INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION
PAPER

1105

AN M-ESTIMATOR FOR TAIL DEPENDENCE IN ARBITRARY DIMENSIONS

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

Submitted to the Annals of Statistics
arXiv: arXiv:1112.0905

AN M-ESTIMATOR FOR TAIL DEPENDENCE IN ARBITRARY DIMENSIONS

By John H.J. Einmahl[‡], Andrea Krajina[*,§] and Johan Segers[†,¶]

Tilburg University[‡], University of Göttingen[§] and Université catholique de Louvain[¶]

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.

1. Introduction. Statistics of multivariate extremes finds important applications in fields like finance, insurance, environmental sciences, aviation safety, hydrology, and meteorology. When considering multivariate extreme events, the estimation of the tail dependence structure is the key part of the statistical inference. This tail dependence structure, represented by the stable tail dependence function l, is becoming rather complex if the dimension increases. Therefore it is customary to model this multivariate function l parametrically, which leads to a semiparametric model. The interest in parametric tail dependence models exists since the early sixties of the 20th century (Gumbel, 1960), but new models are still being proposed (Boldi and Davison, 2007; Cooley, Davis and Naveau, 2010; Ballani and Schlather, 2011). Most of the existing estimators of the parameter, θ, are likelihood-based and their asymptotic behavior is only known in dimension two (Coles and Tawn, 1991; Joe, Smith and Weissman, 1992; Smith, 1994; Ledford and Tawn, 1996; de Haan, Neves and Peng, 2008; Guillotte, Perron and Segers, 2011). For many applications, the bivariate set-up is too restrictive. Also, the likelihood-based estimation methods exclude models that entail a non-differentiable function l, like the widely used factor models, see (1.1) below.

It is the goal of this paper to present and provide a comprehensive treatment of novel M-estimators of θ. The estimators can be used in arbitrary dimension d. Moreover, not relying on the differentiability of l, the estimators are broadly applicable. We establish, again for arbitrary dimension d, the asymptotic normality of our estimators, which yields asymptotic

[*]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).

[†]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.

AMS 2000 subject classifications: Primary 62G32, 62G05, 62G10, 62G20, 60K35; secondary 60F05, 60F17, 60G70

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

confidence regions and tests for the parameter θ. The results in this paper make statistical inference possible for many multivariate extreme value models that either cannot be handled at all by currently available methods or for which statistical theory has only been provided for the bivariate case. Monte Carlo simulation studies confirm that our estimators perform well in practice, see Sections 5 and 6.

The present estimators are a major extension of the method of moments estimators for dimension two (Einmahl, Krajina and Segers, 2008). For applications, the crucial difference is that it is now possible to handle truly multivariate data. Also theoretically, extreme value analysis in dimensions larger than two is quite challenging, which explains why in many papers attention is restricted to the bivariate case. In particular, we establish the asymptotic behavior of the nonparametric estimator of l in arbitrary dimensions and under nonrestrictive smoothness conditions; compare for instance with Drees and Huang (1998) in the bivariate case. Another novel aspect is that the method of moments technique is replaced by general M-estimation, that is, allowing for more estimating equations than the dimension of the parameter space. This more flexible procedure may serve to increase the efficiency of the estimator.

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

$$X_j = \sum_{i=1}^{r} a_{ij} Z_i + \varepsilon_j, \quad j = 1, \ldots, d, \tag{1.1}$$

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 (Fama and French, 1993; Malevergne and Sornette, 2004; Geluk, de Haan and de Vries, 2007). From equation (6.3) below, we see that the stable tail dependence function l of such a factor model is not everywhere differentiable, causing likelihood-based methods to break down.

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 (5.1). 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 $F_1, \ldots, F_d$. For $j = 1, \ldots, d$, write $M_n^{(j)} := \max_{i=1,\ldots,n} X_{ij}$. We say that F is in the max-domain of attraction of an extreme value distribution G if there exist sequences $a_n^{(j)} > 0$, $b_n^{(j)} \in \mathbb{R}$, $j = 1, \ldots, d$, such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n^{(1)} - b_n^{(1)}}{a_n^{(1)}} \leqslant x_1, \ldots, \frac{M_n^{(d)} - b_n^{(d)}}{a_n^{(d)}} \leqslant x_d \right) = G(x), \tag{2.1}$$

for all continuity points $x \in \mathbb{R}^d$ of G. The margins $G_1, \ldots, G_d$ of G must be univariate extreme value distributions and the dependence structure of G is determined by the relation

$$-\log G(x) = l\big(-\log G_1(x_1), \ldots, -\log G_d(x_d)\big)$$

for all points x such that $G_j(x_j) > 0$ for all $j = 1, \ldots, d$. The stable tail dependence function $l : [0,\infty)^d \to [0,\infty)$ can be retrieved from F via

$$l(x) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}\left\{1 - F_1(X_{11}) \leqslant t x_1 \text{ or } \ldots \text{ or } 1 - F_d(X_{1d}) \leqslant t x_d\right\}. \tag{2.2}$$

In fact, the joint convergence in (2.1) is equivalent to convergence of the d marginal distributions together with (2.2).

In this paper, we will only assume the weaker relation (2.2). By itself, (2.2) holds if and only if the random vector $(1/\{1 - F_1(X_{1j})\})_{j=1}^d$ belongs to the max-domain of attraction of the extreme value distribution $G_0(x) = \exp\{-l(1/x_1, \ldots, 1/x_d)\}$ for $x \in (0,\infty)^d$. Alternatively, the existence of the limit in (2.2) is equivalent to multivariate regular variation of the random vector $(1/\{1 - F_1(X_{1j})\})_{j=1}^d$ on the cone $[0,\infty]^d \setminus \{(0, \ldots, 0)\}$ with limit measure or exponent measure μ given by

$$\mu\left(\left\{z \in [0,\infty]^d \colon z_1 \geqslant x_1 \text{ or } \ldots \text{ or } z_d \geqslant x_d\right\}\right) = l(1/x_1, \ldots, 1/x_d)$$

(Resnick, 1987; Beirlant et al., 2004; de Haan and Ferreira, 2006). The measure μ is homogeneous, that is, $\mu(tA) = t^{-1}\mu(A)$, for any $t > 0$ and any relatively compact Borel set $A \subset [0,\infty]^d \setminus \{(0, \ldots, 0)\}$, where $tA := \{tz : z \in A\}$. This homogeneity property yields a decomposition of μ into a radial and an angular part (de Haan and Resnick, 1977; Resnick, 1987). Let $\Delta_{d-1} := \{w \in [0,1]^d : w_1 + \cdots + w_d = 1\}$ be the unit simplex in $\mathbb{R}^d$. Associated to $B \subset \Delta_{d-1}$ and $t > 0$ is the set

$$B_t = \left\{x \in [0,\infty)^d \setminus \{(0, \ldots, 0)\} \colon \textstyle\sum_{j=1}^d x_j \geqslant t,\ x/\sum_{j=1}^d x_j \in B\right\}.$$

By the homogeneity property of the exponent measure, it holds that $\mu(B_t) = t^{-1}\mu(B_1)$. Writing $H(B) = \mu(B_1)$ defines a finite measure H on Δ_{d-1}, called the spectral or angular measure. Any finite measure satisfying the moment conditions

$$\int_{\Delta_{d-1}} w_j H(\mathrm{d}w) = 1, \qquad j = 1, \ldots, d, \tag{2.3}$$

is a spectral measure. Adding up the d constraints in (2.3) shows that H/d is a probability measure.

Sometimes it is more convenient to work with the measure Λ obtained from μ after the transformation $(x_1, \ldots, x_d) \mapsto (1/x_1, \ldots, 1/x_d)$. The measure Λ is also called the exponent measure and it satisfies the homogeneity property $\Lambda(tA) = t\Lambda(A)$, for any $t > 0$ and Borel set $A \subset [0,\infty]^d \setminus \{(\infty, \ldots, \infty)\}$.

There is a one-to-one correspondence between the stable tail dependence function l, the exponent measures μ and Λ, and the spectral measure H. In particular, we have

$$l(x) = \mu\left(\left\{(z_1, \ldots, z_d) \in [0,\infty]^d \colon z_1 \geqslant 1/x_1 \text{ or } \ldots \text{ or } z_d \geqslant 1/x_d\right\}\right) \tag{2.4}$$

$$= \Lambda\left(\left\{(u_1, \ldots, u_d) \in [0,\infty]^d \colon u_1 \leqslant x_1 \text{ or } \ldots \text{ or } u_d \leqslant x_d\right\}\right) \tag{2.5}$$

$$= \int_{\Delta_{d-1}} \max_{j=1,\ldots,d} \{w_j x_j\} H(\mathrm{d}w). \tag{2.6}$$

From the above representations and the moment constraints (2.3), it follows that the function l has the following properties:

- $\max\{x_1,\dots,x_d\} \leqslant l(x) \leqslant x_1+\cdots+x_d$ for all $x \in [0,\infty)^d$; in particular $l(z,0,\dots,0) = \cdots = l(0,\dots,0,z) = z$ 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 $x \in (0,\infty)^d$.

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

$$\begin{aligned}(2.7)\quad &\frac{1}{h}\big(l(x_1,\dots,x_{j-1},x_j+h,x_{j+1},\dots,x_d) - l(x_1,\dots,x_{j-1},x_j,x_{j+1},\dots,x_d)\big)\\ &= \int_{\Delta_{d-1}} \frac{1}{h}\Big(\max\{w_jx_j+w_jh, \max_{s\neq j}\{w_sx_s\}\} - \max\{w_jx_j, \max_{s\neq j}\{w_sx_s\}\}\Big) H(\mathrm{d}w)\\ &\to \int_{\Delta_{d-1}} w_j \mathbf{1}\{w_jx_j \geqslant \max_{s\neq j}\{w_sx_s\}\} H(\mathrm{d}w).\end{aligned}$$

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 (2.7).

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

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

see Huang (1992) and Drees and Huang (1998) for the bivariate case. This definition follows from (2.2), with all the distribution functions replaced by their empirical counterparts, and with t replaced by k/n. Here $k = k_n$ is such that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. The constant $1/2$ in the argument of the indicator function helps to improve the finite-sample properties of the estimator.

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 the copula is constrained only through l, see (2.2).

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

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

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

$$(3.3)\qquad Q_{k,n}(\theta) = \|\varphi(\theta) - \int g\hat{l}_n\|^2 = \sum_{m=1}^{q}\left(\int_{[0,1]^d} g_m(x)\left(\hat{l}_n(x) - l(x;\theta)\right)\mathrm{d}x\right)^2,$$

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.

The fact that our model assumption only concerns a limit relation in the tail shows up in the estimation procedure through the choice of k, which determines the effective sample size. When we study asymptotic properties of either $\hat{l}_n$ or $\hat{\theta}_n$, $k = k_n$ is an intermediate sequence, that is, $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. In practice, the choice of optimal k is a notorious problem, and here we address this issue in the usual way: we present the finite sample results over a wide range of k, see Sections 5 and 6.

REMARK 3.1. The estimator $\hat{\theta}_n$ depends on g. In line with the classical method of moments and for computational feasibility we will choose g to be a vector of low degree polynomials. In Sections 5 and 6 we will see that the thus obtained estimators have a good performance and a wide applicability. Finding an optimal g is very difficult and statistically not very useful since such a g depends on the true, unknown θ_0. For example, when $p = q = 1$, a function g that minimises the asymptotic variance is $(\partial/\partial\theta)l(x;\theta_0)$. For two-dimensional and five-dimensional data, a sensitivity analysis on the choice of g is performed in Section 5. Simple functions like 1 or x_1 lead to estimators that perform approximately the same as the pseudo-estimator based on the optimal g. This supports our choices of g and also suggests that the estimator is not so sensitive to the choice of g.

REMARK 3.2. Since l, part of the model, is parametrically specified, in principle pseudo maximum likelihood estimation could be used. This method however does not apply to many interesting models where l is not differentiable, like the factor model in (1.1). Moreover, no theory is known for dimensions higher than 2, unless the limit relation (2.2) is replaced by an equality for all sufficiently small t. In this paper, the emphasis is on higher dimensions and for a large part on the factor model. Therefore the pseudo MLE is not an available competitor.

4. Asymptotic results. Let $\hat{\Theta}_n$ be the set of minimisers of $Q_{k,n}$ in (3.3), that is,

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

Note that $\hat{\Theta}_n$ may be empty or may contain more than one element. We show that under suitable conditions, a 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.* Recall the definition of the measure Λ from Section 2. 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 $W_{l,j}$, $j = 1,\dots,d$, be the marginal processes

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

Define l_j to be the right-hand partial derivative of l with respect to x_j, where $j = 1, \ldots, d$, see (2.7); if l is differentiable, l_j is equal to the corresponding partial derivative of l. Write

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

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

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

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

Assuming θ is an interior point of Θ and φ is differentiable in θ, let $\dot{\varphi}(\theta) \in \mathbb{R}^{q\times p}$ be the total derivative of φ at θ, and, provided $\dot{\varphi}(\theta)$ is of full rank, put

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

4.2. *Results.* We state the asymptotic results for the M-estimator, $\hat{\theta}_n$, and the asymptotic normality of $\hat{l}_n$. The latter is a result of independent interest, and requires continuous partial derivatives of l, which is not an assumption for the asymptotic normality of the M-estimator. The proofs can be found in Section 7.

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 $n \to \infty$,

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

(ii) If in addition to the assumptions of (i), θ_0 is in the interior of the parameter space, φ is twice continuously differentiable and $\dot{\varphi}(\theta_0)$ is of full rank, then, with probability tending to one, $Q_{k,n}$ in (3.3) 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 $\{\theta \in \Theta : \|\theta - \theta_0\| \leqslant \varepsilon\}$ 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.

THEOREM 4.2 (Asymptotic normality of $\hat{\theta}_n$). *If in addition to the assumptions of Theorem 4.1(ii), the following two conditions hold:*

(C1) $t^{-1}\mathbb{P}\{1 - F_1(X_{11}) \leqslant t x_1 \text{ or } \ldots \text{ or } 1 - F_d(X_{1d}) \leqslant t x_d\} - l(x) = O(t^\alpha)$, uniformly in $x \in \Delta_{d-1}$ as $t \downarrow 0$, for some $\alpha > 0$,

(C2) $k = o\big(n^{2\alpha/(1+2\alpha)}\big)$, for the positive number α of (C1), and $k \to \infty$ as $n \to \infty$;

then as $n \to \infty$, with M as in (4.6),

$$\sqrt{k}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\big(0, M(\theta_0)\big). \tag{4.7}$$

The following consequence of Theorem 4.2 can be used for the construction of confidence regions. Recall from (2.6) that H_θ is the spectral measure corresponding to $l(\cdot;\theta)$. Let χ^2_ν denote the χ^2-distribution with ν degrees of freedom.

COROLLARY 4.3. *If in addition to the conditions of Theorem 4.2, the map $\theta \mapsto H_\theta$ is weakly continuous at θ_0 and if the matrix $M(\theta_0)$ is non-singular, then as $n \to \infty$,*

$$k(\hat{\theta}_n - \theta_0)^T M(\hat{\theta}_n)^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \chi^2_p. \tag{4.8}$$

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,

$$M = \left(\begin{array}{c|c} \cdots & \cdots \\ \hline \cdots & M_2 \end{array}\right) \in \mathbb{R}^{p\times p}. \tag{4.9}$$

COROLLARY 4.4 (Test). *If the assumptions of Corollary 4.3 are satisfied, and $\theta_0 = (\theta_1, \theta_2^*) \in \Theta$ for some θ_1, then as $n \to \infty$,*

$$k(\hat{\theta}_{2n} - \theta_2^*)^T M_2(\hat{\theta}_{1n}, \theta_2^*)^{-1}(\hat{\theta}_{2n} - \theta_2^*) \xrightarrow{d} \chi^2_r. \tag{4.10}$$

The above result can be used for testing for a submodel. For example, we could test for the symmetric logistic model of (5.3) within the asymmetric logistic one, see Section 5.

REMARK 4.5. 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

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

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.

Finally we show 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, a large part of its proof is necessary for the proof of the asymptotic normality of $\hat{\theta}_n$, but we wish to emphasize that the asymptotic normality of $\hat{\theta}_n$ holds *without* any differentiability conditions

on l. Note that under assumption (C3) below, 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 4.6 (Asymptotic normality of $\hat{l}_n$ in arbitrary dimensions). *If in addition to the conditions (C1) and (C2) from Theorem 4.2, the following condition holds:*

(C3) 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 $x_j > 0$;

then for every $T > 0$, as $n \to \infty$,

$$\sup_{x \in [0,T]^d} \left| \sqrt{k} \left(\hat{l}_n(x) - l(x) \right) - B(x) \right| \xrightarrow{\mathbb{P}} 0. \tag{4.11}$$

5. Example 1: Logistic model. The multivariate logistic distribution function with standard Fréchet margins is defined by

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

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

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

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

Sensitivity analysis. Here we observe how for the logistic model the M-estimator changes with different choices of k, and for different functions g. Within this model, $p = 1$ and in the simple case of $p = q = 1$, it is easy to see that the optimal choice for the function g is $(\partial/\partial\theta) l(x; \theta_0)$. Since it depends on the unknown true parameter, this is not a viable option for use in practice, but, as demonstrated below, some simple alternatives result in estimators with basically the same finite-sample behavior.

The following analysis is performed for the logistic model with $\theta_0 = 0.5$, in dimensions 2 and 5. For both settings, we look at 200 replications of samples of size $n = 1500$, and take the threshold parameters $k \in \{40, 80, \ldots, 320\}$. In the bivariate case we compare $g_0(x_1, x_2) = 1$, $g_1(x_1, x_2) = x_1$ and $g_{\text{opt}}(x_1, x_2) = (\partial/\partial\theta) l(x_1, x_2; \theta_0)$ as choices for g. In the five-dimensional case the functions g_0 and g_{opt} are defined analogously, and we compare them to other two functions, $g_1(x) = \sum_{j=1}^5 x_j$ and $g_2(x) = \sum_{j=1}^5 x_j^2$. We use the bias and the Root Mean Squared Error (RMSE) to assess the performance of the estimators. The results are presented in Figure 1 for dimensions $d = 2$ (top) and $d = 5$ (bottom). All of the above choices for g result in similar finite-sample behavior of the estimator, but the simpler function g leads to a somewhat better performance. The RMSEs for some of these g are even lower than the one for g_{opt}, since they yield a smaller bias.

Based on these findings, for the logistic model in dimensions 2 and 5, we advise the use of the simplest choice of g given by $g_0(x) = 1$, for all $x \geqslant 0$. The choice of k is slightly more delicate, but it seems that for $n = 1500$ in dimensions 2 and 5, the choices of $k = 150$ and $k = 100$, respectively, are reasonable.

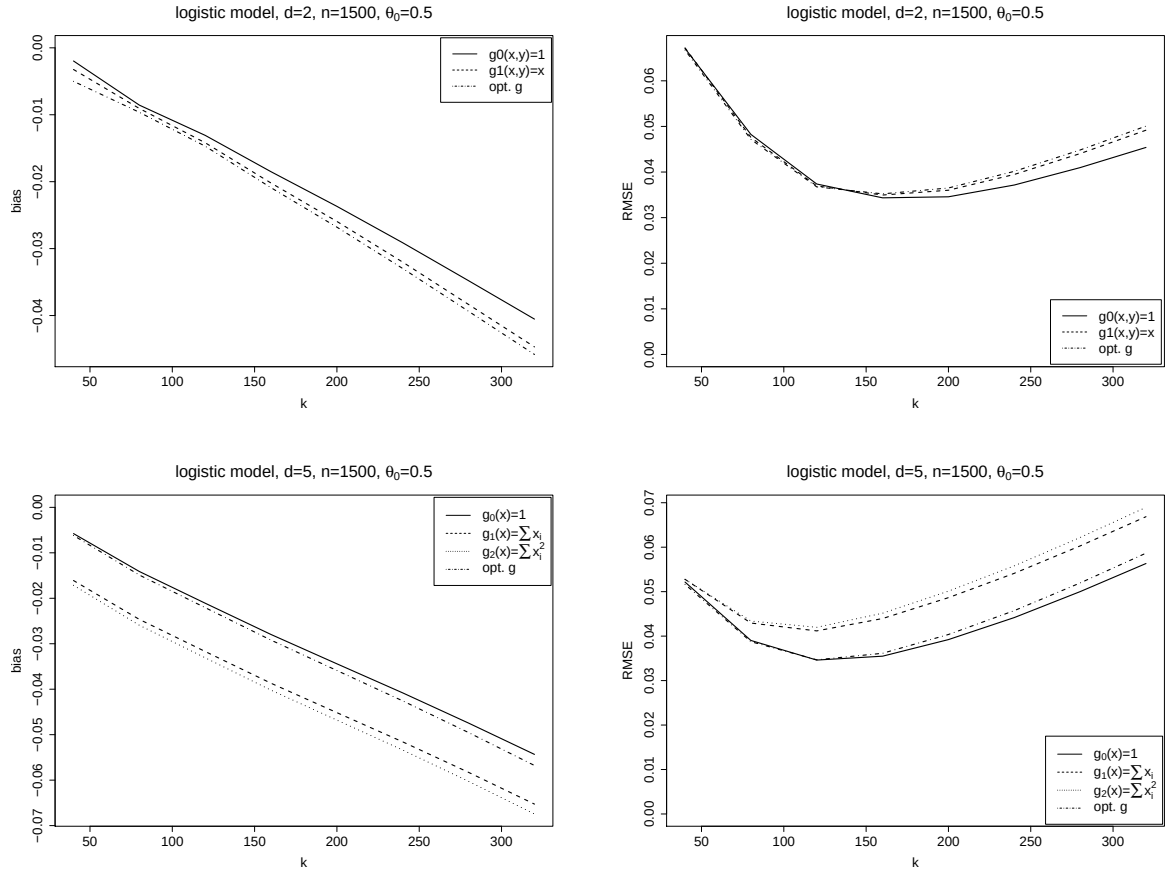


Fig 1: Logistic model: the M-estimator for different functions g in dimension $d = 2$ (top) and $d = 5$ (bottom).

Comparison with maximum likelihood based estimators. For $d = 2$, we also compare the M-estimator with $g \equiv 1$ with the censored maximum likelihood method, see Ledford and Tawn (1996) and with the maximum likelihood estimator introduced in de Haan, Neves and Peng (2008). The latter two we will call the censored MLE and the dHNP MLE respectively. For 200 samples, we compute the censored MLE using the function `fitbvgpd` from the R package `POT`, see Ribatet (2011); the dHNP MLE is calculated as described in the original article. Since the thresholds used in these two methods differ, and since for a different choice of threshold we get a different estimator, the comparison is not straightforward. We consider the M-estimator and the dHNP MLE over the range of k values as used above, and for the censored MLE we take the thresholds such that the expected number of joint exceedances is between 10 and 160, approximately, which amounts to thresholds between 5 and 100. This way we observe all estimators for their best region of thresholds. In Figure 2 we see that the methods perform roughly the same, the RMSEs being of the same order. The lowest RMSE of the censored MLE (0.030) is slightly smaller than the lowest RMSE of the M-estimator (0.034) and the lowest RMSE of the dHNP estimator (0.035), but the M- and the dHNP estimators are much more robust to the choice of the threshold.

Further simulation results. We simulate 500 samples of size $n = 1500$ from a five-dimensional logistic distribution function with $\theta_0 = 0.5$. As suggested by the sensitivity analysis, we opt for $g \equiv 1$ when defining $\hat{\theta}_n$. The bias and the RMSE of this estimator are shown in the upper panels of Figure 3.

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). The lower panels in Figure 3 show 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, for more information, see Frees and Valdez (1998). The scatterplots of the data and their joint ranks are shown in Figure 4. We consider the asymmetric logistic model described below for their tail dependence function and we test whether a more restrictive, 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

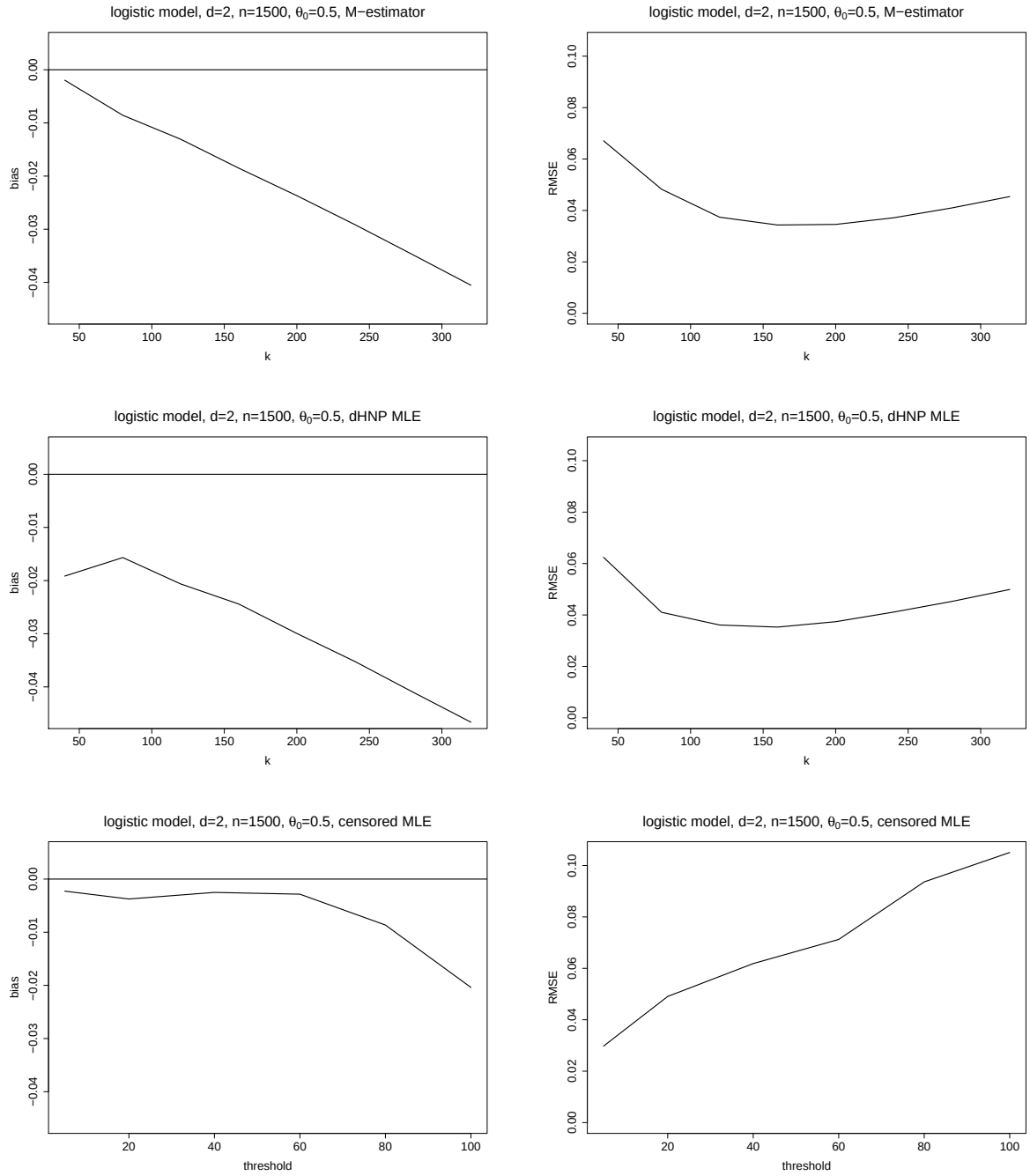
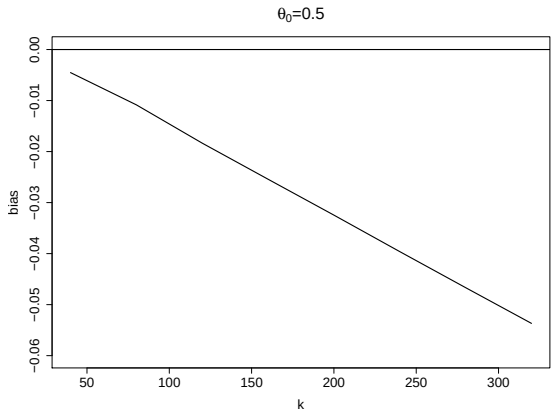


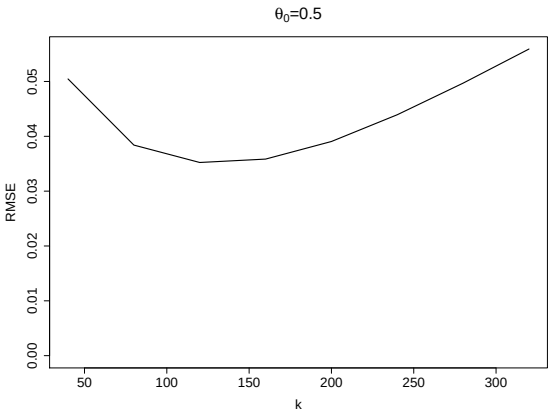
Fig 2: The M-estimator with $g(x, y) = g_0(x, y) = 1$, the MLE from de Haan, Neves and Peng (2008) and the censored MLE, $d = 2$.

to testing whether $\eta_2 = 0$. By Corollary 4.4, the test statistic is given by

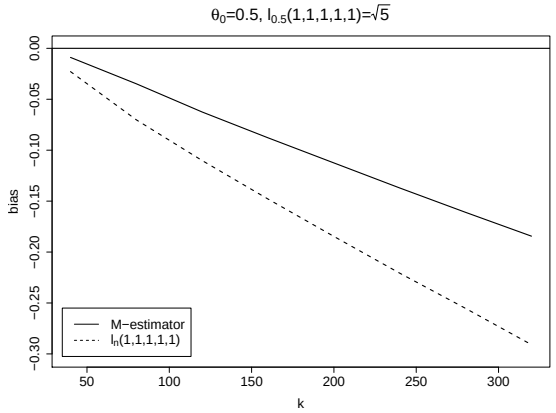
$$S_n := \frac{k\,\hat{\eta}_2^2}{M_2(\hat{\theta}, \hat{\eta}_1, 0)}.$$



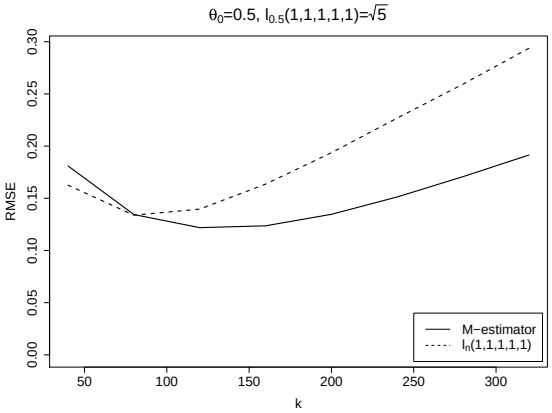
(a) Bias of the M-estimator of θ



(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 3: 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

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

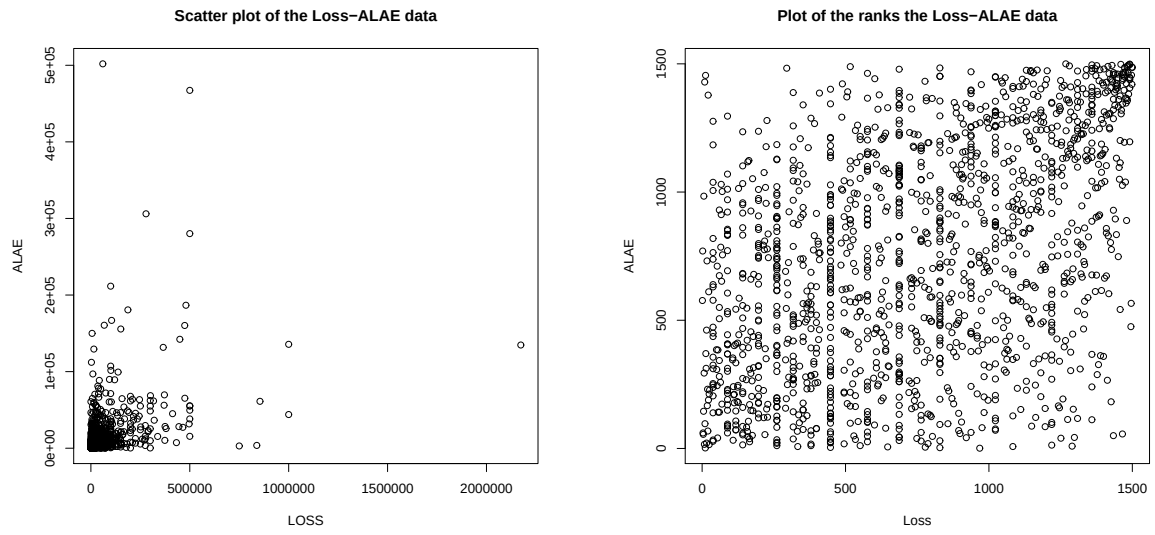


Fig 4: The insurance claims Loss-ALAE data.

with Z_i independent Fréchet(ν) random variables, $\nu > 0$, with ε_j independent random variables which have a lighter right tail than the factors and are independent of them, and with a_{ij} nonnegative 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, de Haan and de Vries (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; this essentially follows from the fact that the ratio of the probabilities of the sum and the maximum of the $a_{ij}Z_i$ exceeding x tends to 1 as $x \to \infty$ (Embrechts, Klüppelberg and Mikosch, 1997, page 38). Let $W_i = Z_i^\nu$, $i = 1, \ldots, r$, and observe 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 variables are increasing transformations of the Y_j variables, the (tail) dependence structures of X and Y coincide. 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*

$$l(x_1,\dots,x_d)=\sum_{i=1}^{r}\max_{j=1,\dots,d}\{b_{ij}x_j\},\qquad (x_1,\dots,x_d)\in[0,\infty)^d, \tag{6.3}$$

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 $x\in[0,\infty)^d$,

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

This H is a discrete measure with r atoms given by

$$\left(\frac{b_{i1}}{\sum_j b_{ij}},\dots,\frac{b_{id}}{\sum_j b_{ij}}\right),\qquad i\in\{1,\dots,r\}, \tag{6.4}$$

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

$$\int_{\Delta_{d-1}} w_j\,H(\mathrm{d}w)=\sum_{i=1}^{r}b_{ij}=1,\qquad j\in\{1,\dots,d\}. \tag{6.5}$$

Every discrete spectral measure can arise in this way. This model for tail dependence is considered also in Ledford and Tawn (1998). Extensions to random fields are considered for instance in Wang and Stoev (2011).

The spectral measure is completely determined by the $r\times d$ parameters b_{ij}, but by the d moment conditions from (6.5), 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, $i=1,\dots,r$, 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

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

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

$$\int_{[0,1]^d} x_k^s \, l(x) \, dx = \sum_{i=1}^{r} \sum_{j=1}^{d} \frac{b_{ij}}{1 + s(1 - \delta_{jk})} \int_0^1 \left(\frac{b_{ij}}{b_{ik}} x \wedge 1 \right)^s \prod_{l=1}^{d} \left(\frac{b_{ij}}{b_{il}} x \wedge 1 \right) dx, \tag{6.6}$$

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

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

The integral on the right-hand side of (6.6) is to be computed numerically. For the factor model, the dependence of the matrix $M(\theta_0)$ on g is too complicated to obtain a general solution for the optimal function g. Since in the previous examples low degree polynomials gave very good results, and since by the previous lemma such a choice simplifies the calculations significantly (numerical integration in dimension 1, instead of in dimension d), we considered such functions g in a sensitivity analysis. It showed that the simplest cases give very good results in terms of root mean squared errors and that the performance of the M-estimator is quite robust to the particular choices of g. Hence, we suggest using simple, low degree polynomials for the functions g. The functions g in the following examples are exactly of that type.

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 5 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 $g_5 \equiv 1$. 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.

Real data: Three-dimensional model with three factors. We consider monthly negative returns (losses) of three industry portfolios (Telecommunications, Finance and Oil) over the period July 1, 1926, until December 31, 2009. See Figure 6(a) for the scatterplot of the data; the sample size $n = 1002$. The data are available at

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french.

We are interested in modelling the losses by a factor model. In the asset pricing literature, see for example Fama and French (1993, 1996), it is common to model the returns by linear factor models of type (6.1), with three underlying economic factors. Based on that line of literature, we also consider a three-factor model for the tails of the three industry portfolios above, see also Kleibergen (2011).

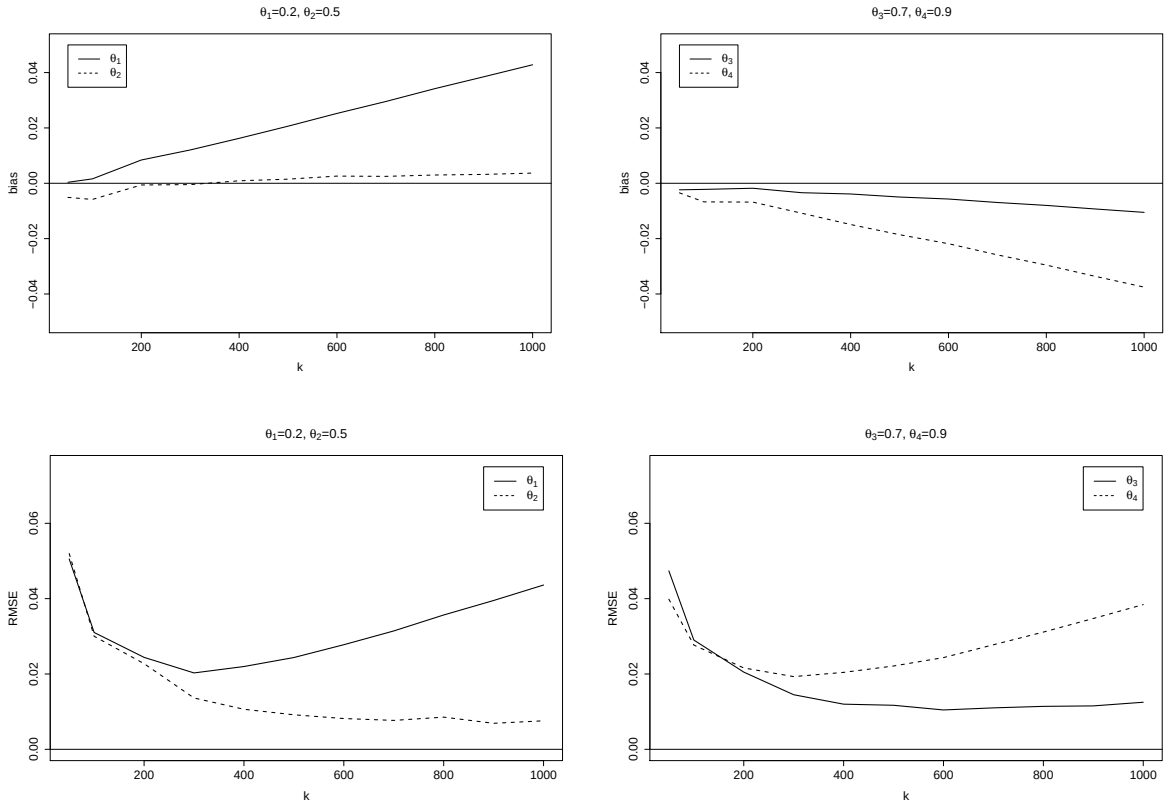
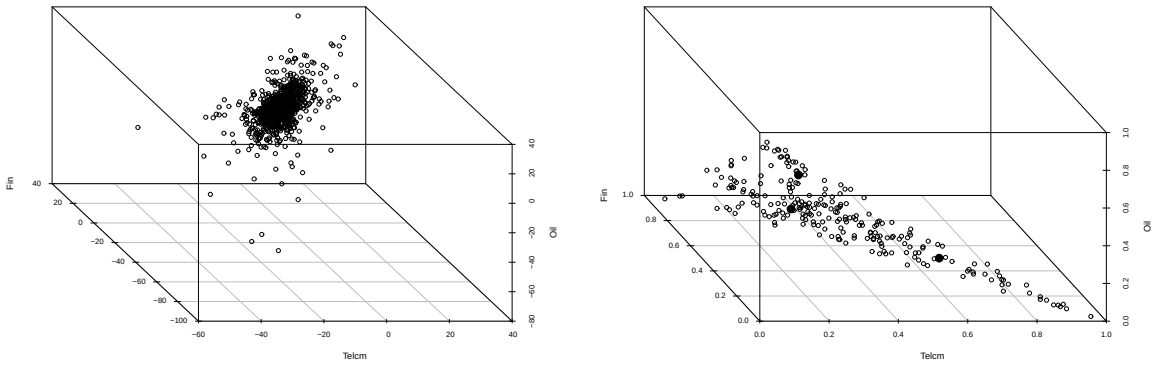


Fig 5: Four-dimensional 2-factor model, estimation of $\theta = (0.2, 0.5, 0.7, 0.9)$.

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 so that they belong to the unit simplex Δ_{3-1}, see Figure 6(b). We compute the 3-means cluster centers for these data. Using equation (6.4), we compute from these three centers the 6-dimensional starting parameter (as described below equation (6.5)) for the minimisation routine. 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. For each k, we estimate the loading of the first two factors. This corresponds to the first two columns of estimated b_{ij} for each k. The third columns follow from the conditions in (6.5).



(a) (b)

Fig 6: (a) Scatterplot of the original data; (b) Plot of the pseudo-data and the three centers.

$k = 60$			$k = 90$		
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 (the first column of estimated coefficients, for each k), but Telecommunications loads more on the second factor (the first lines of estimated coefficients), and Oil more on the third factor (the third lines of estimated coefficients). This indicates that even for only these three portfolios, three factors are required.

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.

Let $\mathcal{H}_{k,n}(\theta) \in \mathbb{R}^{p\times p}$ denote the Hessian matrix of $Q_{k,n}$ as a function of 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

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

LEMMA 7.1. *If $k/n \to 0$ and if the assumptions of Theorem 4.1(ii) are satisfied, 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) $\mathbb{P}\left(\mathcal{H}_{k,n}(\theta)\right.$ is positive definite$\left.\right) \to 1$.

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 thus compact, 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

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

where we can write the minimum on the right-hand side since the set $\{\theta \in \Theta\colon \|\theta - \theta_0\| \leqslant \varepsilon\}$, is closed and thus compact for $0 \leqslant \varepsilon \leqslant \varepsilon_0$. 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. □

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

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

and denote

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

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

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

and note that

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

With probability one, for every x and for every $j \in \{1, \ldots, d\}$, there is at most one i such that $n + \frac{1}{2} - kx_j < R_i^j \leqslant n + 1 - kx_j$. Hence

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

This shows that 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 $n \to \infty$,*

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

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:

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

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

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

PROOF OF THEOREM 4.6. Write

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

Proof of $\sup_{x \in [0,T]^d} |D_1(x) - W_l(x)| \stackrel{\mathbb{P}}{\to} 0$.
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 (C1) and (C2), 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 $\sup_{x\in[0,T]^d} |D_3(x) + \sum_{j=1}^d l_j(x)W_{l,j}(x_j)| \xrightarrow{\mathbb{P}} 0$.

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

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

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

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

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

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

Observe that $0 \leqslant l_1 \leqslant 1$. Also, since l_1 is continuous on $[\delta/2, T] \times [0,T]^{d-1}$, 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) and (C2) from Theorem 4.2 hold, then as* $n \to \infty$,

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

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 4.6 and take $T=1$. Then

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

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 4.6. 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 4.6 such that for all $j=1,\dots,d$,

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

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

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

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 (2.7), and the moment conditions (2.3), it follows that $0\leqslant l_j(x)\leqslant 1$, for all $x\in[0,1]^d$ and all $j=1,\dots,d$.

We get

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

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

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

Hence, we can write

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

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

Let $\nabla Q_{k,n}(\theta)\in\mathbb{R}^{p\times 1}$ be the gradient vector of $Q_{k,n}$ at θ. Put

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

LEMMA 7.4. *If the assumptions of Theorem 4.2 are satisfied, then as $n\to\infty$,*

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

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

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

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

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

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 4.2. 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 4.3. As in Lemma 7.2 in Einmahl, Krajina and Segers (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 the assumption that φ is twice continuously differentiable and $\dot{\varphi}(\theta_0)$ is of full rank, yields that $\theta \mapsto V(\theta)$ is continuous at θ_0. The above assumption 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 (4.3). □

PROOF OF THEOREM 4.4. Theorem 4.2 and the arguments used in the proof of Corollary 4.3 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 (4.10). □

PROOF OF LEMMA 6.1. We have

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

$$= \lim_{t\to\infty} \sum_{i=1}^{r} t \left(1 - \exp\left\{ -\frac{1}{t} \max_{1\leqslant j\leqslant d} \frac{a_{ij}^{\nu} x_j}{\sum_{i=1}^{r} a_{ij}^{\nu}} \right\} \right)$$

$$= \sum_{i=1}^{r} \max_{1\leqslant j\leqslant d} \left\{ \frac{a_{ij}^{\nu} x_j}{\sum_{i=1}^{r} a_{ij}^{\nu}} \right\} =: \sum_{i=1}^{r} \max_{1\leqslant j\leqslant d} \{b_{ij} x_j\}$$

as required. □

Proof of Lemma 6.2. Fix $i \in \{1, \ldots, r\}$. 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 grateful to Axel Bücher for pointing out that the original condition (C3) of Theorem 4.6 was too restrictive. We also like to thank the Associate Editor and two Referees for a thorough reading of the manuscript and for many thoughtful comments that led to this improved version.

References.

Ballani, F. and Schlather, M. (2011). A construction principle for multivariate extreme value distributions. *Biometrika* **98** 633-645.

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

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

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

Cooley, D., Davis, R. and Naveau, P. (2010). The Pairwise Beta Distribution: A Flexible Parametric Multivariate Model for Extremes. *Journal of Multivariate Analysis* **101** 2103-2117.

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

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

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

Drees, H. and Huang, X. (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., Krajina, A. and Segers, J. (2008). A Method of Moments Estimator of Tail Dependence. *Bernoulli* **14** 1003-1026.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, New York.

Fama, E. F. and French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* **33** 3–56.
Fama, E. F. and French, K. R. (1996). Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance* **51** 55–84.
Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal* **2** 1–25.
Geluk, J. L., de Haan, L. and de Vries, C. G. (2007). Weak & strong financial fragility. Technical Report, Tinbergen Institute. TI 2007-023/2.
Guillotte, S., Perron, F. and Segers, J. (2011). Nonparametric Bayesian inference on bivariate extremes. *Journal of the Royal Statistical Society. Series B* **73** 377–406.
Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Society* **55** 698-707.
Huang, X. (1992). Statistics of bivariate extreme values PhD thesis, Tinbergen Institute Research Series.
Joe, H., Smith, R. L. and Weissman, I. (1992). Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society, Series B* **54** 171–183.
Kleibergen, F. (2011). Reality checks for and of factor pricing Technical Report, Department of Economics, Brown University. Preprint available on http://www.econ.brown.edu/fac/Frank_Kleibergen/.
Ledford, A. W. and Tawn, J. A. (1996). Statistics for Near Independence in Multivariate Extreme Values. *Biometrika* **83** 169–187.
Ledford, A. W. and Tawn, J. A. (1998). Concomitant tail behaviour for extremes. *Advances in Applied Probability* **30** 197–215.
Malevergne, Y. and Sornette, D. (2004). Tail Dependence of Factor Models. *The Journal of Risk* **6** 71–116.
Pollard, D. (1984). *Convergence of stochastic processes*. Springer-Verlag, New York.
Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.
Ribatet, M. (2011). POT: Generalized Pareto Distribution and Peaks Over Threshold R package version 1.1-1.
Smith, R. L. (1994). Multivariate threshold methods. In *Extreme Value Theory and Applications* (J. Galambos, J. Lechner and E. Simiu, eds.) 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.
Wang, Y. and Stoev, S. A. (2011). Conditional Sampling for Max-Stable Random Fields. *Advances in Applied Probability* **43** 463-481.

John Einmahl
Department of Econometrics & OR and CentER
Tilburg University
PO Box 90153
5000 LE Tilburg, The Netherlands
E-mail: j.h.j.einmahl@uvt.nl

Andrea Krajina
Institute for Mathematical Stochastics
University of Göttingen
Göttingen, Germany
E-mail: Andrea.Krajina@mathematik.uni-goettingen.de

Johan Segers
ISBA, Université catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve, Belgium
E-mail: johan.segers@uclouvain.be