)}(\lambda_i x) dx \right\}.$$

We use the result provided in Lemma 1 in Williamson (1956); it states that for d -times monotone function f , $\lim_{x \rightarrow \infty} x^k f(x) = 0$ for $k \in \{1, \dots, d\}$. Consequently, we have that

$$\begin{aligned}\alpha_k({}_{\tau}X_i) &= h_i(\lambda_i \tau)^{-1} \left\{ \tau^k h_i^{(0)}(\lambda_i \tau) + \int_{\tau}^{\infty} kx^{k-1} h_i^{(0)}(\lambda_i x) dx \right\} \\ &= h_i(\lambda_i \tau)^{-1} \left\{ \tau^k h_i^{(0)}(\lambda_i \tau) + \frac{(-1)k}{\lambda_i} \int_{\tau}^{\infty} x^{k-1} (-1) \lambda_i h_i^{(0)}(\lambda_i x) dx \right\}.\end{aligned}$$

Notice that the remaining integral is exactly what we began with, only the power of x and the *level* of function h_i have each reduced by one. We reiterate the same procedure; apply the same integration by parts, noting that $h_i^{(-1)}(x) = -\int_x^{\infty} h_i^{(0)}(y) dy$ and therefore has the same relationship as $h_i^{(0)}$ with $h_i^{(1)}$; and finally, use the same result given in Williamson (1956), noting that $\lim_{x \rightarrow \infty} x^{k-1} h_i^{(-1)}(\lambda_i x) = 0$ since $h_i^{(-1)}$ is $(d+1)$ -times monotone. We have that

$$\begin{aligned}\alpha_k({}_{\tau}X_i) &= h_i(\lambda_i \tau)^{-1} \left\{ \tau^k h_i^{(0)}(\lambda_i \tau) \right. \\ &\quad \left. + \frac{(-1)k}{\lambda_i} \left\{ \tau^{k-1} h_i^{(-1)}(\lambda_i \tau) + \frac{(-1)(k-1)}{\lambda_i} \int_{\tau}^{\infty} x^{k-2} (-1) \lambda_i h_i^{(-1)}(\lambda_i x) dx \right\} \right\} \\ &= \dots = h_i(\lambda_i \tau)^{-1} \left\{ (-1)^0 \tau^k h_i^{(0)}(\lambda_i \tau) + (-1)^1 \tau^{k-1} k \frac{h_i^{(-1)}(\lambda_i \tau)}{\lambda_i} \right. \\ &\quad \left. + (-1)^2 \tau^{k-2} k(k-1) \frac{h_i^{(-2)}(\lambda_i \tau)}{\lambda_i^2} + \dots + (-1)^{(k-1)} \tau \frac{k!}{1!} \frac{h_i^{(-(k-1))}(\lambda_i \tau)}{\lambda_i^{(k-1)}} \right. \\ &\quad \left. + \frac{(-1)^k k!}{0! \lambda_i^k} \int_{\tau}^{\infty} (-1) \lambda_i h_i^{(-(k-1))}(\lambda_i x) dx \right\}.\end{aligned}$$

Finally, notice that the remaining integral, where the power of x has reduced to zero, is $h_i^{(-k)}(\lambda_i\tau)$; this produces the result for truncation, in order to show the result without truncation, set $\tau = 0$ and note that $h_i(0) = 1$. ■

As a result of Theorem 1, the mean and variance, denoted μ_2 , of ${}_{\tau}X_i$, if finite, are given by

$$\begin{aligned}\alpha_1({}_{\tau}X_i) &= \tau - \frac{h_i^{(-1)}(\lambda_i\tau)}{\lambda_i h_i(\lambda_i\tau)}, \\ \mu_2({}_{\tau}X_i) &= \frac{2h_i^{(-2)}(\lambda_i\tau)}{\lambda_i^2 h_i(\lambda_i\tau)} - \frac{h_i^{(-1)}(\lambda_i\tau)^2}{\lambda_i^2 h_i(\lambda_i\tau)^2}.\end{aligned}$$

If we let the truncation point be zero, we retrieve the expectation and variance found in Hendriks and Landsman (2017).

2.3 Truncated Quantiles

The use of quantiles is very important for estimating heavy-tailed distributions. We provide a lemma below that is useful when considering the truncated minimum of \mathbf{X} ; see Section 3.2.

Lemma 1 *Consider $\mathbf{X} = (X_1, \dots, X_n)$ with strictly positive components and marginal survival functions $h_i(\lambda_i x_i)$, $\lambda_i > 0, \forall i$, for d -times monotone h_i , $d \geq n$. Let ${}_{\tau}X_i = X_i | X_i > \tau$. The quantile of level p for ${}_{\tau}X_i$, $0 < p < 1$, is given by*

$$q_{{}_{\tau}X_i}(p) = \frac{h_i^{-1}((1-p)h_i(\lambda_i\tau))}{\lambda_i}.$$

Proof. Consider the distribution function of ${}_{\tau}X_i$,

$$F_{{}_{\tau}X_i}(x) = \frac{P(\tau < X_i \leq x)}{P(X_i > \tau)} = \frac{P(X_i > \tau) - P(X_i > x)}{P(X_i > \tau)} = 1 - \frac{P(X_i > x)}{P(X_i > \tau)} = 1 - \frac{h_i(\lambda_i x)}{h_i(\lambda_i \tau)}.$$

Inverting this function produces the desired result. ■

Remark 1 *An interesting example to which the results of Theorem 1 and Lemma 1 apply is found in a multivariate random vector with strictly positive components $X_i > 0$ such that its joint survival function is given by*

$$P(X_1 > x_1, \dots, X_n > x_n) = h \left(\sum_{i=1}^n h^{-1}(h_i(\lambda_i x_i)) \right), \quad x_i \geq 0,$$

for $\lambda_i > 0, \forall i$, where h and $h_i, \forall i$, are d -times monotone, $d \geq n$. We aim to study this distribution in future research.

3 Multivariate Truncation

Presently, we consider applying n -dimensional truncation to the random vector $\mathbf{X} = (X_1, \dots, X_n)$ with distribution generated by d -times monotone h , $d \geq n$, as given in Equation (1). Let ${}_{\tau}\mathbf{X} = \{\mathbf{X} | \mathbf{X} > \boldsymbol{\tau}\}$, where $\boldsymbol{\tau} = \tau \cdot \mathbf{1}_n$, and ${}_{\tau}X_i = \{X_i | \mathbf{X} > \boldsymbol{\tau}\}$. Below, we consider the moments of ${}_{\tau}X_i$ as well as the minimum and maximum of ${}_{\tau}\mathbf{X}$, ${}_{\tau}X_{(1)}$ and ${}_{\tau}X_{(n)}$.

Remark 2 Although subtle, notice that the left subscript, which indicates truncation, is bold in the case of multivariate truncation and plain in the case of univariate truncation. The difference is important, and only in the case of the truncated minimum do we find that $\tau X_{(1)} = \tau X_{(1)}$; see Section 3.2.

3.1 Mixed Moments of the Marginals

To obtain the variance-covariance structure of the multivariate truncated model considered here, we require the mixed moments of the marginal distributions.

Theorem 2 Consider $\mathbf{X} = (X_1, \dots, X_n)$ with distribution generated by d -times monotone h via Equation (1), $d \geq n$. Let $\tau X_i = \{X_i | \mathbf{X} > \boldsymbol{\tau}\}$. If finite,

$$\begin{aligned} \mathbb{E} \left[\prod_{i=1}^n X_i^{k_i} \right] &= (-1)^k h^{(-k)}(0) \prod_{i=1}^n \frac{k_i!}{\lambda_i^{k_i}}, \\ \mathbb{E} \left[\prod_{i=1}^n \tau X_i^{k_i} \right] &= h(\lambda_S \tau)^{-1} \sum_{j_1=0}^{k_1} \cdots \sum_{j_n=0}^{k_n} h^{(-\sum_{i=1}^n j_i)}(\lambda_S \tau) \prod_{i=1}^n \frac{(-1)^{j_i} \tau^{k_i - j_i} k_i!}{(k_i - j_i)! \lambda_i^{j_i}}, \end{aligned}$$

where $\lambda_S = \sum_{i=1}^n \lambda_i$, $k = \sum_{i=1}^n k_i$, $k \in \{1, 2, \dots, d\}$, and $k_i \in \{0\} \cup \mathbb{Z}^+$; furthermore, where $h^{(-k)}(x) = -\int_x^\infty h^{(-(k-1))}(y) dy$ and $h^{(0)}(x) = h(x)$.

Proof. We prove the result directly.

$$\mathbb{E} \left[\prod_{i=1}^n \tau X_i^{k_i} \right] = h(\lambda_S \tau)^{-1} \int_\tau^\infty \cdots \int_\tau^\infty (-1)^n \left\{ \prod_{i=1}^n x_i^{k_i} \lambda_i \right\} h^{(n)} \left(\sum_{l=1}^n \lambda_l x_l \right) dx_n \cdots dx_1.$$

In the case there exist $k_i = 0$, then, without loss of generality, for $m < n$, re-order the X_i such that $k_i > 0$ for $i = 1, \dots, m$ and $k_i = 0$ for $i = m + 1, \dots, n$. Evaluating the integrals that subsequently correspond to $k_i = 0$ is straightforward. Calculating each integral reduces the *level* of the function h by one and changes the relevant portion of its argument from $\lambda_l x_l$ to $\lambda_l \tau$, ultimately to become $\lambda_S \tau$. For each integral we also lose a factor of $-\lambda_i$. Consequently, we have that

$$\mathbb{E} \left[\prod_{i=1}^n \tau X_i^{k_i} \right] = h(\lambda_S \tau)^{-1} \int_\tau^\infty \cdots \int_\tau^\infty (-1)^m \left\{ \prod_{i=1}^m x_i^{k_i} \lambda_i \right\} h^{(m)} \left(\sum_{l=1}^m \lambda_l x_l + \sum_{l=m+1}^n \lambda_l \tau \right) dx_m \cdots dx_1.$$

Consider the integral with respect to x_m , the relevant terms are $h^{(m)}$, $-\lambda_m$, and $x_m^{k_m}$; the remaining terms can be pushed out of this integral. The following remains:

$$\int_\tau^\infty (-1) x_m^{k_m} \lambda_m h^{(m)} \left(\sum_{l=1}^m \lambda_l x_l + \sum_{l=m+1}^n \lambda_l \tau \right) dx_m.$$

Notice that this integral closely resembles the k_m^{th} raw moment of τX_m ; what is remarkable is that the *level* of h is m , rather than one, and that the argument of h includes a positive constant $\sum_{l=1}^{m-1} \lambda_l x_l + \sum_{l=m+1}^n \lambda_l \tau$. This positive constant presents no difficulty, in fact, neither does the level of h . Applying the same techniques used in Theorem 1, we find this integral equal to

$$\sum_{j_m=0}^{k_m} (-1)^{j_m} \tau^{k_m - j_m} \frac{k_m!}{(k_m - j_m)!} \frac{h^{(m-(j_m+1))} \left(\sum_{l=1}^{m-1} \lambda_l x_l + \sum_{l=m}^n \lambda_l \tau \right)}{\lambda_m^{j_m}}.$$

Substituting this back into our equation for $\mathbb{E}\left[\prod_{i=1}^n \tau X_i^{k_i}\right]$ and arranging the terms appropriately, we obtain

$$\mathbb{E}\left[\prod_{i=1}^n \tau X_i^{k_i}\right] = h(\lambda_S \tau)^{-1} \sum_{j_m=0}^{k_m} (-1)^{j_m} \frac{\tau^{k_m-j_m} k_m!}{(k_m-j_m)! \lambda_m^{j_m}} \times \\ \int_{\tau}^{\infty} \cdots \int_{\tau}^{\infty} (-1)^{(m-1)} \left\{ \prod_{i=1}^{m-1} x_i^{k_i} \lambda_i \right\} h^{(m-(j_m+1))} \left(\sum_{l=1}^{m-1} \lambda_l x_l + \sum_{l=m}^n \lambda_l \tau \right) dx_{m-1} \cdots dx_1.$$

We proceed, as above, by addressing the integral with respect to x_{m-1} ; again, we apply the same techniques used in Theorem 1 and obtain

$$\mathbb{E}\left[\prod_{i=1}^n \tau X_i^{k_i}\right] = h(\lambda_S \tau)^{-1} \sum_{j_m=0}^{k_m} \frac{(-1)^{j_m} \tau^{k_m-j_m} k_m!}{(k_m-j_m)! \lambda_m^{j_m}} \sum_{j_{m-1}=0}^{k_{m-1}} \frac{(-1)^{j_{m-1}} \tau^{k_{m-1}-j_{m-1}} k_{m-1}!}{(k_{m-1}-j_{m-1})! \lambda_{m-1}^{j_{m-1}}} \times \\ \int_{\tau}^{\infty} \cdots \int_{\tau}^{\infty} (-1)^{(m-2)} \left\{ \prod_{i=1}^{m-2} x_i^{k_i} \lambda_i \right\} h^{(m-(j_m+1)-(j_{m-1}+1))} \left(\sum_{l=1}^{m-2} \lambda_l x_l + \sum_{l=m-1}^n \lambda_l \tau \right) dx_{m-2} \cdots dx_1.$$

We continue in this way until we obtain

$$\mathbb{E}\left[\prod_{i=1}^n \tau X_i^{k_i}\right] = h(\lambda_S \tau)^{-1} \sum_{j_1=0}^{k_1} \cdots \sum_{j_m=0}^{k_m} h^{(m-\sum_{i=1}^m (j_i+1))} (\lambda_S \tau) \prod_{i=1}^m \frac{(-1)^{j_i} \tau^{k_i-j_i} k_i!}{(k_i-j_i)! \lambda_i^{j_i}}.$$

Notice that the sum and product may include terms corresponding to $i = m+1$ to $i = n$ since they do not alter the expression. Furthermore, notice that the level of h , $m - \sum_{l=1}^m (j_l + 1) = -\sum_{l=1}^m j_l = -\sum_{l=1}^n j_l$, the latter equality since $j_l = 0$ for $l = m+1, \dots, n$. This completes the result in the case of multivariate truncation, in order to show the result without truncation, set $\tau = 0$ and note that $h(0) = 1$. ■

As a result of Theorem 2, we obtain, if finite, the mean and variance of τX_i and the covariance of τX_i and τX_j , $i \neq j$.

$$\alpha_1(\tau X_i) = \tau - \frac{h^{(-1)}(\lambda_S \tau)}{\lambda_i h(\lambda_S \tau)}, \\ \mu_2(\tau X_i) = \frac{2h^{(-2)}(\lambda_S \tau)}{\lambda_i^2 h(\lambda_S \tau)} - \frac{h^{(-1)}(\lambda_S \tau)^2}{\lambda_i^2 h(\lambda_S \tau)^2}, \\ \text{Cov}(\tau X_i, \tau X_j) = \frac{h^{(-2)}(\lambda_S \tau)}{\lambda_i \lambda_j h(\lambda_S \tau)} - \frac{h^{(-1)}(\lambda_S \tau)^2}{\lambda_i \lambda_j h(\lambda_S \tau)^2}.$$

If we let the truncation point be zero, we retrieve the expectation and variance found above, and the covariance found in Hendriks and Landsman (2017).

3.2 Minimum and Maximum of the Truncated Distribution

Consider, again, $\mathbf{X} = (X_1, \dots, X_n)$ with distribution generated by d -times monotone h via Equation (1), $d \geq n$. Let $X_{(1)} = \min(\mathbf{X})$, $\tau \mathbf{X} = \{\mathbf{X} | \mathbf{X} > \tau\}$, and $\tau X_{(1)} = X_{(1)} | X_{(1)} > \tau$. Notice that the distribution of the truncated minimum is equal to the minimum of the truncated joint distribution; that is $\tau X_{(1)} = \tau X_{(1)} = \min(\tau \mathbf{X})$. To see

this, consider the survival probability of the truncated minimum for $x > \tau$,

$$\begin{aligned}
P({}_{\tau}X_{(1)} > x) &= P(X_{(1)} > x | X_{(1)} > \tau) \\
&= P(X_1 > x, \dots, X_n > x | X_1 > \tau, \dots, X_n > \tau) \\
&= P({}_{\tau}X_1 > x, \dots, {}_{\tau}X_n > x) \\
&= P({}_{\tau}X_{(1)} > x).
\end{aligned}$$

It is more convenient to truncate the minimum, since the distribution of the minimum is trivially obtained via survival function,

$$P(X_{(1)} > x) = P(X_1 > x, \dots, X_n > x) = h(\lambda_S x).$$

Therefore, the truncated moments of $X_{(1)}$ can be found via Theorem 1 by replacing λ_i with λ_S . The mean and variance of the truncated minimum are given by

$$\begin{aligned}
\alpha_1({}_{\tau}X_{(1)}) &= \tau - \frac{h^{(-1)}(\lambda_S \tau)}{(\lambda_S)h(\lambda_S \tau)}, \\
\mu_2({}_{\tau}X_{(1)}) &= \frac{2h^{(-2)}(\lambda_S \tau)}{(\lambda_S)^2 h(\lambda_S \tau)} - \frac{h^{(-1)}(\lambda_S \tau)^2}{(\lambda_S)^2 h(\lambda_S \tau)^2}.
\end{aligned}$$

Furthermore, we obtain the quantiles of ${}_{\tau}X_{(1)}$ via Lemma 1,

$$q_{{}_{\tau}X_{(1)}}(p) = \frac{h^{-1}((1-p)h(\lambda_S \tau))}{\lambda_S}.$$

We can also consider the maximum of the truncated distribution, ${}_{\tau}X_{(n)} = \max({}_{\tau}\mathbf{X})$. Notice that in contrast to the truncated minimum, the order of truncating and taking the maximum is not reversible. Furthermore, in order to provide tractable equations for moments of the maximum we require the constraint $\lambda_i = \lambda, \forall i$.

Corollary 3 Consider $\mathbf{X} = (X_1, \dots, X_n)$ with distribution generated by d -times monotone h via Equation (1), $d \geq n$, and $\lambda_i = \lambda$ for all i . Let $X_{(n)} = \max(\mathbf{X})$ and ${}_{\tau}X_{(n)} = \max(\mathbf{X}) | \mathbf{X} > \boldsymbol{\tau}$. If finite, the k^{th} raw moment, for $k \in \{1, 2, \dots, d\}$, of $X_{(n)}$ and ${}_{\tau}X_{(n)}$, $\tau \geq 0$, are given by,

$$\begin{aligned}
\alpha_k(X_{(n)}) &= \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} (-1)^k k! \frac{h^{(-k)}(0)}{(i\lambda)^k}, \\
\alpha_k({}_{\tau}X_{(n)}) &= h(n\lambda\tau)^{-1} \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} \sum_{j=0}^k (-1)^j \tau^{k-j} \frac{k!}{(k-j)!} \frac{h^{(-j)}(n\lambda\tau)}{(i\lambda)^j},
\end{aligned}$$

where $h^{(-k)}(x) = -\int_x^{\infty} h^{(-(k-1))}(y) dy$ and $h^{(0)}(x) = h(x)$.

Proof. We start with the distribution function of the maximum of the truncated random vector and apply the inclusion-exclusion result for probability.

$$\begin{aligned}
P({}_{\tau}X_{(n)} \leq x) &= P(X_{(n)} \leq x | \mathbf{X} > \boldsymbol{\tau}) = \frac{P(\boldsymbol{\tau} < \mathbf{X} \leq x \cdot \mathbf{1}_n)}{P(\mathbf{X} > \boldsymbol{\tau})} \\
&= \sum_{i=0}^n (-1)^i \binom{n}{i} \frac{h(i\lambda x + (n-i)\lambda\tau)}{h(n\lambda\tau)}.
\end{aligned}$$

Differentiate to find the density.

$$f_{\tau X_{(n)}}(x) = h(n\lambda\tau)^{-1} \sum_{i=1}^n (-1)^i \binom{n}{i} h^{(1)}(i\lambda x + (n-i)\lambda\tau) i\lambda, \quad x > \tau.$$

Notice that the summation starts from $i = 1$, since the term corresponding to $i = 0$ is zero. The k^{th} raw moment is given by

$$\begin{aligned} \alpha_k(\tau X_{(n)}) &= h(n\lambda\tau)^{-1} \int_{\tau}^{\infty} x^k \sum_{i=1}^n (-1)^i \binom{n}{i} i\lambda h^{(1)}(i\lambda x + (n-i)\lambda\tau) dx \\ &= h(n\lambda\tau)^{-1} \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} \int_{\tau}^{\infty} x^k (-i\lambda) h^{(1)}(i\lambda x + (n-i)\lambda\tau) dx. \end{aligned}$$

We find ourselves in familiar territory; applying the techniques used in Theorem 1, we find that the integral above is equal to

$$\sum_{j=0}^k (-1)^j \tau^{k-j} \frac{k!}{(k-j)!} \frac{h^{(-j)}(n\lambda\tau)}{(i\lambda)^j},$$

which, when substituted into the equation produces the desired result; this completes the result in the case of truncation, in order to show the result without truncation, set $\tau = 0$ and note that $h(0) = 1$. ■

As a result of Corollary 3, we obtain, if finite, the mean and variance of $\tau X_{(n)}$.

$$\begin{aligned} \alpha_1(\tau X_{(n)}) &= \tau - \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} \frac{h^{(-1)}(n\lambda\tau)}{i\lambda h(n\lambda\tau)}, \\ \mu_2(\tau X_{(n)}) &= \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} \frac{2h^{(-2)}(n\lambda\tau)}{(i\lambda)^2 h(n\lambda\tau)} - \left\{ \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} \frac{h^{(-1)}(n\lambda\tau)}{i\lambda h(n\lambda\tau)} \right\}^2. \end{aligned}$$

4 Examples

We provide six examples of multivariate distributions generated by multiply monotone functions h , which may be referred to as a multivariate Pareto, Clayton, Gumbel, Frank, Ali-Mikhail-Haq (AMH), and exponential-Pareto distribution, respectively. The generators of these distributions, for various suitable levels of θ , are plotted in Figure 1; we return to this figure in more detail below.

4.1 Pareto Distribution

In order to obtain the multivariate Pareto distribution studied in, for example, Chiragiev and Landsman (2007) and Alai *et al.* (2016), we define h such that

$$\begin{aligned} h(x) &= (1+x)^{-\frac{1}{\theta}}, \quad x \geq 0, \quad \theta \in \mathbb{R}^+, \\ h^{-1}(x) &= x^{-\theta} - 1. \end{aligned}$$

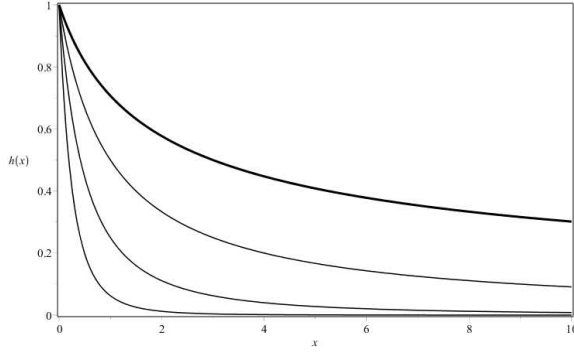
The function h is plotted in Figure 1a for several values of θ , where the greatest value of θ coincides with the heaviest tail, and so on. Of the six examples we consider, the

Pareto has the heaviest tail. For this reason, the first two antiderivatives of h exist only if θ is restricted to values less than one and one-half, respectively; they are given by

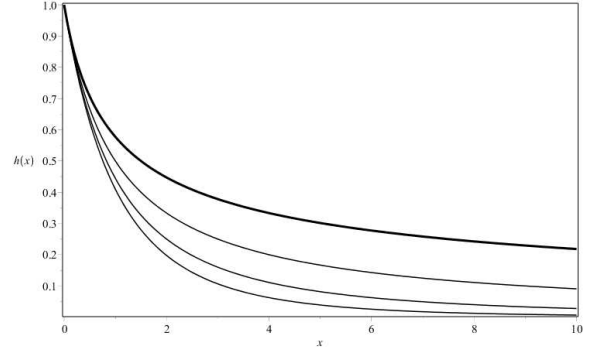
$$h^{(-1)}(x) = \frac{\theta(1+x)^{\frac{\theta-1}{\theta}}}{\theta-1}, \quad 0 < \theta < 1,$$

$$h^{(-2)}(x) = \frac{\theta^2(1+x)^{\frac{2\theta-1}{\theta}}}{(\theta-1)(2\theta-1)}, \quad 0 < \theta < \frac{1}{2}.$$

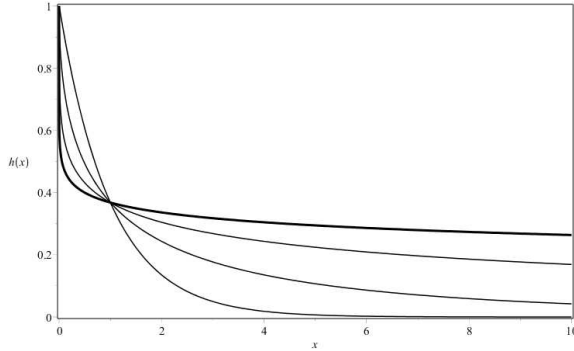
Figure 1: The generator function h .



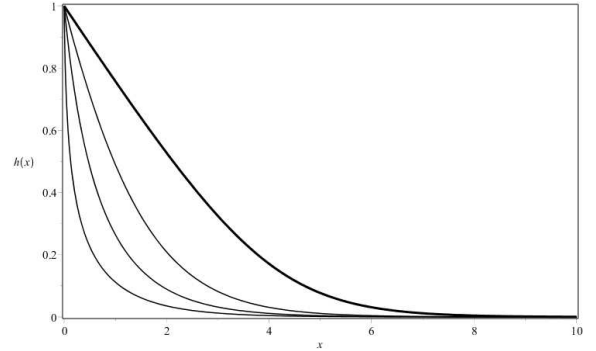
(a) Pareto, $\theta \in \{0.25, 0.50, 1, \mathbf{2}\}$.



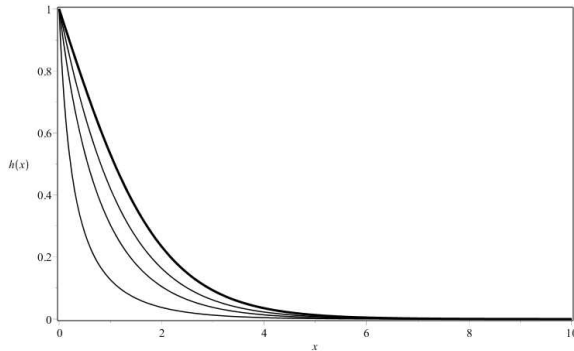
(b) Clayton, $\theta \in \{0.25, 0.50, 1, \mathbf{2}\}$.



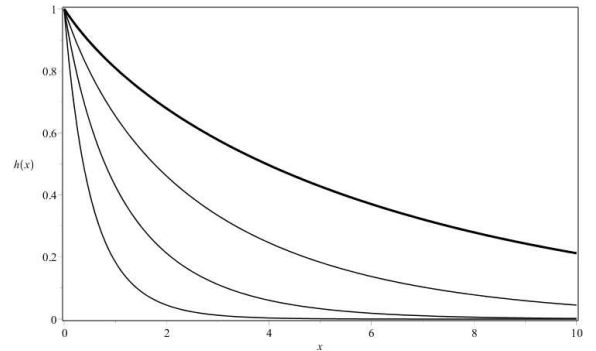
(c) Gumbel, $\theta \in \{1, 2, 4, \mathbf{8}\}$.



(d) Frank, $\theta \in \{-\mathbf{4}, -1, 1, \mathbf{4}\}$.



(e) AMH, $\theta \in \{-\mathbf{0.95}, -0.25, 0.25, 0.75\}$.



(f) Exponential-Pareto, $\theta \in \{1, 2, 4, \mathbf{8}\}$.

Consequently, we find the mean, variance, covariance and correlation of the marginal

distributions of ${}_{\tau}\mathbf{X}$ using Theorems 1 and 2.

$$\begin{aligned}\alpha_1({}_{\tau}X_i) &= \frac{\lambda_i^{-1} + \tau(\lambda_i^{-1}\lambda_S + \theta^{-1} - 1)}{\theta^{-1} - 1}, & 0 < \theta < 1, \\ \mu_2({}_{\tau}X_i) &= \frac{\theta^{-1}(\lambda_i^{-1} + \lambda_i^{-1}\lambda_S\tau)^2}{(\theta^{-1} - 1)^2(\theta^{-1} - 2)}, & 0 < \theta < \frac{1}{2}, \\ \text{Cov}({}_{\tau}X_i, {}_{\tau}X_j) &= \frac{(\lambda_i^{-\frac{1}{2}}\lambda_j^{-\frac{1}{2}} + \lambda_i^{-\frac{1}{2}}\lambda_j^{-\frac{1}{2}}\lambda_S\tau)^2}{(\theta^{-1} - 1)^2(\theta^{-1} - 2)}, & 0 < \theta < \frac{1}{2}, \\ \text{Corr}({}_{\tau}X_i, {}_{\tau}X_j) &= \theta, & 0 < \theta < \frac{1}{2}.\end{aligned}$$

Notice that the correlation between any of the marginal distributions is always θ , provided that $0 < \theta < \frac{1}{2}$, and does not depend on the truncation point τ . Although not particularly enlightening, the correlation for the Pareto distribution can be found in Figure 2a; we provide it for comparison's sake.

If we apply the parametrization used in Alai *et al.* (2016) given by $\lambda_i = \sigma^{-1}, \forall i$ and $\theta = \alpha^{-1}$, we obtain

$$\begin{aligned}\alpha_1({}_{\tau}X_i) &= \frac{\sigma + \tau(n + \alpha - 1)}{\alpha - 1}, & \alpha > 1, \\ \mu_2({}_{\tau}X_i) &= \frac{\alpha(\sigma + \tau n)^2}{(\alpha - 1)^2(\alpha - 2)}, & \alpha > 2, \\ \text{Cov}({}_{\tau}X_i, {}_{\tau}X_j) &= \frac{(\sigma + \tau n)^2}{(\alpha - 1)^2(\alpha - 2)}, & \alpha > 2, \\ \text{Corr}({}_{\tau}X_i, {}_{\tau}X_j) &= \frac{1}{\alpha}, & \alpha > 2.\end{aligned}$$

Notice that the covariance disagrees with what is shown in Theorem 2 of Alai *et al.* (2016), which appears to be the result of an error in the very last step of their proof. However, it seems the only consequence of that error is that their reported correlation should read $\frac{1}{\alpha}$ instead of $\frac{\sigma^2}{(\sigma + \tau n)^2\alpha}$; their estimation results are not affected.

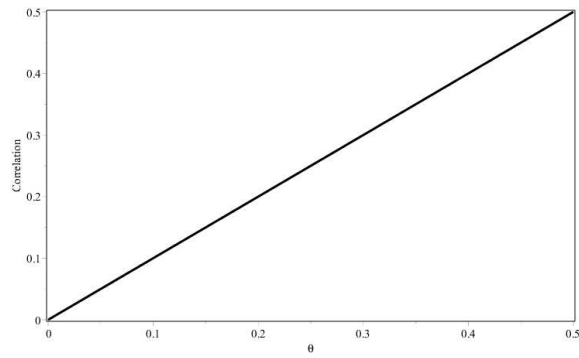
We presently consider three estimation techniques, two based on moments and one based on quantiles. Although beyond the scope of this work, we aim to more thoroughly consider the performance of these estimators and to compare them with maximum likelihood estimators, where applicable, in future research.

Mean-Variance Estimation

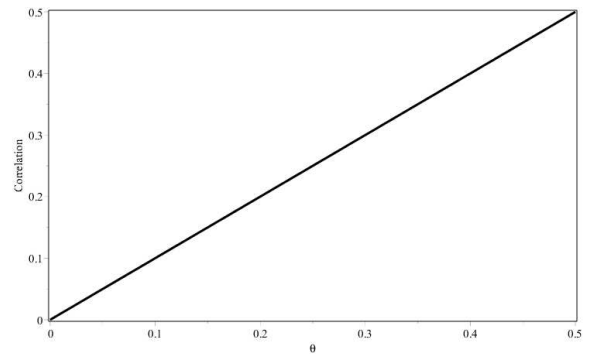
We consider two estimation techniques based on the method of moments, which we describe briefly. Consider data from a multivariate distribution governed by generator h and suppose $\lambda_i = \lambda, \forall i$. Denote by ${}_{\tau}\mathbf{X}^{(j)}$ the j^{th} joint observation of the multivariate truncated distribution, $j \in \{1, \dots, m\}$. Furthermore, let ${}_{\tau}X_{(1)}^{(j)}$ denote the minimum of each joint observation and ${}_{\tau}\mathbf{X}_{(1)}$ the collection of j minima.

The first estimation technique is based on the sample means and sample variances of the joint observations. Denote with a_1 and \tilde{m}_2 the sample mean and unbiased variance.

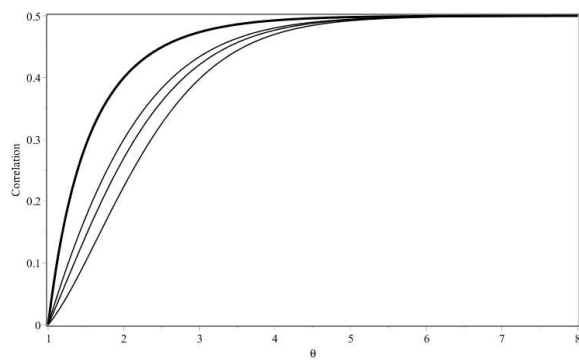
Figure 2: The correlation between marginal distributions for $\tau \in \{0, 1, 2, 5\}$.



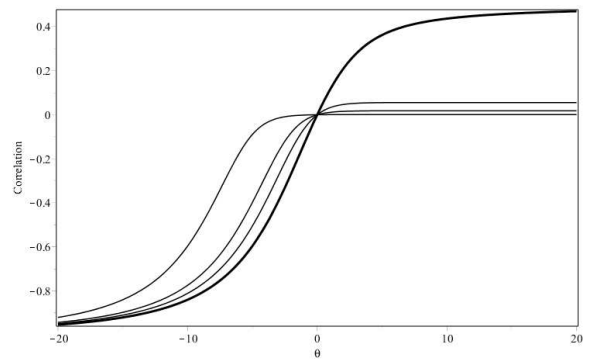
(a) Pareto.



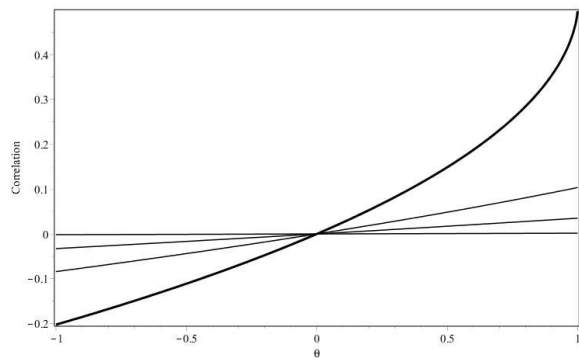
(b) Clayton.



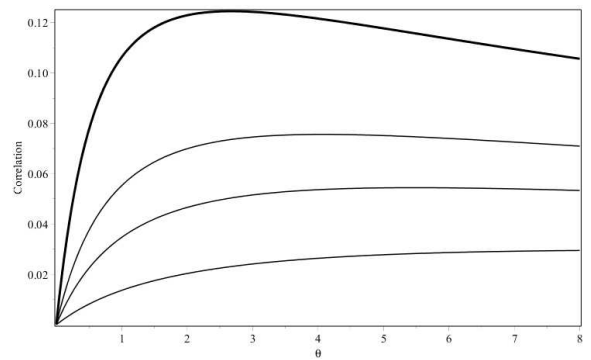
(c) Gumbel.



(d) Frank.



(e) AMH.



(f) Exponential-Pareto.

For each j ,

$$\begin{aligned} a_1(\boldsymbol{\tau}\mathbf{X}^{(j)}) &= \frac{1}{n} \sum_{i=1}^n \tau X_i^{(j)}, \\ \tilde{m}_2(\boldsymbol{\tau}\mathbf{X}^{(j)}) &= \frac{1}{n-1} \sum_{i=1}^n (\tau X_i^{(j)} - a_1(\boldsymbol{\tau}\mathbf{X}^{(j)}))^2. \end{aligned}$$

Trivially, the expectation of $a_1(\boldsymbol{\tau}\mathbf{X}^{(j)})$ is given by $\alpha_1(\tau X_1)$. The expectation of $\tilde{m}_2(\boldsymbol{\tau}\mathbf{X}^{(j)})$ is also readily found.

$$\begin{aligned} \mathbb{E}[a_1(\boldsymbol{\tau}\mathbf{X}^{(j)})] &= \alpha_1(\tau X_1), \\ \mathbb{E}[\tilde{m}_2(\boldsymbol{\tau}\mathbf{X}^{(j)})] &= \alpha_2(\tau X_1) - \mathbb{E}[\tau X_1 \tau X_2]. \end{aligned}$$

The above two expectations on the left-hand sides can be replaced by averages over the samples (i.e. average of the sample means and sample variances). The theoretical expressions on the right-hand sides are obtained from Theorem 2. Hence, this system of equations is able to yield estimates $\hat{\theta}$ and $\hat{\lambda}$.

The second estimation technique is based on the sample average and unbiased variance of the minima

$$\begin{aligned} a_1(\boldsymbol{\tau}\mathbf{X}_{(1)}) &= \frac{1}{m} \sum_{j=1}^m \tau X_{(1)}^{(j)}, \\ \tilde{m}_2(\boldsymbol{\tau}\mathbf{X}_{(1)}) &= \frac{1}{m-1} \sum_{j=1}^m (\tau X_{(1)}^{(j)} - a_1(\boldsymbol{\tau}\mathbf{X}_{(1)}))^2. \end{aligned}$$

Since the sample of minima are independent and identically distributed, we have that

$$\begin{aligned} \mathbb{E}[a_1(\boldsymbol{\tau}\mathbf{X}_{(1)})] &= \alpha_1(\tau X_{(1)}), \\ \mathbb{E}[\tilde{m}_2(\boldsymbol{\tau}\mathbf{X}_{(1)})] &= \mu_2(\tau X_{(1)}). \end{aligned}$$

The sample mean and variance can be used on the left-hand sides above. The theoretical expressions on the right-hand sides are obtained from Theorem 1. Hence, this system of equations is able to yield estimates $\hat{\theta}$ and $\hat{\lambda}$.

Both of the above mean-variance estimation procedures were implemented in Alai *et al.* (2016) specifically for the multivariate Pareto distribution. Notice that we have presented them in the general case, and hence, they can be implemented for the remaining five examples.

Quantile Estimation

In addition to the two mean-variance based estimation procedures above, we also investigate an estimation technique based on the sample quantiles of the observed minima $\boldsymbol{\tau}\mathbf{X}_{(1)}$; see Alai *et al.* (2016) for an application of this estimation procedure to the multivariate Pareto distribution. Notice that the distribution of the minimum is straightforward; from Section 3.2, we know the survival function is given by h with adjusted scale parameter λ_S . This means we can apply Lemma 1 in order to obtain the quantiles. For the Pareto, we have that

$$q_{\boldsymbol{\tau}\mathbf{X}_{(1)}}(p_i) = \lambda_S^{-1} \left((1 - p_i)^{-\theta} (1 + \lambda_S \tau) - 1 \right).$$

Solving for λ_S , we obtain

$$\lambda_S = \frac{(1 - p_i)^{-\theta} - 1}{q(p_i) - (1 - p_i)^{-\theta}\tau}.$$

Together with levels p_1 and p_2 , this produces the following equation that can be used to find $\hat{\theta}$.

$$\frac{q(p_1) - (1 - p_1)^{-\theta}\tau}{(1 - p_1)^{-\theta} - 1} = \frac{q(p_2) - (1 - p_2)^{-\theta}\tau}{(1 - p_2)^{-\theta} - 1}.$$

In order to estimate θ , replace theoretical quantiles, $q(p_i)$, with sample quantiles, $\hat{q}(p_i)$, and solve numerically. Finally, λ_S is estimated using $\hat{\theta}$ and sample quantile with level p_3 .

$$\hat{\lambda}_S = \frac{(1 - p_3)^{-\hat{\theta}} - 1}{\hat{q}(p_3) - (1 - p_3)^{-\hat{\theta}}\tau}.$$

This estimation procedure depends on the function h , as such, we return to it in the remaining five examples.

4.2 Clayton Generator

Closely related to the above multivariate Pareto distribution is the generator of the Clayton copula. If we let h take the form of the Clayton generator, we obtain

$$\begin{aligned} h(x) &= (1 + \theta x)^{-\frac{1}{\theta}}, & x \geq 0, \quad \theta \in \mathbb{R}^+, \\ h^{-1}(x) &= \frac{1}{\theta} \left(x^{-\theta} - 1 \right). \end{aligned}$$

The generator is plotted in Figure 1b for several values of θ , where, as is the case with the Pareto, greater values of θ correspond to heavier tails. The first two antiderivatives are still easily found, but given the nature of the tails, demand restrictions on θ .

$$\begin{aligned} h^{(-1)}(x) &= \frac{(1 + \theta x)^{\frac{\theta-1}{\theta}}}{\theta - 1}, & 0 < \theta < 1, \\ h^{(-2)}(x) &= \frac{(1 + \theta x)^{\frac{2\theta-1}{\theta}}}{(\theta - 1)(2\theta - 1)}, & 0 < \theta < \frac{1}{2}. \end{aligned}$$

These antiderivatives can be used to formulate mean-variance estimators as alluded to above. Using Theorem 2, we find the mean, variance, covariance and correlation of the marginal distributions of $\tau\mathbf{X}$.

$$\begin{aligned} \alpha_1(\tau X_i) &= \frac{\lambda_i^{-1} + \lambda_i^{-1}\lambda_S\tau}{1 - \theta}, & 0 < \theta < 1, \\ \mu_2(\tau X_i) &= \frac{(\lambda_i^{-1} + \theta\lambda_i^{-1}\lambda_S\tau)^2}{(1 - \theta)^2(1 - 2\theta)}, & 0 < \theta < \frac{1}{2}, \\ \text{Cov}(\tau X_i, \tau X_j) &= \frac{\theta(\lambda_i^{-\frac{1}{2}}\lambda_j^{-\frac{1}{2}} + \theta\lambda_i^{-\frac{1}{2}}\lambda_j^{-\frac{1}{2}}\lambda_S\tau)^2}{(1 - \theta)^2(1 - 2\theta)}, & 0 < \theta < \frac{1}{2}, \\ \text{Corr}(\tau X_i, \tau X_j) &= \theta, & 0 < \theta < \frac{1}{2}. \end{aligned}$$

Just as with the Pareto, the correlation in the case of the Clayton generator does not depend on λ_i nor τ ; see Figure 2b for the plot.

Although we do not elaborate on the mean-variance estimators, we do consider the quantile estimation procedure based on the minimum, $\tau X_{(1)}$. From Lemma 1, we obtain the following quantiles, for $0 < p_i < 1$,

$$q_{\tau X_{(1)}}(p_i) = \theta^{-1} \left(((1 - p_i)^{-\theta} - 1) \lambda_S^{-1} + (1 - p_i)^{-\theta} \theta \tau \right).$$

Solving for λ_S , we obtain

$$\lambda_S = \frac{(1 - p_i)^{-\theta} - 1}{\theta(q(p_i) - (1 - p_i)^{-\theta} \tau)},$$

which, with p_1 and p_2 and their corresponding sample quantiles yields an equation for θ able to produce $\hat{\theta}$,

$$\frac{(1 - p_1)^{-\hat{\theta}} - 1}{\hat{q}(p_1) - (1 - p_1)^{-\hat{\theta}} \tau} = \frac{(1 - p_2)^{-\hat{\theta}} - 1}{\hat{q}(p_2) - (1 - p_2)^{-\hat{\theta}} \tau},$$

where $\hat{q}(p_i)$ is the sample quantile. Finally, with p_3 and sample quantile $\hat{q}(p_3)$, we have

$$\hat{\lambda}_S = \frac{(1 - p_3)^{-\hat{\theta}} - 1}{\hat{\theta}(\hat{q}(p_3) - (1 - p_3)^{-\hat{\theta}} \tau)}.$$

Just as with the Pareto, the quantile estimation technique for the Clayton generator estimates θ first, using sample quantiles with levels p_1 and p_2 , and subsequently estimates λ_S using $\hat{\theta}$ and sample quantile with level p_3 .

4.3 Gumbel Generator: Weibull Distribution

Let h be the generator of the Gumbel copula.

$$\begin{aligned} h(x) &= \exp(-x^{\frac{1}{\theta}}), & x \geq 0, \quad \theta \in [1, \infty), \\ h^{-1}(x) &= (-\ln(x))^{\theta}. \end{aligned}$$

The plot of h for several values of θ is given in Figure 1c, where greater values of θ correspond to heavier tails.

Remark 3 *The consequence of letting h be the generator of the Gumbel copula produces the well-known Weibull distribution; a result we use to readily generate observations in the numerical example of Section 5.*

The antiderivatives can be expressed using the incomplete gamma function; we need not restrict θ any further for these to exist. We have, for $\theta \geq 1$,

$$\begin{aligned} h^{(-1)}(x) &= -\theta \Gamma(\theta, x^{\frac{1}{\theta}}), \\ h^{(-2)}(x) &= \theta (\Gamma(2\theta, x^{\frac{1}{\theta}}) - x \Gamma(\theta, x^{\frac{1}{\theta}})), \end{aligned}$$

where $\Gamma(s, x)$ is the incomplete gamma function, given by

$$\Gamma(s, x) = \int_x^{\infty} t^{s-1} e^{-t} dt.$$

We provide some details.

$$h^{(-1)}(x) = - \int_x^\infty \exp(-y^{\frac{1}{\theta}}) dy.$$

Consider the substitution $y = z^\theta$; we have $dy = \theta z^{\theta-1} dz$ and that $x < y < \infty$ implies $x^{\frac{1}{\theta}} < z < \infty$, since θ is positive.

$$\begin{aligned} h^{(-1)}(x) &= - \int_{x^{\frac{1}{\theta}}}^\infty \theta z^{\theta-1} e^{-z} dz \\ &= -\theta \int_{x^{\frac{1}{\theta}}}^\infty z^{\theta-1} e^{-z} dz = -\theta \Gamma(\theta, x^{\frac{1}{\theta}}). \end{aligned}$$

The second antiderivative is dealt with via the same substitution, followed by an application of integration by parts.

$$\begin{aligned} h^{(-2)}(x) &= - \int_x^\infty -\theta \Gamma(\theta, y^{\frac{1}{\theta}}) dy \\ &= \theta \int_{x^{\frac{1}{\theta}}}^\infty \theta z^{\theta-1} \Gamma(\theta, z) dz. \end{aligned}$$

Apply integration by parts, $u = \Gamma(\theta, z)$ and $dv = \theta z^{\theta-1} dz$; we have $du = -z^{\theta-1} e^{-z} dz$ and $v = z^\theta$.

$$\begin{aligned} h^{(-2)}(x) &= \theta \left(z^\theta \Gamma(\theta, z) \Big|_{x^{\frac{1}{\theta}}}^\infty + \int_{x^{\frac{1}{\theta}}}^\infty z^{2\theta-1} e^{-z} dz \right) \\ &= \theta (-x \Gamma(\theta, x^{\frac{1}{\theta}}) + \Gamma(2\theta, x^{\frac{1}{\theta}})), \end{aligned}$$

since $\lim_{z \rightarrow \infty} z^\theta \Gamma(\theta, z) = 0$ by applying L'Hôpital's Rule.

As a result of Theorem 2, we can obtain all the moments of interest required to establish mean-variance estimators. We provide the mean, variance, covariance and correlation of the marginal distributions τX_i below.

$$\begin{aligned} \alpha_1(\tau X_i) &= \tau + \frac{\theta e^\kappa}{\lambda_i} \Gamma(\theta, \kappa), \\ \mu_2(\tau X_i) &= \frac{\theta e^\kappa}{\lambda_i^2} \left(2\Gamma(2\theta, \kappa) - 2\kappa^\theta \Gamma(\theta, \kappa) - \theta e^\kappa \Gamma(\theta, \kappa)^2 \right), \\ \text{Cov}(\tau X_i, \tau X_j) &= \frac{\theta e^\kappa}{\lambda_i \lambda_j} \left(\Gamma(2\theta, \kappa) - \kappa^\theta \Gamma(\theta, \kappa) - \theta e^\kappa \Gamma(\theta, \kappa)^2 \right), \\ \text{Corr}(\tau X_i, \tau X_j) &= \frac{\Gamma(2\theta, \kappa) - \kappa^\theta \Gamma(\theta, \kappa) - \theta e^\kappa \Gamma(\theta, \kappa)^2}{2\Gamma(2\theta, \kappa) - 2\kappa^\theta \Gamma(\theta, \kappa) - \theta e^\kappa \Gamma(\theta, \kappa)^2}, \end{aligned}$$

where $\kappa = \lambda_S \tau^{\frac{1}{\theta}}$. In contrast with the previous two examples, the correlation is not easy to interpret. It is still independent of the λ_i , but is no longer independent of the truncation point τ . In Figure 2c we plot the correlation for several values of τ . The bold line represents the case with no truncation; as the truncation point increases, the correlation decreases. From Figure 2c, it can be seen that the Gumbel generator is able to produce correlation between zero and 0.5, with greater values of θ corresponding to higher positive correlation. The relationship between positive correlation and heavy tails is expected; by defining the joint distribution via functional form h given in Equation

(1), it implies that having strong positive correlation is related to obtaining high joint survival probabilities for large values.

From Lemma 1, we obtain the quantiles of ${}_{\tau}X_{(1)}$ for $0 < p_i < 1$.

$$q_{{}_{\tau}X_{(1)}}(p_i) = \left(-\ln(1 - p_i)\lambda_S^{-\frac{1}{\theta}} + \tau^{\frac{1}{\theta}} \right)^{\theta}.$$

Solving for λ_S , we obtain

$$\lambda_S = \left(\frac{\ln(1 - p_i)}{\tau^{\frac{1}{\theta}} - q(p_i)^{\frac{1}{\theta}}} \right)^{\theta},$$

which, with p_1 and p_2 and corresponding sample quantiles $\widehat{q}(p_1)$ and $\widehat{q}(p_2)$ yield an equation for θ able to produce $\widehat{\theta}$,

$$\frac{\ln(1 - p_1)}{\tau^{\frac{1}{\theta}} - \widehat{q}(p_1)^{\frac{1}{\theta}}} = \frac{\ln(1 - p_2)}{\tau^{\frac{1}{\theta}} - \widehat{q}(p_2)^{\frac{1}{\theta}}}.$$

Finally, with p_3 , sample quantile $\widehat{q}(p_3)$ and estimate $\widehat{\theta}$, we have

$$\widehat{\lambda}_S = \left(\frac{\ln(1 - p_3)}{\tau^{\frac{1}{\widehat{\theta}}} - \widehat{q}(p_3)^{\frac{1}{\widehat{\theta}}}} \right)^{\widehat{\theta}}.$$

In the presence of truncation, the Pareto, Clayton, and Gumbel generators result in estimating θ first, and λ , second. For the remaining examples, we find the reverse. However, before continuing, we consider the special case of the Gumbel generator where $\theta = 1$.

Independence

Let h be given by the Gumbel generator with $\theta = 1$, in which case the marginal distributions are independent.

$$\begin{aligned} h(x) &= e^{-x}, & x &\geq 0, \\ h^{-1}(x) &= -\ln(x), \end{aligned}$$

and hence

$$P(X_1 > x_1, \dots, X_n > x_n) = \prod_{i=1}^n P(X_i > x_i).$$

Furthermore,

$$\begin{aligned} h^{(-1)}(x) &= -e^{-x}, \\ h^{(-2)}(x) &= e^{-x}, \end{aligned}$$

which means, Theorem 2 yields

$$\begin{aligned} \alpha_1({}_{\tau}X_i) &= \tau + \lambda_i^{-1}, \\ \mu_2({}_{\tau}X_i) &= \lambda_i^{-2}. \end{aligned}$$

Finally, the quantiles of ${}_{\tau}X_{(1)}$ for $0 < p_i < 1$ are given by

$$q_{{}_{\tau}X_{(1)}}(p_i) = \tau - \lambda_S^{-1} \ln(1 - p_i).$$

4.4 Frank Generator

Next, we consider the generator of the Frank copula,

$$h(x) = -\theta^{-1} \ln(1 + e^{-x}(e^{-\theta} - 1)), \quad x \geq 0, \quad \theta \in \mathbb{R} \setminus \{0\},$$

$$h^{-1}(x) = -\ln\left(\frac{e^{-\theta x} - 1}{e^{-\theta} - 1}\right).$$

The plot of h for several values of θ is given in Figure 1d. In contrast to the previous examples, the lower curves correspond to greater values of θ . However, from the plot, it becomes immediately clear that we no longer have the ability to capture heavy-tailed phenomena; in other words, all values of θ produce relatively light tails.

One benefit of the light tails is that we find the antiderivatives without having to restrict the values of θ . We have, for $\theta \neq 0$,

$$h^{(-1)}(x) = \frac{\text{Li}_2(-e^{-x}(e^{-\theta} - 1))}{\theta},$$

$$h^{(-2)}(x) = \frac{\text{Li}_3(-e^{-x}(e^{-\theta} - 1))}{\theta},$$

where $\text{Li}_s(z)$ is the polylogarithm function, given by

$$\text{Li}_2(z) = \int_z^0 \frac{\ln(1-t)}{t} dt,$$

$$\text{Li}_{s+1}(z) = -\int_z^0 \frac{\text{Li}_s(t)}{t} dt, \quad s \in \{2, 3, \dots\}.$$

We provide some details.

$$h^{(-1)}(x) = -\int_x^\infty -\theta^{-1} \ln(1 + e^{-y}(e^{-\theta} - 1)) dy.$$

Apply the substitution $z = -e^{-y}(e^{-\theta} - 1)$; we have $dz = e^{-y}(e^{-\theta} - 1) dy$, which means $dy = -z^{-1} dz$. The lower limit of integration is $-e^{-x}(e^{-\theta} - 1)$ and the upper limit is zero.

$$\begin{aligned} h^{(-1)}(x) &= \theta^{-1} \int_x^\infty \ln(1 + e^{-y}(e^{-\theta} - 1)) dy \\ &= \theta^{-1} \int_{-e^{-x}(e^{-\theta}-1)}^0 \frac{\ln(1-z)}{-z} dz \\ &= -\frac{\text{Li}_2(-e^{-x}(e^{-\theta} - 1))}{\theta}. \end{aligned}$$

The second antiderivative is found by, once again, applying the same substitution.

$$\begin{aligned} h^{(-2)}(x) &= \int_x^\infty \frac{\text{Li}_2(-e^{-y}(e^{-\theta} - 1))}{\theta} dy \\ &= \int_{-e^{-x}(e^{-\theta}-1)}^0 \frac{\text{Li}_2(z)}{\theta(-z)} dz \\ &= \frac{\text{Li}_3(-e^{-x}(e^{-\theta} - 1))}{\theta}. \end{aligned}$$

As a result of Theorem 2, we can obtain, amongst others, the mean, variance, covariance and correlation of the marginal distributions τX_i .

$$\begin{aligned}\alpha_1(\tau X_i) &= \tau - \frac{\text{Li}_2(\kappa)}{\lambda_i \ln(1 - \kappa)}, \\ \mu_2(\tau X_i) &= \frac{-2\text{Li}_3(\kappa)}{\lambda_i^2 \ln(1 - \kappa)} - \frac{\text{Li}_2^2(\kappa)}{\lambda_i^2 \ln^2(1 - \kappa)}, \\ \text{Cov}(\tau X_i, \tau X_j) &= \frac{-\text{Li}_3(\kappa)}{\lambda_i \lambda_j \ln(1 - \kappa)} - \frac{\text{Li}_2^2(\kappa)}{\lambda_i \lambda_j \ln^2(1 - \kappa)}, \\ \text{Corr}(\tau X_i, \tau X_j) &= \frac{-\text{Li}_3(\kappa) \ln(1 - \kappa) - \text{Li}_2^2(\kappa)}{-2\text{Li}_3(\kappa) \ln(1 - \kappa) - \text{Li}_2^2(\kappa)},\end{aligned}$$

where $\kappa = -e^{-\lambda_S \tau}(e^{-\theta} - 1)$. As with the Gumbel, the Frank correlation is difficult to interpret except in graphical form; we plot the correlation as a function of θ for various levels of τ in Figure 2d. The bold line corresponds to the case of no truncation, and the correlation decreases (in absolute value) as the truncation point increases. Notice that the range of correlations one is able to capture using the Frank generator goes from -1 to 0.5 ; this added flexibility is a trade-off to the inability to capture heavy-tailed phenomena, one that we also see in the case of the AMH generator below.

The quantiles of $\tau X_{(1)}$ for $0 < p_i < 1$ are given by Lemma 1.

$$q_{\tau X_{(1)}}(p_i) = -\lambda_S^{-1} \ln \left(\frac{e^{1-p_i}(1 + e^{-\lambda_S \tau}(e^{-\theta} - 1)) - 1}{e^{-\theta} - 1} \right),$$

which, unlike in the case of the Pareto, Clayton, and Gumbel examples, cannot be solved for λ_S , rather for θ .

$$\theta = -\ln \left(\frac{e^{\lambda_S \tau}(1 - e^{p_i-1})}{e^{\lambda_S(\tau - q(p_i))} e^{p_i-1} - 1} + 1 \right).$$

By specifying levels p_1 and p_2 , we obtain an equation of λ_S able to yield estimate $\widehat{\lambda}_S$ using sample quantiles $\widehat{q}(p_1)$ and $\widehat{q}(p_2)$.

$$\frac{1 - e^{\widehat{\lambda}_S(\tau - \widehat{q}(p_1))} e^{p_1-1}}{1 - e^{\widehat{\lambda}_S(\tau - \widehat{q}(p_2))} e^{p_2-1}} = \frac{1 - e^{p_1-1}}{1 - e^{p_2-1}}.$$

Finally, together with sample quantile $\widehat{q}(p_3)$, we estimate θ ,

$$\widehat{\theta} = -\ln \left(\frac{e^{\widehat{\lambda}_S \tau}(1 - e^{p_3-1})}{e^{\widehat{\lambda}_S(\tau - \widehat{q}(p_3))} e^{p_3-1} - 1} + 1 \right).$$

4.5 Ali-Mikhail-Haq Generator

We consider the generator for the AMH copula, given by

$$\begin{aligned}h(x) &= \frac{1 - \theta}{e^x - \theta}, \quad x \geq 0, \quad \theta \in [-1, 1), \\ h^{-1}(x) &= \ln \left(\frac{1 - \theta}{x} + \theta \right).\end{aligned}$$

We plot h for several values of θ in Figure 1e, where, like with the Frank generator, lower curves correspond to greater values of θ . Similarly, we find that the AMH generator is unable to capture heavy-tailed phenomena. We have, for $\theta \in [-1, 1) \setminus \{0\}$,

$$\begin{aligned} h^{(-1)}(x) &= \frac{1-\theta}{\theta} \ln(1 - \theta e^{-x}), \\ h^{(-2)}(x) &= \frac{1-\theta}{\theta} \text{Li}_2(\theta e^{-x}), \end{aligned}$$

where $\text{Li}_2(z) = \int_z^0 \frac{\ln(1-t)}{t} dt$ is the dilogarithm function. We provide some details; consider the first antiderivative.

$$h^{(-1)}(x) = - \int_x^\infty \frac{1-\theta}{e^y - \theta} dy.$$

Apply the following substitution, $z - 1 = \theta(e^y - \theta)^{-1}$; this means $dy = \frac{-1}{z(z-1)} dz$, the lower limit of integration is $\theta(e^x - \theta)^{-1} + 1$ and the upper is one. We obtain

$$\begin{aligned} h^{(-1)}(x) &= \frac{1-\theta}{\theta} \int_{\theta(e^x - \theta)^{-1} + 1}^1 z^{-1} dz \\ &= \frac{1-\theta}{\theta} \left(-\ln(\theta(e^x - \theta)^{-1} + 1) \right) \\ &= \frac{1-\theta}{\theta} \ln(1 - \theta e^{-x}). \end{aligned}$$

The second antiderivate is found by applying the substitution $z = \theta e^{-y}$, resulting in $dy = -z^{-1} dz$ below.

$$\begin{aligned} h^{(-2)}(x) &= - \int_x^\infty \frac{1-\theta}{\theta} \ln(1 - \theta e^{-y}) dy \\ &= - \frac{1-\theta}{\theta} \int_{\theta e^{-x}}^0 \frac{\ln(1-z)}{-z} dz \\ &= \frac{1-\theta}{\theta} \text{Li}_2(\theta e^{-x}). \end{aligned}$$

As a result of Theorem 2, we can obtain, among others, the mean, variance, covariance and correlation of the marginal distributions ${}_\tau X_i$.

$$\begin{aligned} \alpha_1({}_\tau X_i) &= \tau - \frac{\ln(1-\kappa)}{\lambda_i} \frac{1-\kappa}{\kappa}, \\ \mu_2({}_\tau X_i) &= \frac{2\text{Li}_2(\kappa)}{\lambda_i^2} \frac{1-\kappa}{\kappa} - \frac{\ln^2(1-\kappa)}{\lambda_i^2} \frac{(1-\kappa)^2}{\kappa^2}, \\ \text{Cov}({}_\tau X_i, {}_\tau X_j) &= \frac{\text{Li}_2(\kappa)}{\lambda_i \lambda_j} \frac{1-\kappa}{\kappa} - \frac{\ln^2(1-\kappa)}{\lambda_i \lambda_j} \frac{(1-\kappa)^2}{\kappa^2}, \\ \text{Corr}({}_\tau X_i, {}_\tau X_j) &= \frac{\text{Li}_2(\kappa)\kappa - \ln^2(1-\kappa)(1-\kappa)}{2\text{Li}_2(\kappa)\kappa - \ln^2(1-\kappa)(1-\kappa)}, \end{aligned}$$

where $\kappa = \theta e^{-\lambda_s \tau}$. We plot the correlation over θ for several values of τ in Figure 2e; the bold line corresponds to $\tau = 0$. We find that both positive and negative correlation can be captured and that the greater the truncation point, the lower, in absolute value, the correlation.

From Lemma 1, we find the quantiles of ${}_{\tau}X_{(1)}$, for $0 < p_i < 1$.

$$q_{{}_{\tau}X_{(1)}}(p_i) = \lambda_S^{-1} \ln \left(\frac{e^{\lambda_S \tau} - \theta p_i}{1 - p_i} \right),$$

which, like in the case of the Frank generator, can be solved for θ , not λ_S .

$$\theta = \frac{e^{\lambda_S \tau} - (1 - p_i)e^{\lambda_S q(p_i)}}{p_i}.$$

Together with levels p_1 and p_2 and their corresponding sample quantiles, we can find an equation to estimate λ_S .

$$\frac{e^{\widehat{\lambda}_S \tau} - (1 - p_1)e^{\widehat{\lambda}_S \widehat{q}(p_1)}}{e^{\widehat{\lambda}_S \tau} - (1 - p_2)e^{\widehat{\lambda}_S \widehat{q}(p_2)}} = \frac{p_1}{p_2}.$$

Finally, together with p_3 and corresponding sample quantile $\widehat{q}(p_3)$, we can estimate θ .

$$\widehat{\theta} = \frac{e^{\widehat{\lambda}_S \tau} - (1 - p_3)e^{\widehat{\lambda}_S \widehat{q}(p_3)}}{p_3}.$$

4.6 Product Exponential-Pareto

Finally, consider the generator given by

$$h(x) = \frac{e^{-\frac{x}{\theta}}}{(1+x)^{\frac{1}{\theta}}}, \quad x \geq 0, \quad \theta \in \mathbb{R}^+.$$

The inverse h^{-1} requires the use of the Lambert function W , it is given by

$$h^{-1}(x) = W(x^{-\theta}e) - 1,$$

where $y = xe^x \iff x = W(y)$. A plot of h for several values of θ is given in Figure 1f. From this plot, we notice the similarity of the exponential-Pareto generator to the Pareto and Clayton generators. However, the tails are not sufficiently heavy to require restrictions on θ when calculating the antiderivatives. We have, for $\theta > 0$,

$$\begin{aligned} h^{(-1)}(x) &= -e^{\frac{1}{\theta}} \theta^{1-\frac{1}{\theta}} \Gamma\left(1 - \frac{1}{\theta}, \frac{x+1}{\theta}\right), \\ h^{(-2)}(x) &= e^{\frac{1}{\theta}} \theta^{2-\frac{1}{\theta}} \left(\Gamma\left(2 - \frac{1}{\theta}, \frac{x+1}{\theta}\right) - \frac{x+1}{\theta} \Gamma\left(1 - \frac{1}{\theta}, \frac{x+1}{\theta}\right) \right). \end{aligned}$$

We provide some details. Consider the first antiderivative.

$$h^{(-1)}(x) = - \int_x^{\infty} e^{-\frac{y}{\theta}} (1+y)^{-\frac{1}{\theta}} dy.$$

Apply the substitution $z = \frac{y+1}{\theta}$; $dy = \theta dz$. We obtain

$$\begin{aligned} h^{(-1)}(x) &= - \int_{\frac{x+1}{\theta}}^{\infty} e^{\frac{1}{\theta}-z} (\theta z)^{-\frac{1}{\theta}} \theta dz \\ &= -e^{\frac{1}{\theta}} \theta^{1-\frac{1}{\theta}} \int_{\frac{x+1}{\theta}}^{\infty} z^{(1-\frac{1}{\theta})-1} e^{-z} dz \\ &= -e^{\frac{1}{\theta}} \theta^{1-\frac{1}{\theta}} \Gamma\left(1 - \frac{1}{\theta}, \frac{x+1}{\theta}\right). \end{aligned}$$

To find the second antiderivative, apply the same substitution and subsequently apply integration by parts.

$$\begin{aligned} h^{(-2)}(x) &= \int_x^\infty e^{\frac{1}{\theta}} \theta^{1-\frac{1}{\theta}} \Gamma\left(1 - \frac{1}{\theta}, \frac{y+1}{\theta}\right) \\ &= e^{\frac{1}{\theta}} \theta^{2-\frac{1}{\theta}} \int_{\frac{x+1}{\theta}}^\infty \Gamma\left(1 - \frac{1}{\theta}, z\right) dz. \end{aligned}$$

Let $u = \Gamma\left(1 - \frac{1}{\theta}, z\right)$, $du = -z^{(1-\frac{1}{\theta})-1} e^{-z} dz$; and $dv = dz, v = z$.

$$\begin{aligned} h^{(-2)}(x) &= e^{\frac{1}{\theta}} \theta^{2-\frac{1}{\theta}} \left(z \Gamma\left(1 - \frac{1}{\theta}, z\right) \Big|_{\frac{x+1}{\theta}}^\infty + \int_{\frac{x+1}{\theta}}^\infty z^{(2-\frac{1}{\theta})-1} e^{-z} dz \right) \\ &= e^{\frac{1}{\theta}} \theta^{2-\frac{1}{\theta}} \left(\Gamma\left(2 - \frac{1}{\theta}, \frac{x+1}{\theta}\right) - \frac{x+1}{\theta} \Gamma\left(1 - \frac{1}{\theta}, \frac{x+1}{\theta}\right) \right), \end{aligned}$$

since $\lim_{z \rightarrow \infty} z \Gamma\left(1 - \frac{1}{\theta}, z\right) = 0$ by applying L'Hôpital's Rule.

As a result of Theorem 2, we obtain, among others, the mean, variance, covariance and correlation of the marginal distributions ${}_{\tau}X_i$.

$$\begin{aligned} \alpha_1({}_{\tau}X_i) &= \tau + \frac{\theta e^{\kappa} \kappa^{\frac{1}{\theta}}}{\lambda_i} \Gamma\left(1 - \frac{1}{\theta}, \kappa\right), \\ \mu_2({}_{\tau}X_i) &= \frac{\theta^2 e^{\kappa} \kappa^{\frac{1}{\theta}}}{\lambda_i^2} \left(2\Gamma\left(2 - \frac{1}{\theta}, \kappa\right) - 2\kappa \Gamma\left(1 - \frac{1}{\theta}, \kappa\right) - e^{\kappa} \kappa^{\frac{1}{\theta}} \Gamma^2\left(1 - \frac{1}{\theta}, \kappa\right) \right), \\ \text{Cov}({}_{\tau}X_i, {}_{\tau}X_j) &= \frac{\theta^2 e^{\kappa} \kappa^{\frac{1}{\theta}}}{\lambda_i \lambda_j} \left(\Gamma\left(2 - \frac{1}{\theta}, \kappa\right) - \kappa \Gamma\left(1 - \frac{1}{\theta}, \kappa\right) - e^{\kappa} \kappa^{\frac{1}{\theta}} \Gamma^2\left(1 - \frac{1}{\theta}, \kappa\right) \right), \\ \text{Corr}({}_{\tau}X_i, {}_{\tau}X_j) &= \frac{\Gamma\left(2 - \frac{1}{\theta}, \kappa\right) - \kappa \Gamma\left(1 - \frac{1}{\theta}, \kappa\right) - e^{\kappa} \kappa^{\frac{1}{\theta}} \Gamma^2\left(1 - \frac{1}{\theta}, \kappa\right)}{2\Gamma\left(2 - \frac{1}{\theta}, \kappa\right) - 2\kappa \Gamma\left(1 - \frac{1}{\theta}, \kappa\right) - e^{\kappa} \kappa^{\frac{1}{\theta}} \Gamma^2\left(1 - \frac{1}{\theta}, \kappa\right)}, \end{aligned}$$

where $\kappa = \frac{\lambda_S \tau + 1}{\theta}$. A plot of the correlation as a function of θ for several values of τ is given in Figure 2f. The correlation in the case of the exponential-Pareto is essentially a mixture of the Pareto correlation and the case of independence; as θ increases, the exponential function dominates and acts to reduce the correlation. This feature is unique to the exponential-Pareto; in the other examples, correlation is a monotone function of θ . Furthermore, an increase in the truncation point τ reduces the correlation.

The quantiles of ${}_{\tau}X_{(1)}$ for $0 < p_i < 1$ are found using Lemma 1.

$$q_{{}_{\tau}X_{(1)}}(p_i) = \frac{W((1-p_i)^{-\theta} (1 + \lambda_S \tau) e^{(1+\lambda_S \tau)}) - 1}{\lambda_S},$$

which, like in the case of the Frank and AMH generators, can be solved for θ , but not λ_S .

$$\theta = \frac{\lambda_S(\tau - q(p_i)) + \ln(1 + \lambda_S \tau) - \ln(1 + \lambda_S q(p_i))}{\ln(1 - p_i)}.$$

Together with levels p_1 and p_2 , we have an equation able to yield estimate $\widehat{\lambda}_S$,

$$\frac{\lambda_S(\tau - q(p_1)) + \ln(1 + \lambda_S \tau) - \ln(1 + \lambda_S q(p_1))}{\lambda_S(\tau - q(p_2)) + \ln(1 + \lambda_S \tau) - \ln(1 + \lambda_S q(p_2))} = \frac{\ln(1 - p_1)}{\ln(1 - p_2)}.$$

Finally, with p_3 and corresponding sample quantile, we obtain

$$\widehat{\theta} = \frac{\widehat{\lambda}_S(\tau - \widehat{q}(p_3)) + \ln(1 + \widehat{\lambda}_S \tau) - \ln(1 + \widehat{\lambda}_S \widehat{q}(p_3))}{\ln(1 - p_3)}.$$

5 Optimal Quantile Selection

In the preceding section we considered three estimation procedures for θ and λ_S for six examples. The last of these estimation procedures is based on the quantiles of the truncated minimum $\tau X_{(1)}$ and requires three sample quantiles. In this section, we investigate how to select these three quantiles optimally.

5.1 Specifying the Objective Functions

We base our optimality criteria on statistical estimation theory developed in Landsman (1996). Consider a sample of independent and identically distributed X_1, \dots, X_n with density $f(x, \vartheta)$, $\vartheta \in \Theta \subset \mathbb{R}$, differentiable with respect to ϑ for almost all $x \in \mathbb{R}$. The Fisher information about parameter ϑ contained in some statistic $T_n(X_1, \dots, X_n)$ is defined as

$$I_{T_n}(\vartheta) = \int_{\mathbb{R}} \left(\frac{\partial \ln f_{T_n}(x, \vartheta)}{\partial \vartheta} \right)^2 f_{T_n}(x, \vartheta) dx.$$

A higher Fisher information is indicative of more precise estimation. Consider sample quantiles $\widehat{q}(p_1), \dots, \widehat{q}(p_k)$ with corresponding levels $0 < p_1 \leq \dots \leq p_k < 1$. Theorem 1 in Landsman (1996) shows that the Fisher information contained in the sample quantiles, $I_{\widehat{q}(p_1), \dots, \widehat{q}(p_k)}(\vartheta)$, is asymptotically equal to $nI_k(p_1, \dots, p_k)$, with $I_k(p_1, \dots, p_k)$ defined as

$$I_k(p_1, \dots, p_k) = \sum_{i=0}^k \frac{(\beta_{i+1} - \beta_i)^2}{p_{i+1} - p_i},$$

where $p_0 = 0$, $p_{k+1} = 1$, $\beta_i = f(q(p_i), \vartheta) \partial q(p_i) / \partial \vartheta$, for $i = 1, \dots, k$ and $\beta_0 = \beta_{k+1} = 0$. We find optimal quantiles p_1^*, \dots, p_k^* , such that I_k is maximized. In the case of using two sample quantiles, i.e. $k = 2$, we obtain the following objective function:

$$I_2(p_1, p_2) = \frac{\beta_1^2}{p_1} + \frac{(\beta_2 - \beta_1)^2}{p_2 - p_1} + \frac{\beta_2^2}{1 - p_2},$$

and in the case of using one sample quantile, we obtain

$$I_1(p_1) = \frac{\beta_1^2}{p_1} + \frac{\beta_1^2}{1 - p_1} = \frac{\beta_1^2}{p_1(1 - p_1)}.$$

The presence of truncation considerably complicates the optimal quantile selection problem. Hence, we presently consider the case where $\tau = 0$; we aim to address the case where $\tau > 0$ in future research.

5.2 The Case with no Truncation

In general, the density and quantile functions of $X_{(1)}$, as introduced in Section 3.2, are given by

$$\begin{aligned} f(x) &= -\lambda_S h^{(1)}(\lambda_S x), \\ q(p) &= \lambda_S^{-1} h^{-1}(\check{p}), \end{aligned}$$

where $\check{p} = 1 - p$. Notice that in the absence of truncation, λ_S is trivially isolated in the equation for $q(p)$.

$$\lambda_S = \frac{h^{-1}(\check{p})}{q(p)}.$$

Consequently, one equation using two sample quantiles $\widehat{q}(p_1)$ and $\widehat{q}(p_2)$ could be formulated to estimate θ .

$$\frac{h^{-1}(\check{p}_1; \theta)}{q(p_1)} = \frac{h^{-1}(\check{p}_2; \theta)}{q(p_2)}.$$

Hence, in the case with no truncation, one can readily estimate θ first using two quantiles, and subsequently estimate λ_S using one quantile. Note that this is also done in the case with truncation for the Pareto, Clayton and Gumbel examples, but not for the Frank, AMH and exponential-Pareto.

In order to obtain optimal quantile levels, we need to derive the β functions outlined above. We have that

$$f(q(p)) = -\lambda_S h^{(1)}(h^{-1}(\check{p})),$$

and the partial derivatives of the quantile function with respect to parameters θ and λ_S are

$$\begin{aligned} \frac{\partial q(p)}{\partial \theta} &= \lambda_S^{-1} \frac{\partial h^{-1}(\check{p})}{\partial \theta}, \\ \frac{\partial q(p)}{\partial \lambda_S} &= -\lambda_S^{-2} h^{-1}(\check{p}), \end{aligned}$$

respectively. As a result, the corresponding β functions with respect to parameters θ and λ_S are given by

$$\begin{aligned} \beta^{(\theta)} &= -h^{(1)}(h^{-1}(\check{p})) \frac{\partial h^{-1}(\check{p})}{\partial \theta}, \\ \beta^{(\lambda_S)} &= \lambda_S^{-1} h^{(1)}(h^{-1}(\check{p})) h^{-1}(\check{p}). \end{aligned}$$

The ability to find optimal quantile levels p_1^* and p_2^* to estimate θ depends on whether $\beta^{(\theta)}$ is proportional to a function that does not depend on θ or λ_S , where the constant of proportionality does not depend on quantile level p . Notice that $\beta^{(\theta)}$ is always proportional to a function that does not depend on λ_S , whether or not on θ depends on the generator h . If h is able to satisfy the criteria, then the objective function $I_2(p_1, p_2)$ can be maximized. In the case of the Pareto, Gumbel and exponential-Pareto generators considered in Section 4, this can be achieved. We obtain, up to constant that does not depend on p ,

$$\begin{aligned} \beta^{(\theta)} &\propto \check{p} \cdot \ln \check{p}, & \text{for the Pareto and exponential-Pareto,} \\ \beta^{(\theta)} &\propto \check{p} \cdot \ln \check{p} \cdot \ln(-\ln \check{p}), & \text{for the Gumbel.} \end{aligned}$$

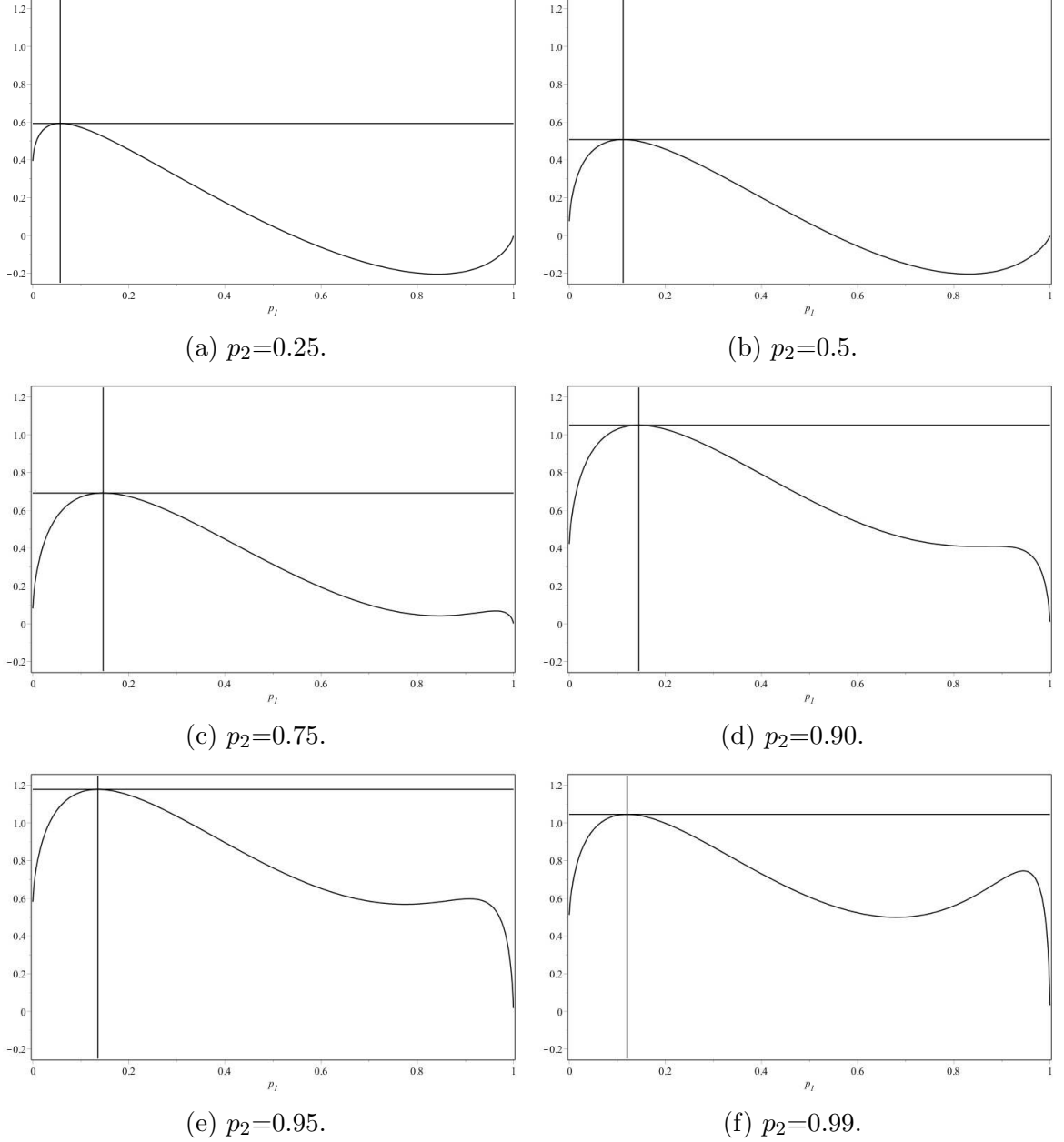
In these examples, we can subsequently determine optimal quantile level p_3^* to estimate λ_S since $\beta^{(\lambda_S)}$ is always proportional to a function that does not depend on λ_S ; whether or not it depends on θ is irrelevant since we can substitute $\widehat{\theta}$ for θ , if required.

The Pareto example has already been investigated in Alai *et al.* (2016), they found optimal p_1^* and p_2^* were given by 0.6385 and 0.9265, respectively. These quantiles would, consequently, serve as optimal for the exponential-Pareto distribution as well. The relevant objective function for the Gumbel generator is given by

$$\begin{aligned} c \cdot I_2(p_1, p_2) &= \frac{(\check{p}_1 \cdot \ln \check{p}_1 \cdot \ln(-\ln \check{p}_1))^2}{p_1} + \frac{(\check{p}_2 \cdot \ln \check{p}_2 \cdot \ln(-\ln \check{p}_2))^2}{\check{p}_2} \\ &\quad + \frac{(\check{p}_2 \cdot \ln \check{p}_2 \cdot \ln(-\ln \check{p}_2) - \check{p}_1 \cdot \ln \check{p}_1 \cdot \ln(-\ln \check{p}_1))^2}{p_2 - p_1}, \end{aligned}$$

for some positive constant c that does not depend on p_1 or p_2 . We find, for the Gumbel generator, that optimal p_1^* and p_2^* are given by 0.1322 and 0.9627, respectively. In Figure 3, we plot $c \cdot I_2$ over p_1 for several values of p_2 ; notice that as p_2 approaches p_2^* , the maximum value of $c \cdot I_2$ increases.

Figure 3: Gumbel, $c \cdot I_2$ versus p_1 for select values of p_2 .



If θ is known or estimated, either with optimal or non-optimal quantiles, λ_S can be optimally estimated for any generator function h . We provide the $\beta^{(\lambda_S)}$ functions for the Pareto, Clayton, and Gumbel generators; note that the Pareto example has already been considered in Alai *et al.* (2016). We have, up to constant that does not depend on p ,

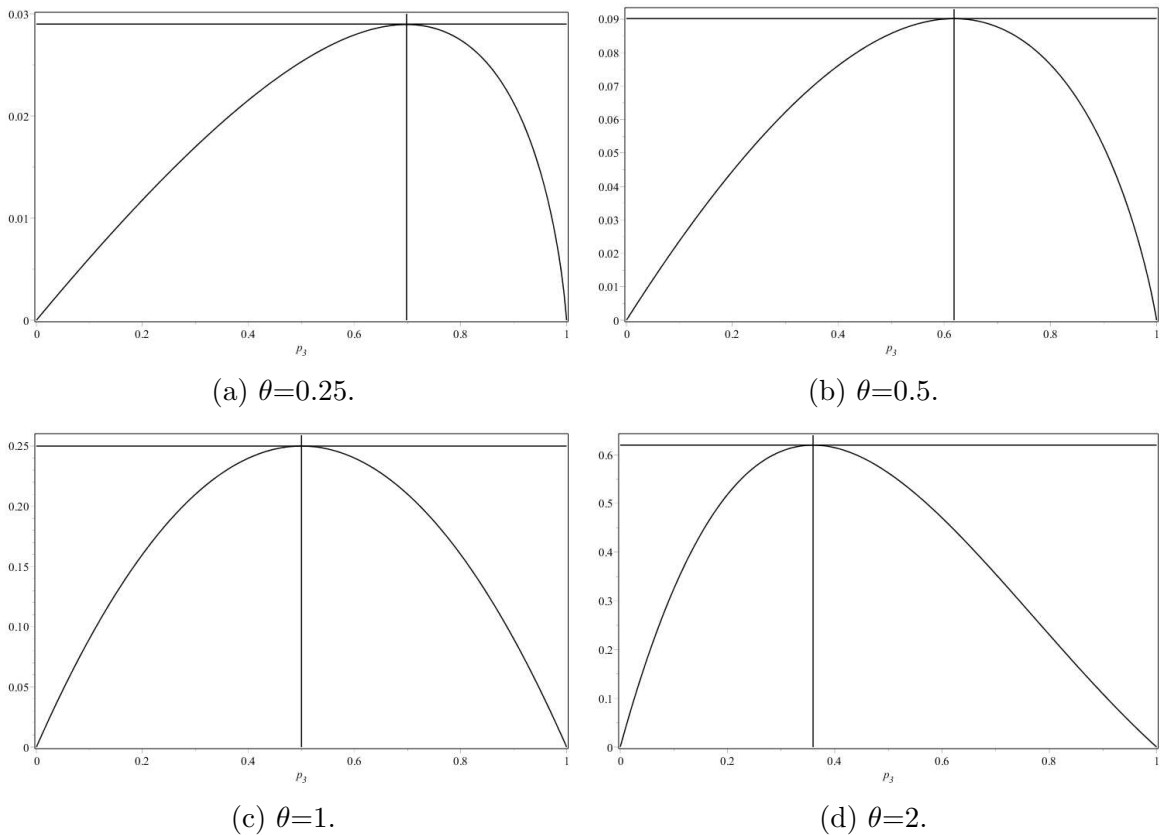
$$\begin{aligned} \beta^{(\lambda_S)} &\propto \check{p} \cdot (1 - \check{p}^\theta), & \text{for the Pareto and Clayton,} \\ \beta^{(\lambda_S)} &\propto \check{p} \cdot \ln \check{p}, & \text{for the Gumbel.} \end{aligned}$$

The above $\beta^{(\lambda_S)}$ can be used to define relevant objective functions $I_1(p)$; since θ is known, we require only one sample quantile to estimate λ_S , hence we use $I_1(p)$. Notice that in the case of the Gumbel generator, θ is not required to determine the optimal quantile level p_3^* , but is required to find estimate λ_S . For the Pareto and Clayton generators, the objective function is given by

$$c \cdot I_1(p_3) = \frac{\check{p}_3(1 - \check{p}_3^\theta)^2}{p_3},$$

for some positive constant c that does not depend on p_3 . In Figure 4, we plot the objective function over p_3 for various levels of θ ; the plots for $\theta = 1$ and $\theta = 2$ are comparable to Figures 3b and 3a, respectively, in Alai *et al.* (2016). Finally, in Figure 5a, we plot p_3^* , for the Pareto and Clayton generators, as a function of θ .

Figure 4: Pareto and Clayton, objective function versus p_3 for select values of θ .

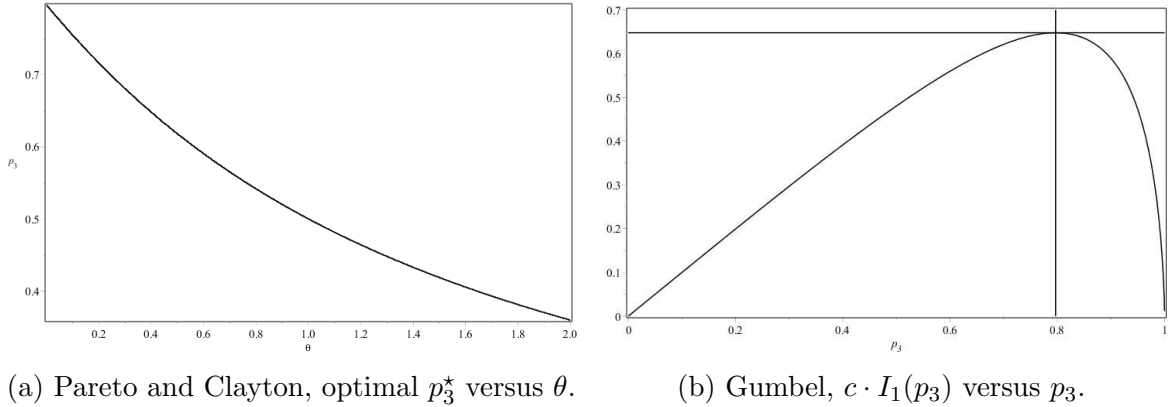


For the Gumbel generator, the objective function is given by

$$c \cdot I_1(p_3) = \frac{\check{p}_3(\ln \check{p}_3)^2}{p_3},$$

for some positive constant c that does not depend on p_3 . Optimal p_3^* is found to be 0.7968; refer to Figure 5b for a plot of $c \cdot I_1(p_3)$.

Figure 5: Optimal p_3^* for the Pareto, Clayton, and Gumbel generators.



5.3 A Numerical Example using the Gumbel Generator

Let ${}_{\tau}X_{(1)} = \min(\mathbf{X}) | \min(\mathbf{X}) > \tau$, where the joint survival function of $\mathbf{X} = (X_1, \dots, X_n)$ is given by

$$P(X_1 > x_1, \dots, X_n > x_n) = h \left(\sum_{i=1}^n \lambda_i x_i \right), \quad x_i \geq 0,$$

for h given by the Gumbel generator function. Consequently, the survival function of $X_{(1)}$ is also given by the Gumbel generator with argument $\lambda_S x$, where $\lambda_S = \sum_{i=1}^n \lambda_i$. In other words, as noted above in Remark 3, $X_{(1)}$ follows a Weibull distribution with scale and shape parameters λ_S and θ , respectively.

Consequently, it is not difficult to simulate samples of ${}_{\tau}X_{(1)}$ and estimate λ_S and θ using sample quantiles. Suppose $\theta = 2$ and $\lambda_S = 1$; refer to Figure 1c for a plot of the survival function. Using these parameter values, it can further be seen from Figure 1c that truncation points of one, two and five represent weak, moderate and strong levels of truncation. In Table 1, we show the estimation results using samples of size 100,000. Each column represents the estimation for one generated sample with truncation point τ , for which four pairs of estimates $\{\theta, \lambda_S\}$ are found. The first, using optimal quantile levels as determined above, i.e. using p_1^* , p_2^* and p_3^* , given by 0.1322, 0.9627 and 0.7968, respectively. The remaining, using some arbitrarily selected quantile levels; the most accurate pair is indicated in bold. It is important to note that each of the four pair of estimates, for a given τ , are based on the same simulated sample. The results of Table 2 and 3 are analogous to Table 1 based on samples of size 1,000, and 10,000,000, respectively. We include the latter for illustrative purposes; we do not anticipate an application with 10,000,000 observations.

From Table 1 it is apparent that a sample size of 100,000 is more than adequate to calibrate the model; although the optimal quantile levels produce the most accurate estimates for $\tau = 0$ and $\tau = 5$, the other quantile levels also perform well. In fact, using quantile levels of 0.5 and 0.95 to estimate θ and 0.95 to estimate λ_S produces the most accurate estimation for $\tau = 1$ and $\tau = 2$; note that this is attributed to random fluctuation since results using a sample size of 10,000,000 always yield the optimal quantile estimates most accurate; see Table 3. Notice that we consider the pair of estimates $\{\hat{\theta}, \hat{\lambda}_S\}$ when determining the most accurate, rather than each estimate in isolation; this is because $\hat{\theta}$ also plays a role in estimating $\hat{\lambda}_S$, rendering the estimates inseparable.

Table 1: Estimation results with a sample size of 100,000.

$\theta = 2; \lambda_S = 1.$	τ	0	1	2	5
$\{p_1^*, p_2^*\}$	$\hat{\theta}$	1.9952	1.9581	1.9933	2.0146
$\{p_3^*\}$	$\hat{\lambda}_S$	1.0075	0.9577	0.9804	1.0221
$\{0.5, 0.95\}$	$\hat{\theta}$	2.0015	1.9834	1.9912	1.9802
$\{0.95\}$	$\hat{\lambda}_S$	1.0178	0.9983	0.9831	0.9572
$\{0.25, 0.75\}$	$\hat{\theta}$	2.0026	1.9565	2.0183	1.9698
$\{0.75\}$	$\hat{\lambda}_S$	1.0103	0.9564	1.0284	0.9365
$\{0.75, 0.95\}$	$\hat{\theta}$	1.9915	1.9755	1.9852	1.9785
$\{0.95\}$	$\hat{\lambda}_S$	1.0067	0.9839	0.9712	0.9534

Table 2: Estimation results with a sample size of 1,000.

$\theta = 2; \lambda_S = 1.$	τ	0	1	2	5
$\{p_1^*, p_2^*\}$	$\hat{\theta}$	2.0113	2.0409	1.9537	2.8449
$\{p_3^*\}$	$\hat{\lambda}_S$	1.0209	1.1160	0.8951	6.4397
$\{0.5, 0.95\}$	$\hat{\theta}$	1.9771	2.0035	1.9177	2.5005
$\{0.95\}$	$\hat{\lambda}_S$	1.0311	0.9768	0.7542	3.2879
$\{0.25, 0.75\}$	$\hat{\theta}$	1.9539	1.8738	1.5142	3.6743
$\{0.75\}$	$\hat{\lambda}_S$	1.0585	0.8496	0.4464	55.9336
$\{0.75, 0.95\}$	$\hat{\theta}$	2.0209	2.1593	2.3160	2.0786
$\{0.95\}$	$\hat{\lambda}_S$	1.0819	1.3081	1.6986	1.1878

In Table 2 we see that the level of truncation plays a role for smaller sample sizes. Although the optimal quantiles produce the most accurate estimation when $\tau = 0$, the optimal quantile estimates for $\tau = 5$ are less impressive. Furthermore, we find a large spread of estimates when $\tau \neq 0$ and hence will need to carefully consider optimal quantile estimation in the presence of truncation in future research. Since optimal quantile levels were chosen based on minimizing the variance of the corresponding estimators, it is no surprise that they are most reliable, if not always most accurate.

Table 3: Estimation results with a sample size of 10,000,000.

$\theta = 2; \lambda_S = 1.$	τ	0	1	2	5
$\{p_1^*, p_2^*\}$	$\hat{\theta}$	2.0005	2.0009	2.0002	2.0053
$\{p_3^*\}$	$\hat{\lambda}_S$	1.0000	1.0017	1.0008	1.0105
$\{0.5, 0.95\}$	$\hat{\theta}$	2.0011	2.0017	2.0029	2.0118
$\{0.95\}$	$\hat{\lambda}_S$	0.9988	1.0041	1.0059	1.0249
$\{0.25, 0.75\}$	$\hat{\theta}$	1.9997	1.9982	1.9983	2.0092
$\{0.75\}$	$\hat{\lambda}_S$	0.9994	0.9974	0.9975	1.0184
$\{0.75, 0.95\}$	$\hat{\theta}$	2.0025	1.9979	2.0009	2.0072
$\{0.95\}$	$\hat{\lambda}_S$	1.0004	0.9971	1.0018	1.0142

6 Conclusion

In this paper, we consider a multivariate distribution based on generator function h . This general approach to constructing a distribution includes, as a special case, the multivariate Pareto used to study joint lifetimes in Alai *et al.* (2016). There is a link between the family of distributions we consider and Archimedean copula generators, which we use to supply examples based on well-known generators from the Clayton, Gumbel, Frank and Ali-Mikhail-Haq copulas. Furthermore, we suggest a new distribution given by combining the Pareto and exponential structure. We provide moment and quantile results in the presence of multivariate truncation and revisit known estimation techniques in order that they may be used to gain insight in the pricing and risk management of life-benefit products such as bulk annuities. Finally, we consider optimal quantile selection for quantile-based estimation procedures and provide a numerical example using a multivariate Weibull distribution.

Acknowledgements

The authors would like to acknowledge the financial support of ARC Linkage Grant Project LP0883398 Managing Risk with Insurance and Superannuation as Individuals Age with industry partners PwC, APRA and the World Bank as well as the support of the Australian Research Council Centre of Excellence in Population Ageing Research (project number CE110001029).

References

- Alai, D. H., Landsman, Z., and Sherris, M. (2016). Modelling lifetime dependence for older ages using a multivariate Pareto distribution. *Insurance: Mathematics and Economics*, **70**, 272–285.
- Arnold, B. C. (1985). *Pareto Distribution*. John Wiley & Sons, New York.
- Asimit, A. V., Furman, E., and Vernic, R. (2010). On a multivariate Pareto distribution. *Insurance: Mathematics and Economics*, **46**(2), 308–316.
- Caramellino, L. and Spizzichino, F. (1994). Dependence and aging properties of lifetimes with Schur-constant survival functions. *Probability in the Engineering and Informational Sciences*, **8**, 103–111.
- Chiragiev, A. and Landsman, Z. (2007). Multivariate Pareto portfolios: TCE-based capital allocation and divided differences. *Scandinavian Actuarial Journal*, **2007**(4), 261–280.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.
- Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, **2**(1), 1–25.
- Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, **88**(423), 1034–1043.
- Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**(3), 543–552.
- Hendriks, H. and Landsman, Z. (2017). A generalization of multivariate Pareto distributions: Tail risk measures, divided differences, asymptotics. *Scandinavian Actuarial Journal*, **2017**(9), 785–803.
- Hofert, M., Mächler, M., and McNeil, A. J. (2012). Archimedean copulas in high dimensions: Estimators and numerical challenges motivated by financial applications. *Journal de la Société Française de Statistique*, **154**(1), 25–63.
- Landsman, Z. (1996). Sample quantiles and additive statistics: Information, sufficiency, estimation. *Journal of Statistical Planning and Inference*, **52**, 93–108.
- Mardia, K. V. (1962). Multivariate Pareto distributions. *The Annals of Mathematical Statistics*, **33**(3), 1008–1015.
- McNeil, A. J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and ℓ_1 -norm symmetric distributions. *The Annals of Statistics*, **37**, 3059–3097.
- McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton: Princeton University Press.

- Nelsen, R. B. (1999). *An Introduction to Copulas*. Springer-Verlag, New York.
- Nelsen, R. B. (2005). Some properties of Schur-constant survival models and their copulas. *Brazilian Journal of Probability and Statistics*, **19**, 179–190.
- Vernic, R. (2011). Tail conditional expectation for the multivariate Pareto distribution of the second kind: Another approach. *Methodology and Computing in Applied Probability*, **13**(1), 121–137.
- Williamson, R. (1956). Multiply monotone functions and their Laplace transforms. *Duke Mathematical Journal*, **23**(2), 189–207.