T E C H N I C A L
R E P O R T

11007

An M-estimator for tail dependence in arbitrary dimensions

EINMAHL, J.H.J., KRAJINA, A. and J. SEGERS

Σ T A T

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

An M-Estimator For Tail Dependence In Arbitrary Dimensions

John H.J. Einmahl

Tilburg University and CentER, Tilburg, The Netherlands.

Andrea Krajina

Institute for Mathematical Stochastics, University of Göttingen, Germany.

Johan Segers

ISBA, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Summary. Consider a random sample in the max-domain of attraction of a multivariate extreme value distribution such that the dependence structure of the attractor belongs to a parametric model. A new estimator for the unknown parameter is defined as the value that minimises the distance between a vector of weighted integrals of the tail dependence function and their empirical counterparts. The minimisation problem has, with probability tending to one, a unique, global solution. The estimator is consistent and asymptotically normal. The spectral measures of the tail dependence models to which the method applies can be discrete or continuous. Examples demonstrate the applicability and the performance of the method.

Keywords: asymptotic statistics, factor model, M-estimation, multivariate extremes, tail dependence.

1. Introduction

As the number of variables increases, modelling tail dependence becomes more complex. For instance, in dimension d there are $d(d-1)/2$ bivariate marginals, which in general can be different up to some consistency requirements. Therefore it is customary to model the tail dependence parametrically. The interest in parametric tail dependence models exists since the early sixties of the 20th century, see for example Gumbel (1960), but new models are still being proposed, see for instance Cooley et al. (2010).

Let F be a continuous d-variate distribution function with marginal distribution functions $F_1, \ldots, F_d$ and quantile functions $F_1^{-1}, \ldots, F_d^{-1}$. Rather than working with a particular parametric model for F, we only assume that the stable tail dependence function $l : [0,\infty)^d \to [0,\infty)$, defined by

$$l(x) = \lim_{t \downarrow 0} t^{-1} \big\{ 1 - F\big(F_1^{-1}(1-tx_1), \ldots, F_d^{-1}(1-tx_d)\big) \big\}, \qquad x \in [0,\infty)^d, \tag{1.1}$$

belongs to some parametric family, $l \in \{l(\cdot;\theta) \colon \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}^p$. The existing estimators of such a θ are all likelihood based and as such, apply only to d times differentiable functions l; see Coles and Tawn (1991); Joe et al. (1992); Smith (1994); Ledford and Tawn (1996); de Haan et al. (2008); Guillotte et al. (2011). Although some of these estimators have been used in dimensions higher than two, their asymptotic properties have been derived in the bivariate case only, and not even for all estimators. A step towards the estimation of high dimensional tail dependence using mixtures of Dirichlet distributions has been proposed in Boldi and Davison (2007).

In Einmahl et al. (2008) the method of moments estimator of the parametric bivariate stable tail dependence function was introduced. Notably, this method does not require the function l to be differentiable. Here we extend that estimator in two directions. First, we consider models in arbitrary dimensions. Second we extend the method of moments estimation to general M-estimation by allowing for more estimating equations than parameters. The first extension addresses an important issue, since the estimation of the tail dependence structure in higher dimensions is a challenge.

If $\theta \in \Theta \subseteq \mathbb{R}^p$ is the unknown parameter, $g\colon [0,1]^d \to \mathbb{R}^q$, $q \geqslant p$, is an auxiliary function and if $\hat{l}_n$ in equation (3.1) below is the nonparametric estimator of l, we define $\hat{\theta}_n$, the M-estimator of θ, as the minimiser of the Euclidean distance in $\mathbb{R}^q$ between

$$\int_{[0,1]^d} g(x)\,\hat{l}_n(x)\,\mathrm{d}x \quad \text{and} \quad \int_{[0,1]^d} g(x)\,l(x;\theta)\,\mathrm{d}x.$$

The unique, global minimiser exists with probability tending to one under minimal conditions. This minimiser is a consistent and asymptotically normal estimator of θ. In passing, the asymptotic normality of $\hat{l}_n$ in arbitrary dimensions is established, which is a result of independent interest.

The absence of smoothness assumptions on l makes it possible to estimate the tail dependence structure of factor models like $X = (X_1, \ldots, X_d)$, with

$$X_j = \sum_{i=1}^{r} a_{ij} Z_i + \varepsilon_j, \quad j = 1, \ldots, d, \tag{1.2}$$

consisting of the following ingredients: nonnegative factor loadings a_{ij} and independent, heavy-tailed random variables Z_i called factors; independent random variables ε_j whose tails are lighter than the ones of the factors and which are independent of them. This kind of factor model is often used in finance, for example in modelling market or credit risk; see Fama and French (1993); Malevergne and Sornette (2004); Geluk et al. (2007).

The organisation of the paper is as follows. The basics of the tail dependence structures in multivariate models are presented in Section 2. The M-estimator is defined in Section 3. Section 4 contains the main theoretical results: consistency and asymptotic normality of the M-estimator, and some consequences of the asymptotic normality result that can be used for construction of confidence regions and for testing. This section also contains the asymptotic normality result for $\hat{l}_n$. In Section 5 we apply the M-estimator to the well-known logistic stable tail dependence function. The tail dependence structure of factor models is studied in Section 6. Both models are illustrated with simulated and real data. The proofs are deferred to Section 7.

2. Tail dependence

We will write points in $\mathbb{R}^d$ as $x := (x_1, \ldots, x_d)$ and random vectors as $X_i := (X_{i1}, \ldots, X_{id})$, for $i = 1, \ldots, n$. Let $X_1, \ldots, X_n$ be independent random vectors in $\mathbb{R}^d$ with common continuous distribution function F and marginal distribution functions $F_1, \ldots, F_d$. We assume that F has a *stable tail dependence function* l, that is, we assume that for all $x = (x_1, \ldots, x_d) \in [0, \infty)^d$ the following limit exists:

$$\lim_{t \downarrow 0} t^{-1} \mathbb{P}\left(1 - F_1(X_{11}) \leqslant t x_1 \text{ or } \ldots \text{ or } 1 - F_d(X_{1d}) \leqslant t x_d\right) = l(x). \tag{2.1}$$

The function $l : [0,\infty)^d \to [0,\infty)$ has the following properties:

- $\max\{x_1,\dots,x_d\} \leqslant l(x) \leqslant x_1+\dots+x_d$ for all $x \in [0,\infty)^d$; in particular $l(z,0,\dots,0) = \dots = l(0,\dots,0,z) = z$ for all $z \geqslant 0$;
- l is convex; and
- l is homogeneous of order one: $l(tx_1,\dots,tx_d) = t\,l(x_1,\dots,x_d)$, for all $t>0$ and all $x \in [0,\infty)^d$.

The function l is connected to the function V in Coles and Tawn (1991) through $l(x) = V(1/x_1,\dots,1/x_d)$ for $x \in (0,\infty)^d$.

Let $\Delta_{d-1} := \{w \in [0,1]^d : w_1+\dots+w_d = 1\}$ be the unit simplex in $\mathbb{R}^d$. A finite Borel measure H on Δ_{d-1} satisfying the d moment conditions

$$\int_{\Delta_{d-1}} w_j H(\mathrm{d}w) = 1, \quad j=1,\dots,d, \tag{2.2}$$

is called a *spectral* or *angular measure*. It follows from the moment conditions that H/d is a probability measure. There is a one-to-one correspondence between the stable tail dependence function and the spectral measure: it holds that there exists a unique spectral measure H such that

$$l(x) = \int_{\Delta_{d-1}} \max_{j=1,\dots,d}\{w_j x_j\} H(\mathrm{d}w). \tag{2.3}$$

It can be shown that there exists a measure Λ on $[0,\infty]^d \setminus \{(\infty,\dots,\infty)\}$ such that

(1) $l(x) = \Lambda\left(\{u \in [0,\infty]^d : u_1 \leqslant x_1 \text{ or } \dots \text{ or } u_d \leqslant x_d\}\right)$,

(2) $\Lambda(tA) = t\Lambda(A)$, for any $t>0$ and any Borel set $A \subset [0,\infty]^d \setminus \{(\infty,\dots,\infty)\}$, with $tA := \{tx : x \in A\}$,

see for example Resnick (1987); Beirlant et al. (2004); de Haan and Ferreira (2006). The measure Λ is called the *exponent measure* and it is yet another way of defining the tail dependence structure. Property (1) connects the exponent measure to the function l. If μ is the measure Λ after the transformation $(x_1,\dots,x_d) \mapsto (1/x_1,\dots,1/x_d)$, the relationship between the spectral measure H and the exponent measure Λ (and μ) is given by

$$H(B) = \mu\left(\left\{x \in [0,\infty)^d \colon \textstyle\sum_{j=1}^d x_j \geqslant 1,\ x/\sum_{j=1}^d x_j \in B\right\}\right),$$

for any Borel set B on Δ_{d-1}. By property (2) we get that for any $t>0$ and any Borel set B on Δ_{d-1},

$$\frac{1}{t} H(B) = \mu\left(\left\{x \in [0,\infty)^d \colon \textstyle\sum_{j=1}^d x_j \geqslant t,\ x/\sum_{j=1}^d x_j \in B\right\}\right),$$

which is a version of the spectral decomposition of the exponent measure, see de Haan and Resnick (1977) or Resnick (1987).

The right-hand partial derivatives of l always exist; indeed, by bounded convergence it follows that for $j=1,\dots,d$, as $h \downarrow 0$,

$$\begin{aligned}
&\frac{1}{h}\big(l(x_1,\dots,x_{j-1},x_j+h,x_{j+1},\dots,x_d) - l(x_1,\dots,x_{j-1},x_j,x_{j+1},\dots,x_d)\big) \\
&\quad = \int_{\Delta_{d-1}} \frac{1}{h}\Big(\max\{w_j x_j + w_j h, \max_{s\neq j}\{w_s x_s\}\} - \max\{w_j x_j, \max_{s\neq j}\{w_s x_s\}\}\Big) H(\mathrm{d}w) \\
&\qquad\qquad \to \int_{\Delta_{d-1}} w_j \mathbf{1}\{w_j x_j \geqslant \max_{s\neq j}\{w_s x_s\}\} H(\mathrm{d}w).
\end{aligned} \tag{2.4}$$

Similarly, the left-hand partial derivatives exist for all $x \in (0,\infty)^d$. By convexity, the function l is almost everywhere continuously differentiable, with its gradient vector of (the right-hand) partial derivatives as in (2.4).

3. Estimation

Let R_i^j denote the rank of X_{ij} among $X_{1j},\dots,X_{nj}$, $i=1,\dots,n$, $j=1,\dots,d$. For $k \in \{1,\dots,n\}$, define a nonparametric estimator of l by

$$\hat{l}_n(x) := \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{R_i^1 > n + \frac{1}{2} - kx_1 \text{ or } \dots \text{ or } R_i^d > n + \frac{1}{2} - kx_d\right\}; \tag{3.1}$$

see Huang (1992) and Drees and Huang (1998) for the bivariate case. When we study asymptotic properties of this estimator, $k = k_n$ is an intermediate sequence, that is, $k \to \infty$ and $k/n \to 0$ as $n \to \infty$.

In the literature, the stable tail dependence function is often modelled parametrically. We impose that the stable tail dependence function l belongs to some parametric family $\{l(\,\cdot\,;\theta) : \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^p$, $p \geqslant 1$. Note that this is still a large, flexible model since there is no restriction on the marginal distributions and also the copula is only modelled through l, see (1.1). We propose an M-estimator of θ. Let $q \geqslant p$. Let $g \equiv (g_1,\dots,g_q)^T : [0,1]^d \to \mathbb{R}^q$ be a column vector of integrable functions such that $\varphi : \Theta \to \mathbb{R}^q$ defined by

$$\varphi(\theta) := \int_{[0,1]^d} g(x) l(x;\theta)\,\mathrm{d}x \tag{3.2}$$

is a homeomorphism between Θ and its image $\varphi(\Theta)$. Let θ_0 denote the true parameter value. The M-estimator $\hat{\theta}_n$ of θ_0 is defined as a minimiser of the criterion function

$$Q_{k,n}(\theta) = \|\varphi(\theta) - \int g\hat{l}_n\|^2 = \sum_{m=1}^{q}\left(\int_{[0,1]^d} g_m(x)\left(\hat{l}_n(x) - l(x;\theta)\right)\mathrm{d}x\right)^2,$$

where $\|\cdot\|$ is the Euclidean norm. In other words, if $\hat{Y}_n = \arg\min_{y\in\varphi(\Theta)} \|y - \int g\hat{l}_n\|$, then $\hat{\theta}_n \in \varphi^{-1}(\hat{Y}_n)$. Later we show that $\hat{\theta}_n$ is, with probability tending to one, unique.

4. Results

Let $\hat{\Theta}_n$ be the set of minimisers of $Q_{k,n}$,

$$\hat{\Theta}_n := \underset{\theta\in\Theta}{\arg\min}\, \|\varphi(\theta) - \textstyle\int g\hat{l}_n\|^2.$$

Note that $\hat{\Theta}_n$ may be empty or may contain more than one element. We show that under suitable conditions, the minimiser exists, that it is unique with probability tending to one, and that it is a consistent and asymptotically normal estimator of θ_0. In addition, we show that the nonparametric estimator $\hat{l}_n$ in (3.1) is asymptotically normal.

4.1. Notation

Let W_Λ be a mean-zero Wiener process indexed by Borel sets of $[0,\infty]^d \setminus \{(\infty,\dots,\infty)\}$ with "time" Λ: its covariance structure is given by

$$\mathbb{E}[W_\Lambda(A_1)\, W_\Lambda(A_2)] = \Lambda(A_1 \cap A_2), \tag{4.1}$$

for any two Borel sets A_1 and A_2 in $[0,\infty]^d \setminus \{(\infty,\dots,\infty)\}$. Define

$$W_l(x) := W_\Lambda(\{u \in [0,\infty]^d \setminus \{(\infty,\dots,\infty)\} : u_1 \leqslant x_1 \text{ or } \dots \text{ or } u_d \leqslant x_d\}). \tag{4.2}$$

Let $W_j,\ j = 1,\dots,d$, be the marginal processes

$$W_j(x_j) := W_l(0,\dots,0,x_j,0,\dots,0), \qquad x_j \geqslant 0. \tag{4.3}$$

Define l_j to be the right-hand partial derivative of l with respect to x_j, where $j = 1,\dots,d$, see (2.4); if l is differentiable, l_j is equal to the corresponding partial derivative of l. Write

$$B(x) := W_l(x) - \sum_{j=1}^{d} l_j(x)\, W_j(x_j), \qquad\qquad \tilde{B} := \int_{[0,1]^d} g(x)\, B(x)\,\mathrm{d}x. \tag{4.4}$$

The distribution of $\tilde{B}$ is zero-mean Gaussian with covariance matrix

$$\begin{aligned}\Sigma &:= \mathbb{E}\left[\int_{[0,1]^d} g(x)\, B(x)\,\mathrm{d}x \cdot \int_{[0,1]^d} g(y)^T\, B(y)\,\mathrm{d}y\right] \\ &= \iint_{([0,1]^d)^2} \mathbb{E}[B(x)B(y)]\, g(x)\, g(y)^T\,\mathrm{d}x\,\mathrm{d}y \in \mathbb{R}^{q\times q}.\end{aligned} \tag{4.5}$$

Note that if l is parametric, Σ depends on the parameter, that is $\Sigma = \Sigma(\theta)$.

Let $\nabla Q_{k,n}(\theta) \in \mathbb{R}^{p\times 1}$ be the gradient vector of $Q_{k,n}$ at θ; for every $x \in [0,1]^d$ let $\nabla l(x;\theta) \in \mathbb{R}^{p\times 1}$ be the gradient vector of $l(x;\cdot)$ in θ; let $\dot\varphi(\theta) \in \mathbb{R}^{q\times p}$ be the total derivative of φ at θ; and put

$$V(\theta) := 4\,\dot\varphi(\theta)^T\,\Sigma(\theta)\,\dot\varphi(\theta) \in \mathbb{R}^{p\times p}.$$

Further let $\mathcal{H}_{k,n}(\theta) \in \mathbb{R}^{p\times p}$ denote the Hessian matrix of $Q_{k,n}$ in θ. Let $\mathcal{H}(\theta)$ be the deterministic, symmetric $p\times p$ matrix whose (i,j)-th element, $i,j \in \{1,\dots,p\}$, is equal to

$$(\mathcal{H}(\theta))_{ij} = 2\left(\frac{\partial}{\partial\theta_i}\varphi(\theta)\right)^T \left(\frac{\partial}{\partial\theta_j}\varphi(\theta)\right) - 2\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\varphi(\theta)\right)^T \big(\varphi(\theta_0) - \varphi(\theta)\big).$$

Observe that

$$\mathcal{H}(\theta_0) = 2\dot\varphi(\theta_0)^T\dot\varphi(\theta_0),$$

and define

$$M(\theta) := \big(\dot\varphi(\theta)^T\dot\varphi(\theta)\big)^{-1}\,\dot\varphi(\theta)^T\,\Sigma(\theta)\,\dot\varphi(\theta)\,\big(\dot\varphi(\theta)^T\dot\varphi(\theta)\big)^{-1} \in \mathbb{R}^{p\times p}. \tag{4.6}$$

4.2. Results

We state the asymptotic results for the M-estimator, $\hat{\theta}_n$, and the asymptotic normality of $\hat{l}_n$. The proofs can be found in Section 7. We require subsets of the following list of conditions:

(C1) θ_0 is in the interior of the parameter space, φ is twice continuously differentiable and $\dot{\varphi}(\theta_0)$ is of full rank;

(C2) $t^{-1}\mathbb{P}\left(1 - F_1(X_{11}) \leqslant tx_1 \text{ or } \ldots \text{ or } 1 - F_d(X_{1d}) \leqslant tx_d\right) - l(x) = O(t^\alpha)$, uniformly in $x \in \Delta_{d-1}$ as $t \downarrow 0$, for some $\alpha > 0$;

(C3) $k = o\big(n^{2\alpha/(1+2\alpha)}\big)$, for the positive number α of (C2) and $k \to \infty$ as $n \to \infty$;

(C4) For all $j = 1, \ldots, d$, the first-order partial derivative of l with respect to x_j exists and is continuous on the set of points x such that $x_j > 0$.

Theorem 4.1 (Existence, uniqueness and consistency of $\hat{\theta}_n$). *Let $g : [0,1]^d \to \mathbb{R}^q$ be integrable.*

(i) If φ is a homeomorphism from Θ to $\varphi(\Theta)$ and if there exists $\varepsilon_0 > 0$ such that the set $\{\theta \in \Theta : \|\theta - \theta_0\| \leqslant \varepsilon_0\}$ is closed, then for every ε such that $\varepsilon_0 \geqslant \varepsilon > 0$, as $n \to \infty$,

$$\mathbb{P}\left(\hat{\Theta}_n \neq \varnothing \text{ and } \hat{\Theta}_n \subseteq \{\theta \in \Theta : \|\theta - \theta_0\| \leqslant \varepsilon\}\right) \to 1.$$

(ii) If in addition to the assumptions of (i), condition (C1) holds, then, with probability tending to one, $Q_{k,n}$ has a unique minimiser $\hat{\theta}_n$. Hence

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0, \quad \text{as } n \to \infty.$$

In part (i) of this theorem we assume that the set $\{\theta \in \Theta : \|\theta - \theta_0\| \leqslant \varepsilon_0\}$ is closed for some $\varepsilon > 0$. This is a generalisation of the usual assumption that Θ is open or closed, and includes a wider range of possible parameter spaces.

We prove the asymptotic normality of $\hat{l}_n$. This result is of independent interest and can be found in the literature for $d = 2$ only and under stronger smoothness conditions on l: see Huang (1992), Drees and Huang (1998), and de Haan and Ferreira (2006). Here it is a necessary part of the proof for asymptotic normality of $\hat{\theta}_n$. Note that under assumption (C4), the process B in (4.4) is continuous, although l_j may be discontinuous at points x such that $x_j = 0$.

The result is stated in an approximation setting, where $\hat{l}_n$ and B are defined on the same probability space obtained by a Skorohod construction. The random quantities involved are only in distribution equal to the original ones, but for convenience this is not expressed in the notation.

Theorem 4.2 (Asymptotic normality of $\hat{l}_n$ in arbitrary dimensions). *If conditions (C2), (C3), and (C4) hold, then for every $T > 0$, as $n \to \infty$,*

$$\sup_{x \in [0,T]^d} \left| \sqrt{k}\left(\hat{l}_n(x) - l(x)\right) - B(x) \right| \xrightarrow{\mathbb{P}} 0. \tag{4.7}$$

Theorem 4.3 (Asymptotic normality of $\hat{\theta}_n$). *If in addition to the assumptions of Theorem 4.1(i), conditions (C1), (C2), and (C3) hold, then as $n \to \infty$,*

$$\sqrt{k}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, M(\theta_0)). \tag{4.8}$$

The following consequence of Theorem 4.3 can be used for the construction of confidence regions.

Corollary 4.4. *If in addition to the conditions of Theorem 4.3, the map $\theta \mapsto H_\theta$ is weakly continuous at θ_0 and if the matrix $M(\theta_0)$ is non-singular, then as $n \to \infty$,*

$$k(\hat{\theta}_n - \theta_0)^T M(\hat{\theta}_n)^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \chi^2_p. \tag{4.9}$$

Let $1 \leqslant r < p$ and $\theta = (\theta_1, \theta_2) \in \Theta \subset \mathbb{R}^p$, where $\theta_1 \in \mathbb{R}^{p-r}$, $\theta_2 \in \mathbb{R}^r$. We want to test $\theta_2 = \theta_2^*$ against $\theta_2 \neq \theta_2^*$, where θ_2^* corresponds to a submodel. Denote $\hat{\theta}_n = (\hat{\theta}_{1n}, \hat{\theta}_{2n})$, and let $M_2(\theta)$ be the $r \times r$ matrix corresponding to the lower right corner of M, as below,

$$M = \left(\begin{array}{c|c} \cdots & \cdots \\ \hline \cdots & M_2 \end{array} \right) \in \mathbb{R}^{p \times p}. \tag{4.10}$$

Corollary 4.5 (Test). *If the assumptions of Corollary 4.4 are satisfied, and $\theta_0 = (\theta_1, \theta_2^*) \in \Theta$ for some θ_1, then as $n \to \infty$,*

$$k(\hat{\theta}_{2n} - \theta_2^*)^T M_2(\hat{\theta}_{1n}, \theta_2^*)^{-1}(\hat{\theta}_{2n} - \theta_2^*) \xrightarrow{d} \chi^2_r. \tag{4.11}$$

The above result can be used for testing for a submodel. For example, we could test for the symmetric logistic model within the asymmetric logistic one, see Section 5.

Remark 4.6. The matrices M and M_2 are needed for the computation of the confidence regions and the test statistics. However, computing these matrices can be challenging. To compute M, we first need the $q \times p$ matrix $\dot{\varphi}(\theta)$, whose (i,j)-th element is given by $\int g_i(x)(\partial/\partial\theta_j)l(x;\theta)\mathrm{d}x$. The expression itself will depend on the model in use, but usually the (right-hand) partial derivatives of l can be computed explicitly, whereas the integral is to be computed numerically in most cases. Secondly, we need to calculate the covariance of the process $\tilde{B}$. We see from (4.5) that the most difficult part will be the expression $\mathbb{E}[B(x)B(y)]$. It holds that

$$\begin{aligned}\mathbb{E}[B(x)B(y)] = \mathbb{E}[W_l(x)W_l(y)] - \sum_{j=1}^d l_j(y)\,\mathbb{E}[W_l(x)W_j(y_j)] - \sum_{i=1}^d l_i(x)\,\mathbb{E}[W_l(y)W_i(x_i)] \\ + \sum_{i=1}^d \sum_{j=1}^d l_i(x)l_j(y)\,\mathbb{E}[W_i(x_i)W_j(y_j)].\end{aligned}$$

Using (4.1), (4.2), (4.3), and the relation between Λ and l, we can express this in l and its partial derivatives. Numerical integration is then performed to obtain Σ.

5. Example 1: Logistic model

The multivariate logistic distribution function with standard Fréchet margins is defined by

$$F(x_1,\dots,x_d;\theta) = \exp\left\{-\left(\textstyle\sum_{j=1}^d x_j^{-1/\theta}\right)^\theta\right\},$$

for $x_1 > 0,\dots,x_d > 0$ and $\theta \in [0,1]$, with the proper limit interpretation for $\theta = 0$. The corresponding stable tail dependence function is given by

$$l(x_1,\dots,x_d;\theta) = \left(x_1^{1/\theta} + \cdots + x_d^{1/\theta}\right)^\theta. \tag{5.1}$$

Introduced in Gumbel (1960), it is one of the oldest parametric models of tail dependence.

Simulation study: Five-dimensional logistic model. We simulate 500 samples of size $n = 3000$ from a five-dimensional logistic distribution function with $\theta_0 = 0.5$. We obtain $\hat\theta_n$, the M-estimator of θ_0, by choosing $g_1 \equiv 1$ and $g_2(x) = 2^5 x_1 \cdot \cdots \cdot x_5$. The bias and the Root Mean Squared Error (RMSE) of this estimator are shown in the upper panels of Figure 1.

Also, we consider the estimation of $l(1,1,1,1,1;\theta)$, based on this M-estimator $\hat\theta_n$. From (5.1) it follows that $l(1,1,1,1,1;\theta) = 5^\theta$. The estimator of this quantity is then $5^{\hat\theta_n}$. Since $\theta_0 = 0.5$, the true parameter is $\sqrt{5}$. We compare the bias and the RMSE of this estimator and of the nonparametric estimator $\hat l_n(1,1,1,1,1)$, see (3.1). Figure 1, lower panels, shows that the M-estimator performs better than the nonparametric estimator for almost every k.

Real data: Testing and estimation. We use the bivariate Loss-ALAE data set, consisting of 1500 insurance claims, comprising losses and allocated loss adjustment expenses, see Frees and Valdez (1998). The scatterplots of the data and their joint ranks are shown in Figure 2. We consider the asymmetric logistic model described below for their tail dependence function and we test whether a smaller symmetric logistic model suffices to describe the tail dependence of these data. The asymmetric logistic tail dependence function was introduced in Tawn (1988) as an extension of the logistic model. In dimension $d = 2$ it is given by

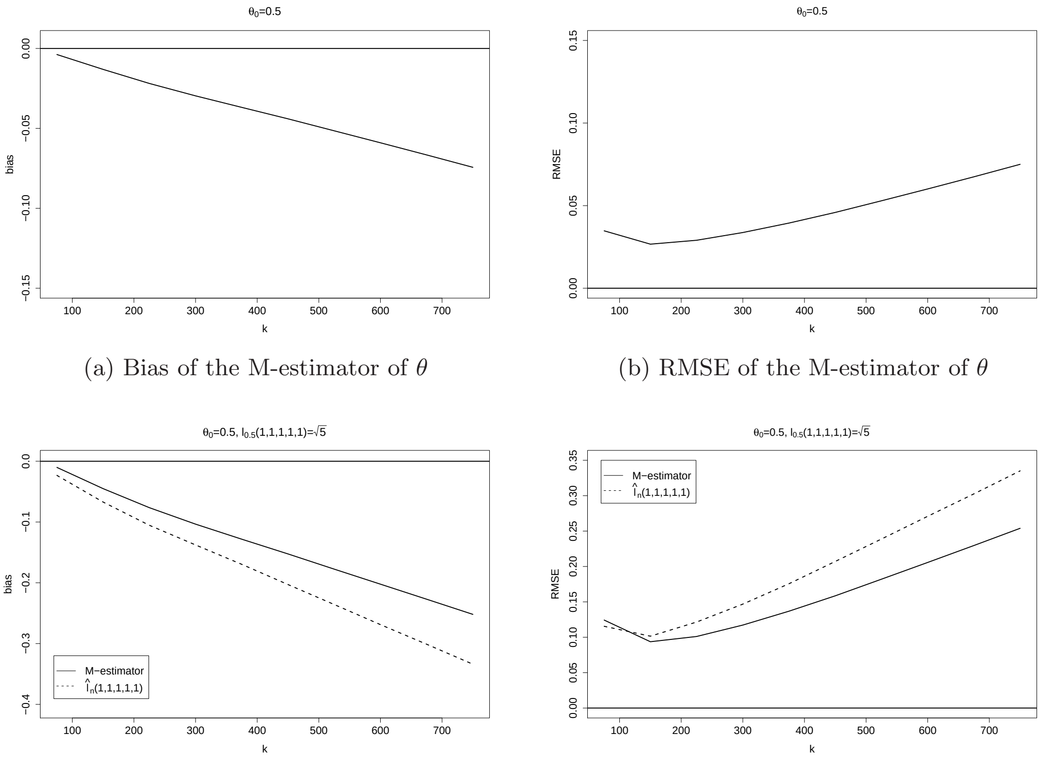
$$l(x,y;\theta,\psi_1,\psi_2) = (1-\psi_1)x + (1-\psi_2)y + \left((\psi_1 x)^{1/\theta} + (\psi_2 y)^{1/\theta}\right)^\theta, \tag{5.2}$$

with the dependence parameter $\theta \in [0,1]$ and the asymmetry parameters $\psi_1,\psi_2 \in [0,1]$. This model yields a spectral measure H with atoms at $(1,0)$ and $(0,1)$ whenever $\psi_1 < 1$ and $\psi_2 < 1$. When $\psi_1 = \psi_2 =: \psi$, we have the symmetric tail dependence function

$$l(x,y;\theta,\psi) = (1-\psi)(x+y) + \psi\left(x^{1/\theta} + y^{1/\theta}\right)^\theta. \tag{5.3}$$

For the given data, we test whether the use of this symmetric model is justified, as opposed to the wider asymmetric logistic model. Setting $\eta_1 := (\psi_1+\psi_2)/2 \in [0,1]$ and $\eta_2 := (\psi_1-\psi_2)/2 \in [-1/2,1/2]$, we reparametrize the model in (5.2) so that testing for symmetry amounts to testing whether $\eta_2 = 0$. By Corollary 4.5, the test statistic is given by

$$S_n := \frac{k\,\hat\eta_2^2}{M_2(\hat\theta,\hat\eta_1,0)}.$$



(a) Bias of the M-estimator of θ

(b) RMSE of the M-estimator of θ

(c) Bias of the estimators of $l(1,1,1,1,1;0.5)$

(d) RMSE of the estimators of $l(1,1,1,1,1;0.5)$

Fig. 1: Logistic model, $d = 5$, $\theta_0 = 0.5$, $l(1,1,1,1,1;\theta_0) = \sqrt{5}$.

The table below shows the obtained values of S_n for the Loss-ALAE data for selected values of k:

k	50	100	150	200	250
S_n	0.041	0.139	0.294	0.477	0.681

Since the critical value is 3.84, the null hypothesis is clearly not rejected. Hence we adopt the symmetric tail dependence model (5.3) and we compute the M-estimates of $(\theta, \eta_1) = (\theta, \psi)$, the auxiliary functions being $g_1(x,y) = x$ and $g_2(x,y) = 2(x+y)$. For $k = 150$, we obtain $(\hat{\theta}, \hat{\psi}) = (0.65, 0.95)$ with estimated standard errors 0.032 for $\hat{\theta}$ and 0.014 for $\hat{\psi}$.

6. Example 2: Factor model

Consider the r-factor model, $r \in \mathbb{N}$, in dimension d: $X' = (X'_1, \ldots, X'_d)$ and

$$X'_j = \sum_{i=1}^{r} a_{ij} Z_i + \varepsilon_j, \qquad j \in \{1, \ldots, d\}, \tag{6.1}$$

with Z_i independent Fréchet(ν) random variables, ε_j independent random variables which are lighter tailed than the factors and independent of them, $\nu > 0$, and a_{ij} nonnegative

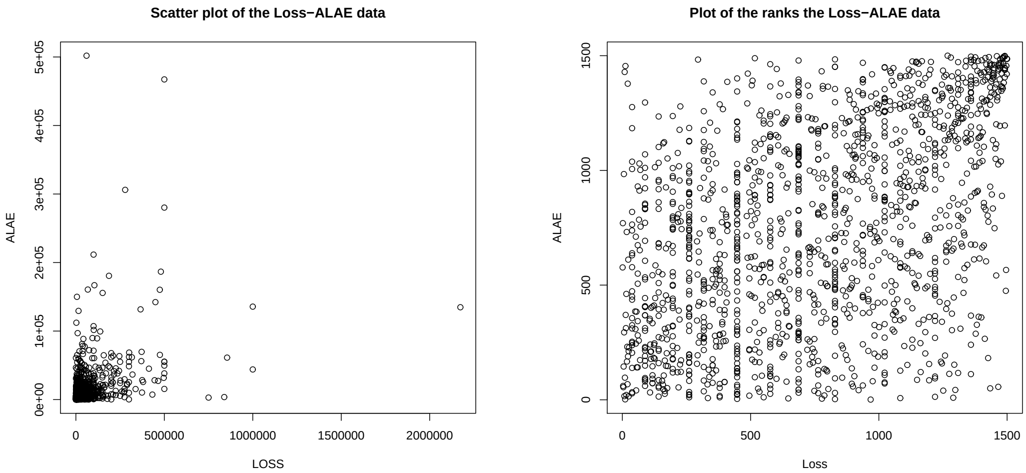


Fig. 2: The insurance claims Loss-ALAE data.

constants such that $\sum_j a_{ij} > 0$ for all i. Factor models of this type are common in various applications, for examples in finance see Fama and French (1993); Malevergne and Sornette (2004); Geluk et al. (2007). However, for the purpose of studying the tail properties, it is more convenient to consider the (max) factor model: $X = (X_1, \ldots, X_d)$ and

$$X_j = \max_{i=1,\ldots,r} \{a_{ij} Z_i\}, \qquad j \in \{1, \ldots, d\}, \tag{6.2}$$

with a_{ij} and Z_i as above. Note that X' and X have the same tail dependence function l. Let $W_i = Z_i^\nu$, $i = 1, \ldots, r$, and note that the W_i are standard Fréchet random variables. Define a d-dimensional random vector $Y = (Y_1, \ldots, Y_d)$ by

$$Y_j := X_j^\nu = \max_{i=1,\ldots,r} \{a_{ij}^\nu W_i\}, \qquad j \in \{1, \ldots, d\}.$$

It is easily seen that, as $x \to \infty$,

$$1 - F_{Y_j}(x) = 1 - \exp\left\{ -\frac{\sum_{i=1}^r a_{ij}^\nu}{x} \right\} \sim \frac{\sum_{i=1}^r a_{ij}^\nu}{x}.$$

Since the X_j are increasing transformations of the Y_j, the (tail) dependence structure of X and Y is the same. We will determine the tail dependence function l and the spectral measure H of X.

Lemma 6.1. *Let X follow a factor model given by* (6.1) *or* (6.2). *Then its stable tail dependence function is given by*

$$l(x_1, \ldots, x_d) = \sum_{i=1}^r \max_{j=1,\ldots,d} \{b_{ij} x_j\}, \quad (x_1, \ldots, x_d) \in [0, \infty)^d,$$

where $b_{ij} := a_{ij}^\nu / \sum_{i=1}^r a_{ij}^\nu$.

Next, we are looking for a measure H on the unit simplex $\Delta_{d-1} = \{w \in [0,\infty)^d : w_1 + \cdots + w_d = 1\}$ such that for all $x \in [0,\infty)^d$,

$$\sum_{i=1}^{r} \max_{j=1,\ldots,d} \{b_{ij} x_j\} = l(x_1,\ldots,x_d) = \int_{\Delta_{d-1}} \max_{j=1,\ldots,d} \{w_j x_j\} \, H(\mathrm{d}w).$$

This H is a discrete measure with r atoms given by

$$\left(\frac{b_{i1}}{\sum_j b_{ij}}, \ldots, \frac{b_{id}}{\sum_j b_{ij}} \right), \qquad i \in \{1,\ldots,r\}, \tag{6.3}$$

the atom receiving mass $\sum_j b_{ij}$, which is positive by assumption. Note that H is indeed a spectral measure, for

$$\int_{\Delta_{d-1}} w_j \, H(\mathrm{d}w) = \sum_{i=1}^{r} b_{ij} = 1, \qquad j \in \{1,\ldots,d\}. \tag{6.4}$$

Every discrete spectral measure can arise in this way.

The spectral measure is completely determined by the $r \times d$ parameters b_{ij}, but by the d moment conditions from (6.4), the actual number of parameters is $p = (r-1)d$. The parameter vector $\theta \in \mathbb{R}^p$, which is to be estimated, can be constructed in many ways. For identification purposes, the definition of θ should be unambiguous. We opt for the following approach. Consider the matrix of the coefficients b_{ij},

$$\begin{pmatrix} b_{11} & \cdots & b_{r1} \\ \vdots & \ddots & \vdots \\ b_{1d} & \cdots & b_{rd} \end{pmatrix} \in \mathbb{R}^{d \times r}.$$

The coefficients corresponding to the i-th factor, $i = 1,\ldots,r$, are in the i-th column of this matrix. We define θ by stacking the above columns in decreasing order of their sums, leaving out the column with the lowest sum. (If two columns have the same sum, we order them then in decreasing order lexicographically.)

The definition of the M-estimator of θ involves integrals of the form

$$\int_{[0,1]^d} g_m(x) l(x) \, \mathrm{d}x = \sum_{i=1}^{r} \int_{[0,1]^d} g_m(x) \max_{j=1,\ldots,d} \{b_{ij} x_j\} \, \mathrm{d}x,$$

where $g_m : [0,1]^d \to \mathbb{R}$ is integrable and $m = 1,\ldots,q$. A possible choice is $g_m(x) = x_k^s$, where $k \in \{1,\ldots,d\}$ and $s \geqslant 0$.

Lemma 6.2. *If l is the tail dependence function of a factor model such that all $b_{ij} > 0$, then*

$$\int_{[0,1]^d} x_k^s l(x) \, dx = \sum_{i=1}^{r} \sum_{j=1}^{d} \frac{b_{ij}}{1 + s(1-\delta_{jk})} \int_0^1 \left(\frac{b_{ij}}{b_{ik}} x \wedge 1 \right)^s \prod_{l=1}^{d} \left(\frac{b_{ij}}{b_{il}} x \wedge 1 \right) dx,$$

where δ_{jk} is 1 if $j = k$ and 0 if $j \neq k$.

The integral on the right-hand side is to be computed numerically.

We illustrate the performance of the M-estimator on two factor models: a four-dimensional model with 2 factors ($p = 1 \times 4 = 4$), and a three-dimensional model with 3 factors ($p = 2 \times 3 = 6$).

Simulation study: Four-dimensional model with two factors. We simulated 500 samples of size $n = 5000$ from a four-dimensional model

$$\begin{aligned} X_1 &= 0.2Z_1 \vee 0.8Z_2 \\ X_2 &= 0.5Z_1 \vee 0.5Z_2 \\ X_3 &= 0.7Z_1 \vee 0.3Z_2 \\ X_4 &= 0.9Z_1 \vee 0.1Z_2, \end{aligned}$$

with independent standard Fréchet factors Z_1 and Z_2. We have $\theta = (0.2, 0.5, 0.7, 0.9)$.

In Figure 3 we show the bias and the RMSE of the M-estimator based on $q = 5$ moment equations, with auxiliary functions $g_i(x) = x_i$, for $i = 1, 2, 3, 4$ and $g_5 \equiv 1$. Estimation in this particular example benefited from the extension of the method of moments estimator to the M-estimator. Adding a fifth moment equation via $g_5 \equiv 1$ reduced the RMSE of the estimator in most cases and for most values of k. The M-estimator performs very well. For relatively small k, the four components of θ are estimated equally well, whereas for larger k the estimator performs somewhat better for parameter values in the "middle" of the interval $(0, 1)$ than for values near 0 or 1.

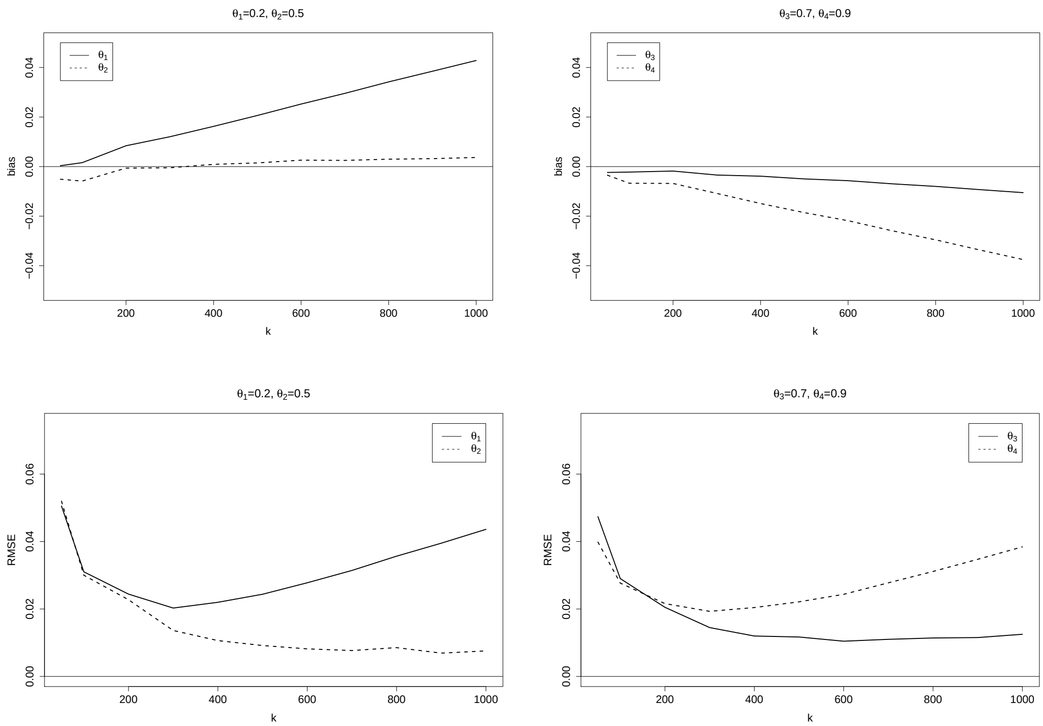


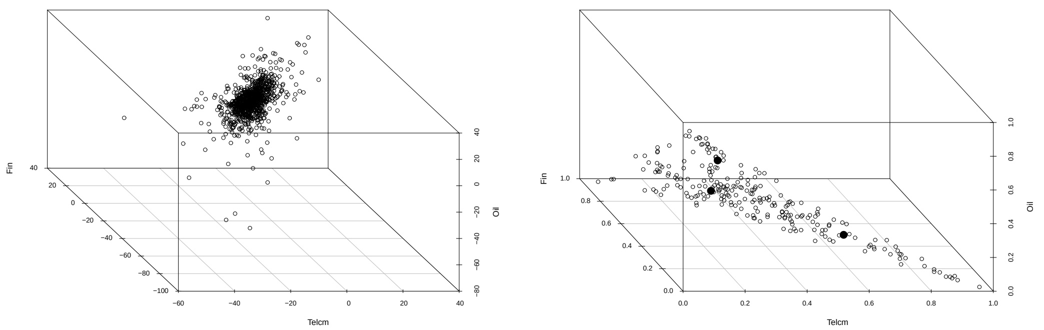
Fig. 3: Four-dimensional 2-factor model, estimation of $\theta = (0.2, 0.5, 0.7, 0.9)$.

Real data: Three-dimensional model with three factors. We consider monthly returns of three industry portfolios (Telecommunications, Finance and Oil) over the period July 1, 1963, until December 31, 2009. The data are available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french`. We are interested in modelling the losses (negative returns) by a factor model. See Figure 4(a) for the scatterplot of the losses; the sample size $n = 1002$.

Based on Fama and French (1993), in which three main stock-market factors are proposed, we consider a three-factor model for the three industry portfolios above. To estimate the parameter vector with $p = 2 \times 3 = 6$ components, we need to find a minimum of a 6-dimensional nonlinear criterion function. To solve such a difficult minimisation problem, it is important to have good starting values. We find a starting parameter vector by applying the 3-means clustering algorithm (see for example Pollard (1984), page 9) to the following pseudo-data: we transform the data (Telcm, Fin, Oil) to

$$(n/(n+1-R_{Ti}), n/(n+1-R_{Fi}), n/(n+1-R_{Oi}))\,, \quad i = 1, \ldots, n,$$

where R_{Ti}, R_{Fi} and R_{Oi} are the ranks of the components of the i-th observation. Only the entries such that the sum of their values is greater than the threshold $n/75$ are taken into account, and subsequently normalized such that they belong to the unit simplex Δ_{3-1}, see Figure 4(b). Then we compute the 3-means cluster centers for these data. Using equation (6.3), we compute from these three centers the 6-dimensional starting parameter [as described below equation (6.4)] for the minimisation routine.



(a) (b)

Fig. 4: (a) Scatterplot of the original data; (b) Plot of the pseudo-data and the three centers.

For the criterion function we use $q = 7$ functions g_i as follows: $g_i(x) = x_i$ for $i = 1, 2, 3$, $g_i(x) = x_{i-3}^2$ for $i = 4, 5, 6$, and $g_7 \equiv 1$. For different choices of k, we obtain the estimates presented in the table below. The ones in parentheses follow from the other ones by the moment conditions.

$k = 60$			$k = 90$		
0.394	0.593	(0.013)	0.344	0.616	(0.040)
0.691	0.211	(0.098)	0.701	0.216	(0.083)
0.358	0.062	(0.580)	0.368	0.052	(0.580)
$k = 120$			$k = 150$		
0.387	0.586	(0.027)	0.388	0.581	(0.031)
0.695	0.215	(0.090)	0.699	0.211	(0.090)
0.348	0.058	(0.594)	0.364	0.086	(0.550)

Estimates for the factor loadings b_{ij} in the three-factor model fitted to the tail of the Telcm/Fin/Oil data.

Observe that the estimates do hardly depend on the choice of k. We see that all three portfolios load substantially on the first factor, but Telecommunications loads more on the second factor and Oil more on the third factor. This shows that even for only these three portfolios three factors are required.

Remark 6.3. The examples we have presented show good performance and wide applicability of the estimator. Its performance, however, depends on the function g. The optimal choice of g is a difficult issue, which is beyond the scope of the present paper. The choices of g in Sections 5 and 6 is driven by computational feasibility, cf. Lemma 6.2.

7. Proofs

The asymptotic properties of the nonparametric estimator $\hat{l}_n$ are required for the proofs of the asymptotic properties of the M-estimator $\hat{\theta}_n$. Consistency of $\hat{l}_n$, see (7.1), for dimension $d = 2$ was shown in Huang (1992), cf. Drees and Huang (1998). In particular, it holds that for every $T > 0$, as $n \to \infty$, $k \to \infty$ and $k/n \to 0$,

$$\sup_{(x_1,x_2)\in[0,T]^2} |\hat{l}_n(x_1, x_2) - l(x_1, x_2)| \xrightarrow{\mathbb{P}} 0.$$

The proof translates straightforwardly to general dimension d, and together with integrability of g yields consistency of $\int g\hat{l}_n$ for $\int gl = \varphi(\theta_0)$. For the proof of Theorem 4.1, a technical result is needed.

Lemma 7.1. *If $k/n \to 0$ and if in addition to the assumptions of Theorem 4.1 condition (C4) holds, then as $n \to \infty$ and $k \to \infty$, on some closed neighbourhood of θ_0:*

(i) $\mathcal{H}_{k,n}(\theta) \xrightarrow{\mathbb{P}} \mathcal{H}(\theta)$ uniformly in θ, and
(ii) $\mathbb{P}\,(\mathcal{H}_{k,n}(\theta)$ is positive definite$) \to 1$.

PROOF. (i) The Hessian matrix of $Q_{k,n}$ in θ is a $p \times p$ matrix $\mathcal{H}_{k,n}(\theta)$ with elements $(\mathcal{H}_{k,n}(\theta))_{ij} = \partial^2 Q_{k,n}(\theta)/\partial\theta_j\partial\theta_i$ for $i,j \in \{1,\dots,p\}$ given by

$$\begin{aligned}(\mathcal{H}_{k,n}(\theta))_{ij} &= 2\sum_{m=1}^{q}\int_{[0,1]^d} g_m(x)\frac{\partial}{\partial\theta_j}l(x;\theta)\mathrm{d}x \cdot \int_{[0,1]^d} g_m(x)\frac{\partial}{\partial\theta_i}l(x;\theta)\mathrm{d}x \\ &\quad -2\sum_{m=1}^{q}\int_{[0,1]^d} g_m(x)\frac{\partial^2}{\partial\theta_j\partial\theta_i}l(x;\theta)\mathrm{d}x \cdot \int_{[0,1]^d} g_m(x)(\hat{l}_n(x) - l(x;\theta))\mathrm{d}x \\ &= 2\left(\frac{\partial}{\partial\theta_i}\varphi(\theta)\right)^T\left(\frac{\partial}{\partial\theta_j}\varphi(\theta)\right) - 2\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\varphi(\theta)\right)^T \\ &\quad \cdot\left(\int_{[0,1]^d} g(x)\hat{l}_n(x)\mathrm{d}x - \varphi(\theta)\right).\end{aligned}$$

The consistency of $\int g\hat{l}_n$ for $\varphi(\theta_0)$ implies

$$\begin{aligned}(\mathcal{H}_{k,n}(\theta))_{ij} &\xrightarrow{\mathbb{P}} 2\left(\frac{\partial}{\partial\theta_i}\varphi(\theta)\right)^T\left(\frac{\partial}{\partial\theta_j}\varphi(\theta)\right) - 2\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\varphi(\theta)\right)^T(\varphi(\theta_0) - \varphi(\theta)) \\ &=: (\mathcal{H}(\theta))_{ij}\,.\end{aligned}$$

Since we assumed that there exists $\varepsilon_0 > 0$ such that the set $\{\theta \in \Theta : \|\theta - \theta_0\| \leqslant \varepsilon_0\} =: B_{\varepsilon_0}(\theta_0)$ is closed, and since φ is assumed to be twice continuously differentiable, the second derivatives of φ are uniformly bounded on $B_{\varepsilon_0}(\theta_0)$, and hence, the convergence above is uniform on $B_{\varepsilon_0}(\theta_0)$.

(ii) For $\theta = \theta_0$ we get

$$(\mathcal{H}(\theta_0))_{ij} = 2\left(\frac{\partial}{\partial\theta_i}\varphi(\theta)\Big|_{\theta=\theta_0}\right)^T\left(\frac{\partial}{\partial\theta_j}\varphi(\theta)\Big|_{\theta=\theta_0}\right),$$

that is,

$$\mathcal{H}(\theta_0) = 2\dot{\varphi}(\theta_0)^T\dot{\varphi}(\theta_0).$$

Since $\dot{\varphi}(\theta_0)$ is assumed to be of full rank, $\mathcal{H}(\theta_0)$ is positive definite. For θ close to θ_0, $\mathcal{H}(\theta)$ is also positive definite. Due to the uniform convergence of $\mathcal{H}_{k,n}(\theta)$ to $\mathcal{H}(\theta)$ on $B_{\varepsilon_0}(\theta_0)$, the matrix $\mathcal{H}_{k,n}(\theta)$ is also positive definite on $B_{\varepsilon_0}(\theta_0)$ with probability tending to one. □

PROOF OF THEOREM 4.1. (i) Fix $\varepsilon > 0$ such that $0 < \varepsilon \leqslant \varepsilon_0$. Since φ is a homeomorphism, there exists $\delta > 0$ such that $\theta \in \Theta$ and $\|\varphi(\theta) - \varphi(\theta_0)\| \leqslant \delta$ implies $\|\theta - \theta_0\| \leqslant \varepsilon$. In other words, for every $\theta \in \Theta$ such that $\|\theta - \theta_0\| > \varepsilon$, we have $\|\varphi(\theta) - \varphi(\theta_0)\| > \delta$. Hence, on the event

$$A_n = \{\|\varphi(\theta_0) - \textstyle\int g\hat{l}_n\| \leqslant \delta/2\},$$

for every $\theta \in \Theta$ with $\|\theta - \theta_0\| > \varepsilon$, necessarily

$$\|\varphi(\theta) - \textstyle\int g\hat{l}_n\| \geqslant \|\varphi(\theta) - \varphi(\theta_0)\| - \|\varphi(\theta_0) - \int g\hat{l}_n\| > \delta - \delta/2 = \delta/2 \geqslant \|\varphi(\theta_0) - \int g\hat{l}_n\|.$$

As a consequence, on the event A_n, we have

$$\inf_{\theta:\|\theta-\theta_0\|>\varepsilon}\|\varphi(\theta) - \textstyle\int g\hat{l}_n\| > \min_{\theta:\|\theta-\theta_0\|\leqslant\varepsilon}\|\varphi(\theta) - \int g\hat{l}_n\|.$$

Hence, on the event A_n, the "argmin" set $\hat{\Theta}_n$ is non-empty and is contained in the closed ball of radius ε centered at θ_0. Finally, $\mathbb{P}(A_n) \to 1$ by weak consistency of $\int g\hat{l}_n$ for $\int gl = \varphi(\theta_0)$.

(ii) In the proof of (i) we have seen that with probability tending to one the proposed M-estimator exists and it is contained in a closed ball around θ_0. In Lemma 7.1 we have shown that the criterion function is with probability tending to one strictly convex on such a closed ball around θ_0, and hence, with probability tending to one, the minimiser of the criterion function is unique. $\square$

For $i = 1, \dots, n$ let

$$U_i := (U_{i1}, \dots, U_{id}) := (1 - F_1(X_{i1}), \dots, 1 - F_d(X_{id})),$$

and denote

$$\begin{aligned} Q_{nj}(u_j) &:= U_{\lceil nu_j \rceil : n, j},\ j = 1, \dots, d, \\ S_{nj}(x_j) &:= \frac{n}{k} Q_{nj}\left(\frac{kx_j}{n}\right),\ j = 1, \dots, d, \\ S_n(x) &:= (S_{n1}(x_1), \dots, S_{nd}(x_d)), \end{aligned}$$

where $U_{1:n,j} \leqslant \dots \leqslant U_{n:n,j}$ are the order statistics of $U_{1j}, \dots, U_{nj}$, $j = 1, \dots, d$, and $\lceil a \rceil$ is the smallest integer not smaller than a. Write

$$\begin{aligned} V_n(x) &:= \frac{n}{k}\mathbb{P}\left(U_{11} \leqslant \frac{kx_1}{n} \text{ or } \dots \text{ or } U_{1d} \leqslant \frac{kx_d}{n}\right), \\ T_n(x) &:= \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{U_{i1} < \frac{kx_1}{n} \text{ or } \dots \text{ or } U_{id} < \frac{kx_d}{n}\right\}, \\ \hat{L}_n(x) &:= \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{U_{i1} < \frac{k}{n}S_{n1}(x_1) \text{ or } \dots \text{ or } U_{id} < \frac{k}{n}S_{nd}(x_d)\right\}, \\ &= \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{R_i^1 > n + 1 - kx_1 \text{ or } \dots \text{ or } R_i^d > n + 1 - kx_d\right\}, \end{aligned}$$

and note that

$$\hat{L}_n(x) = T_n(S_n(x)).$$

Since

$$\sup_{x \in [0,1]^d} \sqrt{k}\left|\hat{l}_n(x) - \hat{L}_n(x)\right| \leqslant \frac{d}{\sqrt{k}} \to 0, \tag{7.1}$$

the asymptotic properties of $\hat{l}_n$ and $\hat{L}_n$ are the same. With the notation $v_n(x) = \sqrt{k}(T_n(x) - V_n(x))$, we have the following result.

Proposition 7.2. *Let $T > 0$ and denote $A_x := \{u \in [0, \infty]^d : u_1 \leqslant x_1 \text{ or} \cdots \text{or } u_d \leqslant x_d\}$. There exists a sequence of processes $\tilde{v}_n$ such that for all n $\tilde{v}_n \stackrel{d}{=} v_n$ and there exist a Wiener process $W_l(x) := W_\Lambda(A_x)$ such that as $n \to \infty$,*

$$\sup_{x \in [0,2T]^d} |\tilde{v}_n(x) - W_l(x)| \xrightarrow{\mathbb{P}} 0. \tag{7.2}$$

The result follows from Theorem 3.1 in Einmahl (1997). From the proofs there it follows that a single Wiener process, instead of the sequence in the original statement of the theorem, can be used, and that convergence holds almost surely, instead of in probability, once the Skorohod construction is introduced. From now on, we work on this new (Skorohod) probability space, but keep the old notation, without the tildes. In particular we have convergence of the marginal processes:

$$\sup_{x_j\in[0,2T]} |v_{nj}(x) - W_j(x_j)| \to 0 \text{ a.s.}, \; j = 1, \ldots, d,$$

where $v_{nj}(x_j) := v_n((0, \ldots, 0, x_j, 0, \ldots, 0))$. The Vervaat (1972) lemma implies

$$\sup_{x_j\in[0,2T]} |\sqrt{k}(S_{nj}(x_j) - x_j) + W_j(x_j)| \to 0 \text{ a.s.}, \; j = 1, \ldots, d. \tag{7.3}$$

Proof of Theorem 4.2. Write

$$\begin{aligned}
&\sqrt{k}\left(\hat{L}_n(x) - l(x)\right) \\
&\quad = \sqrt{k}\big(T_n(S_n(x)) - V_n(S_n(x))\big) + \sqrt{k}\big(V_n(S_n(x)) - l(S_n(x))\big) + \sqrt{k}\big(l(S_n(x)) - l(x)\big) \\
&\quad = D_1(x) + D_2(x) + D_3(x).
\end{aligned}$$

Proof of $\sup_{x\in[0,T]^d} |D_1(x) - W_l(x)| \xrightarrow{\mathbb{P}} 0$.
We have

$$D_1(x) = \sqrt{k}\,(T_n(S_n(x)) - V_n(S_n(x))) = v_n(S_n(x)).$$

It holds that

$$\begin{aligned}
&\sup_{x\in[0,T]^d} |D_1(x) - W_l(x)| \\
\leqslant\; &\sup_{x\in[0,T]^d} |D_1(x) - W_l(S_n(x))| + \sup_{x\in[0,T]^d} |W_l(S_n(x)) - W_l(x)|\,.
\end{aligned}$$

Because of (7.3), this is with probability tending to one less than or equal to

$$\sup_{y\in[0,2T]^d} |v_n(y) - W_l(y)| + \sup_{x\in[0,T]^d} |W_l(S_n(x)) - W_l(x)|\,.$$

Both terms tend to zero in probability, the first one by Proposition 7.2, the second one because of the uniform continuity of W_l and (7.3).

Proof of $\sup_{x\in[0,T]^d} |D_2(x)| \xrightarrow{\mathbb{P}} 0$.

Because of (7.3), with probability tending to one, $\sup_{x\in[0,T]^d} |D_2(x)|$ is less than or equal to $\sup_{y\in[0,2T]^d} \sqrt{k}|V_n(y) - l(y)|$, which in turn, because of conditions (C2) and (C3), is equal to

$$\sqrt{k}O\left(\left(\frac{k}{n}\right)^{\alpha}\right) = O\left(\left(\frac{k}{n^{2\alpha/(1+2\alpha)}}\right)^{\frac{1}{2}+\alpha}\right) = o(1).$$

Proof of $\sup_{x\in[0,T]^d} |D_3(x) + \sum_{j=1}^d l_j(x) W_j(x_j)| \xrightarrow{\mathbb{P}} 0$.

Due to the existence of the first derivatives, we can use the mean value theorem to write

$$\frac{1}{\sqrt{k}} D_3(x) \quad = \quad l\left(S_n(x)\right) - l(x) = \sum_{j=1}^{d} (S_{nj}(x_j) - x_j) \cdot l_j(\xi_n),$$

with ξ_n between x and $S_n(x)$. Therefore

$$\sup_{x\in[0,T]^d} |D_3(x) + \sum_{j=1}^{d} l_j(x) W_j(x_j)| \leqslant \sum_{j=1}^{d} |l_j(\xi_n)\sqrt{k}(S_{nj}(x_j) - x_j) + l_j(x) W_j(x_j)|.$$

Note that all the terms on the right-hand side of the above inequality can be dealt with in the same way. Therefore we consider only the first term. For $\delta \in (0, T)$, this term is bounded by

$$\begin{aligned} & \sup_{x\in[0,T]^d} |l_1(\xi_n)| \cdot \sup_{x_1\in[0,T]} |\sqrt{k}(S_{n1}(x_1) - x_1) + W_1(x_1)| \\ & \qquad + \sup_{x\in[\delta,T]\times[0,T]^{d-1}} |l_1(\xi_n) - l_1(x)| \cdot \sup_{x_1\in[0,T]} |W_1(x_1)| \\ & \qquad + \sup_{x\in[0,\delta]\times[0,T]^{d-1}} |l_1(\xi_n) - l_1(x)| \cdot \sup_{x_1\in[0,\delta]} |W_1(x_1)| \\ & =: \quad D_4 \cdot D_5 + D_6 \cdot D_7 + D_8 \cdot D_9. \end{aligned}$$

Observe that $0 \leqslant l_1 \leqslant 1$. Also, since l_1 is continuous on $[0, \delta/2] \times [0, T]^{d-1}$, it is uniformly continuous on that region. We have $D_5 \xrightarrow{\mathbb{P}} 0$ by (7.3), so $D_4 \cdot D_5 \xrightarrow{\mathbb{P}} 0$. The uniform continuity of l_1 and the fact that almost surely $D_7 < \infty$ yield $D_6 \cdot D_7 \xrightarrow{\mathbb{P}} 0$. Finally, for every $\varepsilon > 0$, we can find a δ such that, with probability at least $1 - \varepsilon$, $D_9 < \varepsilon$ and hence $D_8 \cdot D_9 < \varepsilon$.

Applying (7.1) completes the proof. □

PROPOSITION 7.3. *If the conditions (C1), (C2) hold, then as $n \to \infty$ and $k \to \infty$,*

$$\sqrt{k} \int_{[0,1]^d} g(x) \left(\hat{l}_n(x) - l(x)\right) dx \xrightarrow{d} \tilde{B}. \tag{7.4}$$

PROOF. Throughout the proof we write $l(x)$ instead of $l(x; \theta_0)$. Also, since l does not need to be differentiable, we will use notation $l_j(x)$, $j = 1, \dots, d$, to denote the right-hand partial derivatives here. Let $D_1(x)$, $D_2(x)$, $D_3(x)$ be as in the proof of Theorem 4.2 and take $T = 1$. Then

$$\begin{aligned} & \left|\sqrt{k}\Big(\int_{[0,1]^d} g(x)\hat{L}_n(x)\mathrm{d}x - \int_{[0,1]^d} g(x) l(x)\mathrm{d}x\Big) - \tilde{B}\right| \\ & \qquad \leqslant \sup_{x\in[0,1]^d} |D_1(x) - W_l(x)| \int_{[0,1]^d} |g(x)|\mathrm{d}x + \sup_{x\in[0,1]^d} |D_2(x)| \int_{[0,1]^d} |g(x)|\mathrm{d}x \\ & \qquad\qquad + \int_{[0,1]^d} |g(x,y)| \cdot \left| D_3(x) + \sum_{j=1}^{d} l_j(x) W_j(x_j) \right| \mathrm{d}x. \end{aligned}$$

The first two terms on the right hand side converge to zero in probability due to integrability of g and uniform convergence of $D_1(x)$ and $D_2(x)$, which was shown in the proof of Theorem 4.2. The third term needs to be treated separately, as the condition on continuity (and existence) of partial derivatives is no longer assumed to hold.

Let ω be a point in the Skorohod probability space introduced before the proof of Theorem 4.2 such that for all $j = 1, \ldots, d$,

$$\sup_{x_j \in [0,1]} |W_j(x_j)| < +\infty \quad \text{and} \quad \sup_{x_j \in [0,1]} |\sqrt{k}(S_{nj}(x_j) - x_j) + W_j(x_j)| \to 0.$$

For such ω we will show by means of dominated convergence that

$$\int_{[0,1]^d} |g(x)| \cdot \left| \sqrt{k}\,(l(S_n(x)) - l(x)) + \sum_{j=1}^{d} l_j(x) W_j(x_j) \right| \mathrm{d}x \to 0. \tag{7.5}$$

Proof of the pointwise convergence. If l is differentiable, convergence of the above integrand to zero follows from the definition of partial derivatives and (7.3). Since this might fail only on a set of Lebesgue measure zero, the convergence of the integrand to zero holds almost everywhere on $[0,1]^d$.

Proof of the domination. Note that from expressions for (one-sided) partial derivatives (2.4), and the moment conditions (2.2) it follows that $0 \leqslant l_j(x) \leqslant 1$, for all $x \in [0,1]^d$ and all $j = 1, \ldots, d$.

We get

$$\begin{aligned} &|g(x)| \cdot \left| \sqrt{k}\,(l\,(S_n(x)) - l(x)) + \sum_{j=1}^{d} l_j(x) W_j(x_j) \right| \\ &\qquad \leqslant |g(x)| \cdot \left(\sqrt{k} |l(S_n(x)) - l(x)| + \sum_{j=1}^{d} |W_j(x_j)| \right). \end{aligned}$$

Using the definition of function l and uniformity of $1 - F_j(X_{1j})$, we have for all $j = 1, \ldots, d$

$$|l(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_d) - l(x_1, \ldots, x_{j-1}, x'_j, x_{j+1}, \ldots, x_d)| \leqslant |x_j - x'_j|.$$

Hence, we can write

$$\begin{aligned} \sup_{x \in [0,1]^d} \sqrt{k} |l(S_n(x)) - l(x)| \;\leqslant\; & \sup_{x \in [0,1]^d} \sqrt{k} |l(S_n(x)) - l(x_1, S_{n2}(x_2), \ldots, S_{nd}(x_d))| \\ & + \sup_{x \in [0,1]^d} \sqrt{k} |l(x_1, S_{n2}(x_2), S_{n3}(x_3), \ldots, S_{nd}(x_d)) \\ & \quad - l(x_1, x_2, S_{n3}(x_3), \ldots, S_{nd}(x_d))| \\ & + \cdots \\ & + \sup_{x \in [0,1]^d} \sqrt{k} |l(x_1, \ldots, x_{d-1}, S_{nd}(x_d)) - l(x)| \\ \leqslant\; & \sum_{j=1}^{d} \sup_{x_j \in [0,1]} \sqrt{k} |S_{nj}(x_j) - x_j| = O(1). \end{aligned}$$

Since for all $j = 1, \ldots, d$ we have $\sup_{x_j \in [0,1]} |W_j(x_j)| < +\infty$, the proof of (7.5) is complete. This together with (7.1) finishes the proof of the proposition. $\square$

LEMMA 7.4. *If in addition to assumptions of Theorem 4.1, conditions (C1), (C2), (C4) hold, then as $n \to \infty$ and $k \to \infty$,*

$$\sqrt{k}\nabla Q_{k,n}(\theta_0) \xrightarrow{d} N(0, V(\theta_0)).$$

PROOF. The gradient vector of $Q_{k,n}$ with respect to θ in θ_0 is

$$\nabla Q_{k,n}(\theta_0) = \left(\frac{\partial}{\partial \theta_1} Q_{k,n}(\theta)\Big|_{\theta=\theta_0}, \ldots, \frac{\partial}{\partial \theta_p} Q_{k,n}(\theta)\Big|_{\theta=\theta_0} \right)^T,$$

where for $i = 1, \ldots, p$,

$$\frac{\partial}{\partial \theta_i} Q_{k,n}(\theta)\Big|_{\theta=\theta_0} = -2 \sum_{m=1}^{q} \int_{[0,1]^d} g_m(x) \frac{\partial}{\partial \theta_i} l(x;\theta)\Big|_{\theta=\theta_0} \mathrm{d}x \cdot \int_{[0,1]^d} g_m(x)(\hat{l}_n(x) - l(x;\theta_0))\mathrm{d}x.$$

Using vector notation we obtain

$$\nabla Q_{k,n}(\theta_0) = -2\dot{\varphi}(\theta_0)^T \cdot \int_{[0,1]^d} g(x)(\hat{l}_n(x) - l(x;\theta_0))\mathrm{d}x.$$

Equation (7.1) and the proof of Proposition 7.3 imply that

$$\sqrt{k}\nabla Q_{k,n}(\theta_0) = -2\dot{\varphi}(\theta_0)^T \cdot \int_{[0,1]^d} g(x)\sqrt{k}\left(\hat{l}_n(x) - l(x;\theta_0)\right) \mathrm{d}x \xrightarrow{d} -2\dot{\varphi}(\theta_0)^T \tilde{B}.$$

The limit distribution of $\sqrt{k}\nabla Q_{k,n}(\theta_0)$ is therefore zero-mean Gaussian with covariance matrix $V(\theta_0) = 4\dot{\varphi}(\theta_0)^T \Sigma(\theta_0)\dot{\varphi}(\theta_0)$. □

PROOF OF THEOREM 4.3. Consider the function $f(t) := \nabla Q_{k,n}(\theta_0 + t(\hat{\theta}_n - \theta_0))$, $t \in [0,1]$. The mean value theorem yields

$$\nabla Q_{k,n}(\hat{\theta}_n) = \nabla Q_{k,n}(\theta_0) + \mathcal{H}_{k,n}(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0),$$

for some $\tilde{\theta}_n$ between θ_0 and $\hat{\theta}_n$. First note that with probability tending to one, $0 = \nabla Q_{k,n}(\hat{\theta}_n)$, which follows from the fact that $\hat{\theta}_n$ is a minimiser of $Q_{k,n}$ and that with probability tending to one $\hat{\theta}_n$ is in an open ball around θ_0. By the consistency of $\hat{\theta}_n$ we have that $\tilde{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$, and since the convergence of $\mathcal{H}_{k,n}$ to $\mathcal{H}$ is uniform on a neighbourhood of θ_0, we get that $\mathcal{H}_{k,n}(\tilde{\theta}_n) \xrightarrow{\mathbb{P}} \mathcal{H}(\theta_0)$. Hence, $\sqrt{k}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, M(\theta_0))$. □

PROOF OF COROLLARY 4.4. As in Lemma 7.2 in Einmahl et al. (2008), we can see that if $\theta \mapsto H_\theta$ is weakly continuous at θ_0, then $\theta \mapsto \Sigma(\theta)$ is continuous at θ_0. This, together with condition (C4), yields that $\theta \mapsto V(\theta)$ is continuous at θ_0. Assumption (C4) also implies that $\theta \mapsto \mathcal{H}(\theta)$ is continuous at θ_0, which, with the positive definiteness of $\mathcal{H}(\theta)$ in a neighbourhood of θ_0, shows that if $\theta \mapsto H_\theta$ is weakly continuous at θ_0, then $\theta \mapsto M(\theta) = \mathcal{H}(\theta)^{-1}V(\theta)\mathcal{H}(\theta)^{-1}$ is continuous at θ_0. Hence, we obtain

$$M(\hat{\theta}_n)^{-1/2}\sqrt{k}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I_p),$$

which yields (4.4). □

Proof of Theorem 4.5. Theorem 4.3 and the arguments used in the proof of Corollary 4.4 imply that as $n \to \infty$,

$$M_2^{-1/2}(\hat{\theta}_1, \theta_2^*)\sqrt{k}(\hat{\theta}_2 - \theta_2^*) \xrightarrow{d} N(0, I_r), \tag{7.6}$$

and hence (4.11). □

Proof of Lemma 6.1. We have

$$\begin{aligned}
l(x_1, \dots, x_d) &= \lim_{t\to\infty} t\mathbb{P}\left(1 - F_1(X_1) \leqslant x_1/t \text{ or } \dots \text{ or } 1 - F_d(X_d) \leqslant x_d/t\right) \\
&= \lim_{t\to\infty} t\mathbb{P}\left(1 - F_{Y_1}(Y_1) \leqslant x_1/t \text{ or } \dots \text{ or } 1 - F_{Y_d}(Y_d) \leqslant x_d/t\right) \\
&= \lim_{t\to\infty} t\mathbb{P}\left(Y_1 \geqslant \frac{t\sum_{i=1}^r a_{i1}^\nu}{x_1} \text{ or } \dots \text{ or } Y_d \geqslant \frac{t\sum_{i=1}^r a_{id}^\nu}{x_d}\right) \\
&= \lim_{t\to\infty} t\mathbb{P}\left(\bigcup_{1\leqslant j\leqslant d}\bigcup_{1\leqslant i\leqslant r}\left\{W_i \geqslant \frac{t\sum_{i=1}^r a_{ij}^\nu}{a_{ij}^\nu x_j}\right\}\right) \\
&= \lim_{t\to\infty} t\mathbb{P}\left(\bigcup_{1\leqslant i\leqslant r}\left\{W_i \geqslant \min_{1\leqslant j\leqslant d}\frac{t\sum_{i=1}^r a_{ij}^\nu}{a_{ij}^\nu x_j}\right\}\right) \\
&= \lim_{t\to\infty} t\sum_{i=1}^r \mathbb{P}\left(W_i \geqslant \min_{1\leqslant j\leqslant d}\frac{t\sum_{i=1}^r a_{ij}^\nu}{a_{ij}^\nu x_j}\right) \\
&= \lim_{t\to\infty} \sum_{i=1}^r t\left(1 - \exp\left\{-\frac{1}{t}\max_{1\leqslant j\leqslant d}\frac{a_{ij}^\nu x_j}{\sum_{i=1}^r a_{ij}^\nu}\right\}\right) \\
&= \sum_{i=1}^r \max_{1\leqslant j\leqslant d}\left\{\frac{a_{ij}^\nu x_j}{\sum_{i=1}^r a_{ij}^\nu}\right\} =: \sum_{i=1}^r \max_{1\leqslant j\leqslant d}\{b_{ij}x_j\}
\end{aligned}$$

as required. □

Proof of Lemma 6.2. Fix $i \in \{1, \dots, r\}$. We have

$$\int_{[0,1]^d} x_k^s \max_{1\leqslant j\leqslant d}\{b_{ij}x_j\}\mathrm{d}x = \sum_{j=1}^d \int_{[0,1]^d} x_k^s(b_{ij}x_j)\mathbf{1}\left(b_{ij}x_j \geqslant \max_{l\neq j}\{b_{il}x_l\}\right)\mathrm{d}x.$$

Write the integral as a double integral, the outer integral with respect to $x_j \in [0,1]$ and the inner integral with respect to $x_{-j} = (x_l)_{l\neq j} \in \mathbb{R}^{d-1}$ over the relevant domain. We find

$$\int_{[0,1]^d} x_k^s \max_{1\leqslant j\leqslant d}\{b_{ij}x_j\}\,\mathrm{d}x = \sum_{j=1}^d \int_0^1 b_{ij}x_j \int_{0<x_l<\frac{b_{ij}}{b_{il}}x_j\wedge 1} x_k^s\,\mathrm{d}x_{-j}\,\mathrm{d}x_j.$$

After some long, but elementary computations, this simplifies to the stated expression. □

Acknowledgements

We are very grateful to Axel Bücher for pointing out that our original condition (C4) of Theorem 4.2 was too restrictive. Johan Segers gratefully acknowledges financial support from IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy) and from the contract "Projet d'Actions de Recherche Concertées" no. 07/12/002 of the Communauté française de Belgique, granted by the Académie universitaire Louvain. Andrea Krajina gratefully acknowledges financial support from the Open Competition grant from the Netherlands Organisation for Scientific Research (NWO) and from the Deutsches Forschungsgemeinschaft (DFG) grant SNF FOR 916; her research was mainly performed at Tilburg University and Eurandom (Eindhoven).

References

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Chichester: Wiley.

Boldi, M.-O. and A. Davison (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society: Series B 69*, 217–229.

Coles, S. and J. Tawn (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society, Series B 53*, 377–392.

Cooley, D., R. Davis, and N. P. (2010). The pairwise beta distribution: A flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis 101*, 2103–2117.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction*. New York: Springer.

de Haan, L., C. Neves, and L. Peng (2008). Parametric tail copula estimation and model testing. *Journal of Multivariate Analysis 99*, 1260–1275.

de Haan, L. and S. Resnick (1977). Limit theory for multidimensional sample extremes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete 40*, 317–337.

Drees, H. and X. Huang (1998). Best attainable rates of convergence for estimates of the stable tail dependence function. *Journal of Multivariate Analysis 64*, 25–47.

Einmahl, J. H. J. (1997). Poisson and Gaussian approximation of weighted local empirical processes. *Stoch. Proc. Appl. 70*, 31–58.

Einmahl, J. H. J., A. Krajina, and J. Segers (2008). A method of moments estimator of tail dependence. *Bernoulli 14*, 1003–1026.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 3–56.

Frees, E. and E. Valdez (1998). Understanding relationships using copulas. *North American Actuarial Journal 2*, 1–25.

Geluk, J. L., L. de Haan, and C. G. de Vries (2007). Weak & strong financial fragility. Technical report, Tinbergen Institute. TI 2007-023/2.

Guillotte, S., F. Perron, and J. Segers (2011). Nonparametric Bayesian inference on bivariate extremes. *Journal of the Royal Statistical Society. Series B*. (to appear) arXiv:0911.3270 [math.ST].

Gumbel, E. (1960). Bivariate exponential distributions. *Journal of the American Statistical Society 55*, 698–707.

Huang, X. (1992). *Statistics of bivariate extreme values*. Ph. D. thesis, Tinbergen Institute Research Series.

Joe, H., R. L. Smith, and I. Weissman (1992). Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society, Series B 54*, 171–183.

Ledford, A. W. and J. A. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika 83*, 169–187.

Malevergne, Y. and D. Sornette (2004). Tail dependence of factor models. *The Journal of Risk 6*, 71–116.

Pollard, D. (1984). *Convergence of stochastic processes*. New York: Springer-Verlag.

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. New York: Springer.

Smith, R. L. (1994). Multivariate threshold methods. In J. Galambos, J. Lechner, and E. Simiu (Eds.), *Extreme Value Theory and Applications*, pp. 225–248. Kluwer.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika 75*, 397–415.

Vervaat, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrscheinlichkeitstheorie verw. Gebiete 23*, 249–253.