Extremes (2018) 21:205–233
https://doi.org/10.1007/s10687-017-0303-7

A continuous updating weighted least squares estimator of tail dependence in high dimensions

John H. J. Einmahl[1] · **Anna Kiriliouk**[2] · **Johan Segers**[2]

Received: 27 October 2016 / Revised: 13 July 2017 / Accepted: 11 August 2017 /
Published online: 31 August 2017

Abstract Likelihood-based procedures are a common way to estimate tail dependence parameters. They are not applicable, however, in non-differentiable models such as those arising from recent max-linear structural equation models. Moreover, they can be hard to compute in higher dimensions. An adaptive weighted least-squares procedure matching nonparametric estimates of the stable tail dependence function with the corresponding values of a parametrically specified proposal yields a novel minimum-distance estimator. The estimator is easy to calculate and applies to a wide range of sampling schemes and tail dependence models. In large samples, it is asymptotically normal with an explicit and estimable covariance matrix. The minimum distance obtained forms the basis of a goodness-of-fit statistic whose asymptotic distribution is chi-square. Extensive Monte Carlo simulations confirm the excellent finite-sample performance of the estimator and demonstrate that it is a strong competitor to currently available methods. The estimator is then applied to disentangle sources of tail dependence in European stock markets.

✉ Anna Kiriliouk
anna.kiriliouk@uclouvain.be

John H. J. Einmahl
j.h.j.einmahl@uvt.nl

Johan Segers
johan.segers@uclouvain.be

[1] Department of Econometrics & OR and CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

[2] Institut de Statistique, Biostatistique et Sciences Actuarielles, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium

Keywords Brown–Resnick process · Extremal coefficient · Max-linear model · Multivariate extremes · Stable tail dependence function

AMS 2000 Subject Classifications Primary–62G32 · 62G05 · 62G10 · 62G20; Secondary–60F05 · 60G70

1 Introduction

Extreme value analysis has been applied to measure and manage financial and actuarial risks, assess natural hazards stemming from heavy rainfall, wind storms, and earthquakes, and control processes in the food industry, internet traffic, aviation, and other branches of human activity. The extension from univariate to multivariate data gives rise to the concept of tail dependence. The latter can and will be represented here by the stable tail dependence function, denoted by ℓ (Huang 1992; Drees and Huang 1998), or tail dependence function for short. Estimating this tail dependence function is the subject of this paper. Fitting tail dependence models for spatial phenomena observed at finitely many sites constitutes an interesting special case.

In high(er) dimensions, the class of tail dependence functions becomes rather unwieldy, and therefore we follow the common route of modelling it parametrically. Note that this is far from imposing a fully parametric model on the data generating process. In particular, we only assume a domain-of-attraction condition at the copula level. Parametric models for tail dependence have their origins in Gumbel (1960), and many models have since then been proposed, see, e.g., Coles and Tawn (1991), and more recently, Kabluchko et al. (2009).

Likelihood-based procedures are perhaps the most common way to estimate tail dependence parameters (Davison et al. 2012; Wadsworth and Tawn 2014; Huser et al. 2016). Likelihood methods, however, are not applicable to models involving non-differentiable tail dependence functions. Such functions arise in max-linear models (Wang and Stoev 2011), in particular factor models (Einmahl et al. 2012) or structural equation models based on directed acyclic graphs (Gissibl and Klüppelberg 2017). Moreover, likelihoods can be hard to compute, especially in higher dimensions. This is why current likelihood methods are usually based on composite likelihoods, relying on pairs or triples of variables only, not exploiting information from higher-dimensional tuples.

It is the goal of this paper to estimate the true parameter vector θ_0 of the tail dependence function ℓ and to assess the goodness-of-fit of the parametric model. The parameter estimator is obtained by comparing, at finitely many points in the domain of ℓ, some initial, typically nonparametric, estimator of the latter with the corresponding values of the parametrically specified proposals, and retaining the parameter value yielding the best match. The method is generic in the sense that it applies to many parametric models, differentiable or not, and to many initial estimators, not only the usual empirical tail dependence function but also, for instance, bias-corrected versions thereof (Fougères et al. 2015; Beirlant et al. 2016). Further, the method avoids integration or differentiation of functions of many variables and can therefore handle joint dependence between many variables simultaneously, more easily than the likelihood methods mentioned earlier and the M-estimator approach in Einmahl et al. (2016). This feature is particularly interesting for inferring on higher-order interactions, going beyond mere distance-based dependence models such as those frequently employed for spatial extremes. Finally, in those situations where likelihood methods are applicable, the new estimator is a strong competitor.

The distance between the initial estimator and the parametric candidates is measured through weighted least squares. The weight matrix may depend on the unknown parameter θ and is hence estimated simultaneously. The construction of the estimator bears some similarity with the continuous updating generalized method of moments (Hansen et al. 1996); the present estimator, however, is substantially different and does not use moments. Our flexible estimation procedure is related to that in Einmahl et al. (2016), but the continuous updating procedure is new in multivariate extreme value statistics.

We show that the weighted least squares estimator for the tail dependence function is consistent and asymptotically normal, provided that the initial estimator enjoys these properties too, as is the case for the empirical tail dependence function and its recently proposed bias-corrected variations. The asymptotic covariance matrix is a function of the unknown parameter and can thus be estimated by a plug-in technique. We also provide novel goodness-of-fit tests for the parametric tail dependence model based on a comparison between the nonparametric and the parametric estimators. Under the null hypothesis that the tail dependence model is correctly specified, the test statistics are asymptotically chi-square distributed.

The paper is organized as follows. In Section 2 we present the estimator, the goodness-of-fit statistic, and their asymptotic distributions. Section 3 reports on a Monte Carlo simulation study involving a variety of models, as well as a finite-sample comparison of our estimator with estimators based on composite likelihoods. An application to European stock market data is presented in Section 4, where we try to disentangle sources of tail dependence stemming from the country of origin (Germany versus France) and the economic sector (chemicals versus insurance), fitting a structural equation model. All proofs are deferred to Appendix A. In Appendix B we verify the main conditions on the models considered in Sections 3.1–3.3.

2 Inference on tail dependence parameters

2.1 Setup

Let $X_i = (X_{i1}, \ldots, X_{id})$, $i \in \{1, \ldots, n\}$, be random vectors in $\mathbb{R}^d$ with a common cumulative distribution function F and marginal cumulative distribution functions $F_1, \ldots, F_d$. The (stable) tail dependence function $\ell : [0, \infty)^d \to [0, \infty)$ is defined as

$$\ell(x) := \lim_{t \downarrow 0} t^{-1}\, \mathbb{P}[1 - F_1(X_{11}) \le t x_1 \text{ or} \ldots \text{ or } 1 - F_d(X_{1d}) \le t x_d], \qquad (2.1)$$

for $x \in [0, \infty)^d$, provided the limit exists, as we will assume throughout. Existence of the limit is a necessary, but not sufficient, condition for F to be in the max-domain of attraction of a d-variate Generalized Extreme Value distribution. Closely related to ℓ is the exponent measure function $V(z) = \ell(1/z_1, \ldots, 1/z_d)$, for $z \in (0, \infty]^d$. For more background on multivariate extreme value theory, see for instance Beirlant et al. (2004) or de Haan and Ferreira (2006).

The function ℓ is convex and homogeneous of order one, that is, $\ell(cx) = c\ell(x)$ for $c > 0$. Moreover, it satisfies

$$\max(x_1, \ldots, x_d) \le \ell(x) \le x_1 + \cdots + x_d, \qquad x \in [0, \infty)^d, \tag{2.2}$$

where the lower bound corresponds to perfect tail dependence and the upper bound to asymptotic independence. If $d = 2$, the above properties characterize the class of all d-variate tail dependence functions, but not if $d \ge 3$ (Molchanov 2008; Ressel 2013). For any dimension $d \ge 2$, the collection of d-variate tail dependence functions is infinite-dimensional. This poses challenges to inference on tail dependence, especially in higher dimensions.

The usual way of dealing with this problem consists of considering parametric models for ℓ, a number of which are presented in Section 3. Henceforth we assume that ℓ belongs to a parametric family $\{\ell(\cdot\,;\theta) : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^p$. Let θ_0 denote the true parameter vector, that is, let θ_0 denote the unique point in Θ such that $\ell(x) = \ell(x;\theta_0)$ for all $x \in [0, \infty)^d$. Our aim is to estimate the parameter θ_0 and to test the goodness-of-fit of the model.

Extremal coefficients are popular summary measures of tail dependence (de Haan 1984; Smith 1990; Schlather and Tawn 2003). For non-empty $J \subset \{1, \ldots, d\}$, let $e_J \in \mathbb{R}^d$ be defined by

$$(e_J)_j := \begin{cases} 1 & \text{if } j \in J, \\ 0 & \text{if } j \in \{1, \ldots, d\} \setminus J. \end{cases} \tag{2.3}$$

The extremal coefficients are defined by

$$\ell_J := \ell(e_J) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}[\max_{j \in J} F_j(X_{1j}) \ge 1 - t]. \tag{2.4}$$

By Eq. 2.2, it follows that $1 \le \ell_J \le |J|$. The extremal coefficients ℓ_J can be interpreted as assigning to each subset J the effective number of tail independent variables among $(X_{1j})_{j \in J}$.

Comparing initial and parametric estimators of the extremal coefficients is a special case of the inference method that we propose. In fact, Smith (1990) already proposes an estimator based on pairwise ($|J| = 2$) extremal coefficients; see also de Haan and Pereira (2006) and Oesting et al. (2015).

2.2 Continuous updating weighted least squares estimator

Let $\widehat{\ell}_{n,k}$ denote an initial estimator of ℓ based on $X_1, \ldots, X_n$; some possibilities will be described in Subsection 2.5. The estimators $\widehat{\ell}_{n,k}$ that we will consider depend on an intermediate sequence $k = k_n \in (0, n]$, that is,

$$k \to \infty \quad \text{and } k/n \to 0, \qquad \text{as } n \to \infty. \tag{2.5}$$

The sequence k will determine the tail fraction of the data that we will use for inference, see for instance Subsection 2.5.

Let $c_1, \ldots, c_q \in [0, \infty)^d$, with $c_m = (c_{m1}, \ldots, c_{md})$ for $m = 1, \ldots, q$, be q points in which we will evaluate ℓ and $\widehat{\ell}_{n,k}$. Consider the $q \times 1$ column vectors

$$\begin{aligned}
\widehat{L}_{n,k} &:= \left(\widehat{\ell}_{n,k}(c_m)\right)_{m=1}^{q}, \\
L(\theta) &:= \left(\ell(c_m; \theta)\right)_{m=1}^{q}, \qquad (2.6) \\
D_{n,k}(\theta) &:= \widehat{L}_{n,k} - L(\theta), \qquad (2.7)
\end{aligned}$$

where $\theta \in \Theta$. The points $c_1, \ldots, c_q$ need to be chosen in such a way that the map $L : \Theta \to \mathbb{R}^q$ is one-to-one, i.e., θ is identifiable from the values of $\ell(c_1; \theta), \ldots, \ell(c_q; \theta)$. In particular, we will assume that $q \geq p$, where p is the dimension of the parameter space Θ. Since $\ell(ce_{\{j\}}) = c$ for any tail dependence function ℓ, any $c \in [0, \infty)$ and any $j \in \{1, \ldots, d\}$, we will choose the points c_m in such a way that each point has at least two positive coordinates.

For $\theta \in \Theta$, let $\Omega(\theta)$ be a symmetric, positive definite $q \times q$ matrix with ordered eigenvalues $0 < \lambda_1(\theta) \leq \ldots \leq \lambda_q(\theta)$ and define

$$f_{n,k}(\theta) := \|D_{n,k}(\theta)\|^2_{\Omega(\theta)} := D^T_{n,k}(\theta)\, \Omega(\theta)\, D_{n,k}(\theta). \qquad (2.8)$$

Our continuous updating weighted least squares estimator for θ_0 is defined as

$$\widehat{\theta}_{n,k} := \operatorname*{arg\,min}_{\theta \in \Theta} f_{n,k}(\theta) = \operatorname*{arg\,min}_{\theta \in \Theta} \left\{ D_{n,k}(\theta)^T\, \Omega(\theta)\, D_{n,k}(\theta) \right\}. \qquad (2.9)$$

The set of minimizers could be empty or could have more than one element. The present notation, suggesting that there exists a unique minimizer, will be justified in Theorem 1. If all points c_m are chosen as e_{J_m} in Eq. 2.3 for some collection $J_1, \ldots, J_q$ of q different subsets of $\{1, \ldots, d\}$, each subset having at least two elements, then we will refer to our estimator as an extremal coefficients estimator.

We will address the optimal choice of $\Omega(\theta)$ below. The simplest choice for $\Omega(\theta)$ is the identity matrix I_q, yielding an ordinary least-squares estimator

$$\widehat{\theta}_{n,k} := \operatorname*{arg\,min}_{\theta \in \Theta} \sum_{m=1}^{q} \left(\widehat{\ell}_{n,k}(c_m) - \ell(c_m; \theta)\right)^2. \qquad (2.10)$$

This special case of our estimator is similar to the estimator proposed in Fougères et al. (2016) in the more specific context of fitting max-stable distributions to a random sample from such a distribution.

2.3 Consistency and asymptotic normality

If L is differentiable at an interior point $\theta \in \Theta$, its total derivative will be denoted by $\dot{L}(\theta) \in \mathbb{R}^{q \times p}$. Differentiability of the map $\theta \mapsto L(\theta)$ is a basic smoothness condition on the model; we do not assume differentiability of the map $x \mapsto \ell(x; \theta)$.

Theorem 1 (Existence, uniqueness and consistency) *Let $\{\ell(\,\cdot\,; \theta) : \theta \in \Theta\}$, with $\Theta \subset \mathbb{R}^p$, be a parametric family of d-variate stable tail dependence functions. Let $c_1, \ldots, c_q \in [0, \infty)^d$ be $q \geq p$ points such that the map $L : \theta \mapsto (\ell(c_m; \theta))_{m=1}^{q}$ is a homeomorphism from Θ to $L(\Theta)$. Let the true d-variate distribution function F have stable tail dependence function $\ell(\,\cdot\,; \theta_0)$ for some interior point $\theta_0 \in \Theta$. Assume* *that L is twice continuously differentiable on a neighbourhood of θ_0 and that $\dot{L}(\theta_0)$ is of full rank; also assume that $\Omega : \Theta \to \mathbb{R}^{q\times q}$ is twice continuously differentiable on a neighbourhood of θ_0. Assume $\lambda_1 := \inf_{\theta\in\Theta} \lambda_1(\theta) > 0$. Finally assume, for $m = 1, \dots, q$, and for a positive sequence $k = k_n$ satisfying (2.5),*

$$\widehat{\ell}_{n,k}(c_m) \xrightarrow{p} \ell(c_m; \theta_0), \qquad \text{as } n \to \infty. \tag{2.11}$$

Then with probability tending to one, the minimizer $\widehat{\theta}_{n,k}$ in Eq. 2.9 *exists and is unique. Moreover,*

$$\widehat{\theta}_{n,k} \xrightarrow{p} \theta_0, \qquad \text{as } n \to \infty.$$

Typically, the above conditions on the parametric family, on Ω, and on the initial estimator, can be verified. In Appendix B this is done for the parametric models considered in Sections 3.1–3.3.

Theorem 2 (Asymptotic normality) *If in addition to the assumptions of Theorem 1, the estimator $\widehat{\ell}_{n,k}$ satisfies*

$$\sqrt{k}\, D_{n,k}(\theta_0) = \left(\sqrt{k}\left\{\widehat{\ell}_{n,k}(c_m) - \ell(c_m; \theta_0)\right\}\right)_{m=1}^{q} \xrightarrow{d} \mathcal{N}_q\big(0, \Sigma(\theta_0)\big), \quad \text{as } n \to \infty, \tag{2.12}$$

for some $q \times q$ covariance matrix $\Sigma(\theta_0)$, then, as $n \to \infty$,

$$\sqrt{k}\,(\widehat{\theta}_{n,k} - \theta_0) = (\dot{L}^T \Omega \dot{L})^{-1} \dot{L}^T \Omega \sqrt{k}\, D_{n,k}(\theta_0) + o_p(1) \xrightarrow{d} \mathcal{N}_p\big(0, M(\theta_0)\big), \tag{2.13}$$

where the $p \times p$ covariance matrix $M(\theta_0)$ is defined by

$$M(\theta_0) := (\dot{L}^T \Omega \dot{L})^{-1}\, \dot{L}^T \Omega\, \Sigma(\theta_0)\, \Omega \dot{L}\, (\dot{L}^T \Omega \dot{L})^{-1},$$

and the matrices $\dot{L}$ and Ω are evaluated at θ_0.

Provided $\Sigma(\theta_0)$ is invertible, we can choose Ω in such a way that the asymptotic covariance matrix $M(\theta_0)$ is minimal, say $M_{\mathrm{opt}}(\theta_0)$, i.e., the difference $M(\theta_0) - M_{\mathrm{opt}}(\theta_0)$ is positive semi-definite. The minimum is attained at $\Omega(\theta_0) = \Sigma(\theta_0)^{-1}$ and the matrix $M(\theta_0)$ becomes simply

$$M_{\mathrm{opt}}(\theta_0) = \left(\dot{L}(\theta_0)^T\, \Sigma(\theta_0)^{-1}\, \dot{L}(\theta_0)\right)^{-1}, \tag{2.14}$$

see for instance Abadir and Magnus (2005, page 339). Now extend the covariance matrix $\Sigma(\theta_0)$ to the whole parameter space Θ by letting the map $\theta \mapsto \Sigma(\theta)$ be such that $\Sigma(\theta)$ is an invertible covariance matrix and $\Sigma^{-1} : \Theta \to \mathbb{R}^{q\times q}$ satisfies the assumptions on Ω.

Corollary 1 (Optimal weight matrix) *If the assumptions of Theorem 2 are satisfied and $\widehat{\theta}_{n,k}$ is the estimator based on the weight matrix $\Omega(\theta) = \Sigma(\theta)^{-1}$, then, with M_{opt} as in* Eq. 2.14*, we have*

$$\sqrt{k}(\widehat{\theta}_{n,k} - \theta_0) \xrightarrow{d} \mathcal{N}_p\big(0, M_{\mathrm{opt}}(\theta_0)\big), \qquad \text{as } n \to \infty. \tag{2.15}$$

The asymptotic covariance matrices M and M_{opt} in Eqs. 2.13 and 2.15, respectively, depend on the unknown parameter vector θ_0 through the matrices $\dot{L}(\theta)$, $\Omega(\theta)$ and $\Sigma(\theta)$ evaluated at $\theta = \theta_0$. If these matrices vary continuously with θ, then it is a standard procedure to construct confidence regions and hypothesis tests, cf. (Einmahl et al. 2012, Corollaries 4.3 and 4.4).

2.4 Goodness-of-fit testing

It is of obvious importance to be able to test the goodness-of-fit of the parametric family of tail dependence functions that we intend to use. The basis for such a test is $D_{n,k}(\widehat{\theta}_{n,k})$, the difference vector between the initial and parametric estimators of $\ell(c_m)$ at the estimated value of the parameter.

Corollary 2 *Under the assumptions of Theorem 2, we have*

$$\sqrt{k}\, D_{n,k}(\widehat{\theta}_{n,k}) = (I_q - P(\theta_0))\,\sqrt{k}\, D_{n,k}(\theta_0) + o_p(1)$$
$$\xrightarrow{d} \mathcal{N}_q\big(0,\, (I_q - P(\theta_0))\,\Sigma(\theta_0)\,(I_q - P(\theta_0))^T\big), \quad \text{as } n \to \infty, \tag{2.16}$$

where $P := \dot{L}(\dot{L}^T \Omega \dot{L})^{-1}\, \dot{L}^T \Omega$ *has rank* p *and* $I_q - P$ *has rank* $q - p$.

The easiest case in which Eq. 2.16 can be exploited is when $\Sigma(\theta)$ is invertible and $\Omega(\theta) = \Sigma(\theta)^{-1}$. Then it suffices to consider the minimum attained by the criterion function $f_{n,k}$ in Eq. 2.8, i.e., the test statistic is just $f_{n,k}(\widehat{\theta}_{n,k}) = \min_{\theta \in \Theta} f_{n,k}(\theta)$. Observe that it is important here that we allow Ω to depend on θ.

Corollary 3 *Let* $q > p$. *If the assumptions of Corollary 1 are satisfied, in particular if* $\Omega(\theta) = \Sigma(\theta)^{-1}$, *then*

$$k\, f_{n,k}(\widehat{\theta}_{n,k}) \xrightarrow{d} \chi^2_{q-p}, \qquad \text{as } n \to \infty.$$

If $\Omega(\theta)$ is different from $\Sigma(\theta)^{-1}$, for instance when $\Sigma(\theta)$ is not invertible, a goodness-of-fit test can still be based upon Eq. 2.16 by considering the spectral decomposition of the limiting covariance matrix. For convenience, we suppress the dependence on θ. Let

$$(I_q - P)\,\Sigma\,(I_q - P)^T = V D V^T$$

where $V = (v_1, \ldots, v_q)$ is an orthogonal $q \times q$ matrix, $V^T V = I_q$, the columns of which are orthonormal eigenvectors, and D is diagonal, $D = \operatorname{diag}(\nu_1, \ldots, \nu_q)$, with $\nu_1 \geq \ldots \geq \nu_q = 0$ the corresponding eigenvalues, at least p of which are zero, the rank of $I_q - P$ being $q - p$. Let $s \in \{1, \ldots, q - p\}$ be such that $\nu_s > 0$ and consider the $q \times q$ matrix

$$A := V_s D_s^{-1} V_s^T$$

where $D_s = \operatorname{diag}(\nu_1, \ldots, \nu_s)$ is an $s \times s$ diagonal matrix and where $V_s = (v_1, \ldots, v_s)$ is a $q \times s$ matrix having the first s eigenvectors as its columns.

Corollary 4 *If the assumptions of Theorem 2 hold and if $s \in \{1, \ldots, q-p\}$ is such that, in a neighbourhood of θ_0, $\nu_s(\theta) > 0$ and the matrix $A(\theta)$ depends continuously on θ, then*

$$k\, D_{n,k}(\widehat{\theta}_{n,k})^T\, A(\widehat{\theta}_{n,k})\, D_{n,k}(\widehat{\theta}_{n,k}) \xrightarrow{d} \chi_s^2, \qquad \text{as } n \to \infty.$$

Remark 1 If $\Sigma(\theta)$ is invertible for all θ, then we can set $s = q - p$ and $\Omega(\theta) = \Sigma(\theta)^{-1}$. The difference between the two test statistics in Corollaries 3 and 4 then converges to zero in probability, i.e., the two tests are asymptotically equivalent under the null hypothesis.

2.5 Choice of the initial estimator

Our estimator in Eq. 2.9 is flexible enough to allow for various initial estimators, perhaps based on exceedances over high thresholds or rather on vectors of componentwise block maxima extracted from a multivariate time series (Bücher and Segers 2014). Here we will focus on the former case, and more specifically on the empirical tail dependence function and a variant thereof.

For simplicity, we assume that the random vectors X_i, $i \in \{1, \ldots, n\}$, are not only identically distributed but also independent, so that they are a random sample from F. Let R_{ij}^n denote the rank of X_{ij} among $X_{1j}, \ldots, X_{nj}$ for $j = 1, \ldots, d$. For convenience, assume that F is continuous.

2.5.1 Empirical stable tail dependence function

A natural estimator of $\ell(x)$ is obtained by replacing F and $F_1, \ldots, F_d$ in Eq. 2.1 by their empirical counterparts and replacing t by k/n, yielding

$$\widetilde{\ell}'_{n,k}(x) := \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left\{R_{i1}^n > n + 1 - kx_1 \text{ or} \ldots \text{ or} R_{id}^n > n + 1 - kx_d\right\}. \tag{2.17}$$

This estimator, the empirical stable tail dependence function, was introduced for $d = 2$ in Huang (1992) and studied further in Drees and Huang (1998). A slight modification of it allows for better finite-sample properties,

$$\widetilde{\ell}_{n,k}(x) := \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left\{R_{i1}^n > n + 1/2 - kx_1 \text{ or} \ldots \text{ or} R_{id}^n > n + 1/2 - kx_d\right\}. \tag{2.18}$$

By Einmahl et al. (2012, Theorem 4.6), this estimator satisfies Eq. 2.12 under conditions controlling the rate of convergence in Eq. 2.1 and the growth rate of the intermediate sequence $k = k_n$. The first-order partial derivatives $\dot{\ell}_j(x; \theta_0)$ of $x \mapsto \ell(x; \theta_0)$ are assumed to exist and to be continuous in neighbourhoods of the points c_m for which $c_{mj} > 0$.

In this case, the entries of the matrix $\Sigma(\theta)$ in Eq. 2.12, for θ in the interior of Θ, are, for $i, j \in \{1, \ldots, q\}$, given by

$$\Sigma_{i,j}(\theta) = \mathbb{E}[B(c_i)\, B(c_j)], \tag{2.19}$$

with $B(c_i) := W_\ell(c_i) - \sum_{j=1}^{d} \dot{\ell}_j(c_i) W_\ell(c_{ij}\, e_j)$ and with $(W_\ell(x) : x \in [0,\infty)^d)$ a zero-mean Gaussian process with covariance function $\mathbb{E}[W_\ell(x)\, W_\ell(y)] = \ell(x) + \ell(y) - \ell(x \vee y)$, the maximum being taken componentwise. For points c_i of the form e_J in Eq. 2.3, the expectation in Eq. 2.19 can be calculated as follows: for non-empty subsets J and K of $\{1,\ldots,d\}$,

$$\begin{aligned}\mathbb{E}[B(e_J)\, B(e_K)] = \ell_J + \ell_K - \ell_{J\cup K} &- \sum_{j\in J} \dot{\ell}_{j,J}\, (1 + \ell_K - \ell_{\{j\}\cup K}) \\ &- \sum_{k\in K} \dot{\ell}_{k,K}\, (\ell_J + 1 - \ell_{J\cup\{k\}}) \\ &+ \sum_{j\in J}\sum_{k\in K} \dot{\ell}_{j,J}\dot{\ell}_{k,K}\, (2 - \ell_{\{j,k\}}),\end{aligned}$$

where $\ell_J := \ell(e_J;\theta_0)$ and $\dot{\ell}_{j,J} := \dot{\ell}_j(e_J;\theta_0)$.

2.5.2 Bias-corrected estimator

A drawback of $\widetilde{\ell}_{n,k}$ in Eq. 2.18 is its possibly quickly growing bias as k increases. Recently, two bias-corrected estimators have been proposed. We consider here the kernel-type estimator of Beirlant et al. (2016), which is partly based on (the one in) Fougères et al. (2015).

Consider first a rescaled version of $\widetilde{\ell}'_{n,k}$ in Eq. 2.17, defined as $\widetilde{\ell}_{n,k,a}(x) := a^{-1}\widetilde{\ell}'_{n,k}(ax)$ for $a > 0$. Then define the weighted average

$$\breve{\ell}_{n,k}(x) := \frac{1}{k}\sum_{j=1}^{k} K(a_j)\, \widetilde{\ell}_{n,k,a_j}(x), \qquad a_j := \frac{j}{k+1},\ j \in \{1,\ldots,k\}, \tag{2.20}$$

where K is a kernel function, i.e., a positive function on $(0, 1)$ such that $\int_0^1 K(u)\, du = 1$.

In addition to Eq. 2.1, we assume there exist a positive function α on $(0,\infty)$ tending to 0 as $t \downarrow 0$ and a non-zero function M on $[0,\infty)^d$ such that for all $x \in [0,\infty)^d$,

$$\lim_{t\downarrow 0} \frac{1}{\alpha(t)}[t^{-1}\mathbb{P}\{1 - F_1(X_{11}) \le tx_1 \text{ or} \ldots \text{ or} 1 - F_d(X_{1d}) \le tx_d\} - \ell(x)] = M(x). \tag{2.21}$$

Moreover, we assume a third-order condition on ℓ (Beirlant et al. 2016, equation (3)). In Beirlant et al. (2016, Theorem 1) the asymptotic distribution of $\breve{\ell}_{n,k}$ in Eq. 2.20 is derived under these three assumptions and for intermediate sequences $k = k_n$ growing faster than the ones considered above. A non-zero asymptotic bias term arises and the idea is to estimate and remove it, thereby obtaining a possibly more accurate estimator.

In order to achieve this bias reduction, the rate function, α, and its index of regular variation, β, need to be estimated. Consider another intermediate sequence

$k_1 = k_{1,n}$ such that $k/k_1 \to 0$. The bias-corrected estimator is then defined as

$$\overline{\ell}_{n,k,k_1}(x) := \frac{\breve{\ell}_{n,k}(x) - (k_1/k)^{\widehat{\beta}_{k_1}(x)} \widehat{\alpha}_{k_1}(x) \frac{1}{k} \sum_{j=1}^{k} K(a_j) a_j^{-\widehat{\beta}_{k_1}(X)}}{\frac{1}{k} \sum_{j=1}^{k} K(a_j)},$$

where $\widehat{\alpha}_{k_1}$ and $\widehat{\beta}_{k_1}$ are the estimators of α and β defined in Beirlant et al. (2016). Under the mentioned conditions, asymptotic normality as in Eq. 2.12 holds, where the limiting random vector is equal in distribution to $\int_0^1 K(u) u^{-1/2} \, \mathrm{d}u$ times the one corresponding to $\widetilde{\ell}_{n,k}$. Here, the growth rate of k here can be taken faster than when using $\widetilde{\ell}_{n,k}$.

A simple choice for K is a power kernel, i.e, $K(t) = (\tau + 1)t^{\tau}$ for $t \in (0, 1)$ and $\tau > -1/2$. Then $\int_0^1 K(u) \, u^{-1/2} \, \mathrm{d}u = (2 + \tau)/(1 + 2\tau)$. Note that this factor tends to 1 if $\tau \to \infty$. In practice, we take $\tau = 5$ as recommended in Beirlant et al. (2016).

3 Simulation studies

We conduct simulation studies for data in the max-domain of attraction of the logistic model, the Brown–Resnick process and the max-linear model. For each model, we report the empirical bias, standard deviation, and root mean squared error (RMSE) of our estimators. We also study the finite-sample performance of the goodness-of-fit statistic of Corollary 3. All simulations were done in the R statistical software environment (R Core Team 2015). The programs used to calculate our estimator are available in the R package tailDepFun (Kiriliouk 2016).

3.1 Logistic model: comparison with likelihood methods

The d-dimensional logistic model has stable tail dependence function

$$\ell(x_1, \ldots, x_d; \theta) = \left(x_1^{1/\theta} + \cdots + x_d^{1/\theta}\right)^{\theta}, \qquad \theta \in [0, 1].$$

The domain-of-attraction condition (2.1) holds for instance if F has continuous margins and its copula is Archimedean with generator $\phi(t) = 1/(t^{\theta} + 1)$, also known as the outer power Clayton copula (Hofert et al. 2015).

In Huser et al. (2016), a comprehensive comparison of likelihood estimators for θ has been performed based on random samples from this copula. We compare those results to our extremal coefficients estimator, i.e., the weighted least squares estimator based on points c_m of the form e_J, with J ranging in the collection

$$\mathcal{Q}_a := \left\{J \subset \{1, \ldots, d\} : |J| = a\right\} \tag{3.1}$$

for $a \in \{2, 3\}$. Moreover, we let $\Omega(\theta)$ be the identity matrix, since by exchangeability of the model, a weighting procedure can bring no improvements.

Following Huser et al. (2016, Section 4.2), we simulated 10 000 random samples of size $n = 10\,000$ from the outer power Clayton copula. For the likelihood-based estimators, the margins are standardized to the unit Pareto scale via the rank transformation

$$X^*_{ij} := \frac{n}{n + 1/2 - R^n_{ij}}, \qquad i \in \{1, \ldots, n\}, \ j \in \{1, \ldots, d\}.$$

Again as in Huser et al. (2016, Section 4.2), we take dimension $d \in \{2, 5, 10, 15, 20, 25, 30\}$ and parameter $\theta \in \{0.3, 0.6, 0.9, 0.95\}$. Note that in the likelihood setting, this is a very demanding experiment, and three of the ten likelihood-based estimators considered in Huser et al. (2016) are only computed for $d \in \{2, 5, 10\}$. In Huser et al. (2016), threshold probabilities are set to 0.98, corresponding to $k = 200$ in our setup.

Figure 1 shows the RMSE of three estimators based on the empirical tail dependence function: the two extremal coefficients estimators mentioned above and the pairwise M-estimator of Einmahl et al. (2016) as implemented in the R package tailDepFun (Kiriliouk 2016). When dependence is strong, $\theta = 0.3$, the estimator based on $\mathcal{Q}_3$ performs best, whereas when dependence is weak, $\theta = 0.9$ or

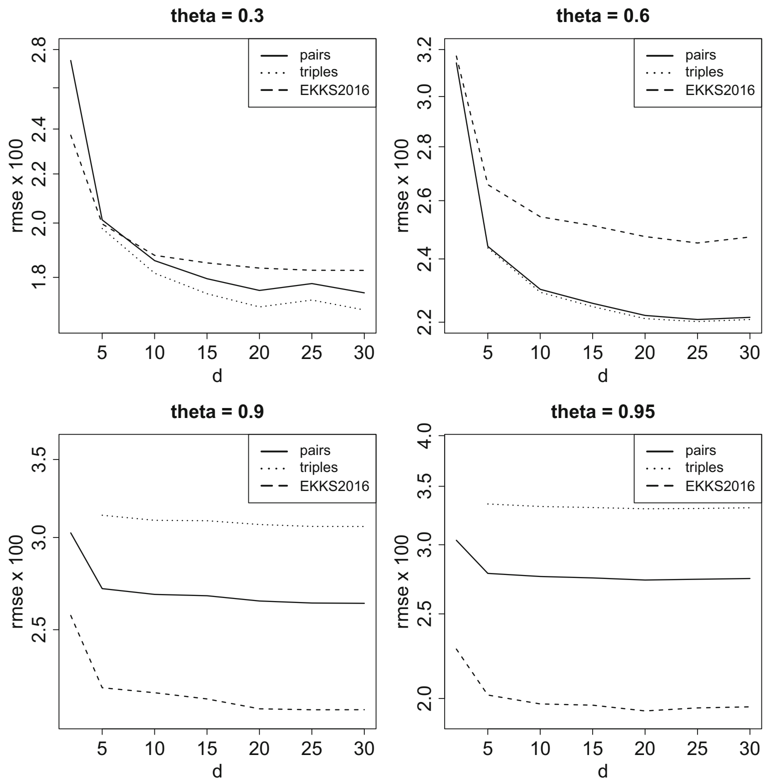


Fig. 1 Logistic model: RMSE (on a logarithmic scale) for the estimators; 10 000 samples of size $n = 10\,000$

$\theta = 0.95$, the estimator based on $\mathcal{Q}_2$ performs better than the estimator based on $\mathcal{Q}_3$. Note also that when the dependence is not too weak, the estimators based on extremal coefficients perform better than the pairwise M-estimator of Einmahl et al. (2016). Finally, our estimation procedures have almost constant RMSE as the dimension increases, in line with the pairwise composite likelihood methods studied in Huser et al. (2016).

Comparing these results to the ten likelihood-based estimators in Huser et al. (2016, Figure 4), we see that our estimators are strong competitors in the sense that they rank highly when comparing RMSEs, and are not dominated by one of the likelihood-based estimators. More precisely, for $\theta = 0.3$, only the likelihood estimators based on the Poisson process representation (Coles and Tawn 1991) and the multivariate Generalized Pareto distribution outperform our estimators; for $\theta = 0.6$, the same two likelihood estimators outperform ours, but only for $d \geq 15$; finally, for $\theta = 0.9$ and $\theta = 0.95$ only the pairwise censored likelihood estimator (Huser and Davison 2014) has a smaller RMSE than our estimators.

3.2 Brown–Resnick process

The Brown–Resnick process on a planar set $\mathcal{S} \subset \mathbb{R}^2$ is given by

$$Y(s) = \max_{i \in \mathbb{N}} \xi_i \exp\{\epsilon_i(s) - \gamma(s)\}, \qquad s \in \mathcal{S}, \tag{3.2}$$

where $\{\xi_i\}_{i \geq 1}$ is a Poisson process on $(0, \infty)$ with intensity measure $\xi^{-2}\,d\xi$ and $\{\epsilon_i(\,\cdot\,)\}_{i \geq 1}$ are independent copies of a Gaussian process ϵ with stationary increments such that $\epsilon(0) = 0$ and with variance $2\gamma(\,\cdot\,)$ and semi-variogram $\gamma(\,\cdot\,)$. In Kabluchko et al. (2009) it is shown that the Brown–Resnick process with $\gamma(s) = (\|s\|/\rho)^\alpha$ is the only possible limit of (rescaled) maxima of stationary and isotropic Gaussian random fields; here $\rho > 0$ and $0 < \alpha \leq 2$.

For d locations $s_1, \ldots, s_d \in \mathcal{S}$, the distribution of the random vector $(Y(s_i))_{i=1}^d$ is max-stable with tail dependence function ℓ depending on $\gamma(\,\cdot\,)$. From Huser and Davison (2013), we obtain the following representation for the extremal coefficients ℓ_J in (2.4). Let $\Phi_a(\,\cdot\,; R)$ denote the cumulative distribution function of the $\mathcal{N}_a(0, R)$ distribution. Then we have

$$\ell_J = \sum_{j \in J} \Phi_{|J|-1}(\eta^{(j)}; R^{(j)}), \qquad J \subset \{1, \ldots, d\},\ J \neq \emptyset, \tag{3.3}$$

where $\eta^{(j)} = \left(\sqrt{\gamma(s_j - s_i)/2}\right)_{i \in J \setminus \{j\}} \in \mathbb{R}^{|J|-1}$ and where $R^{(j)}$ is a $(|J|-1) \times (|J|-1)$ correlation matrix with entries given by

$$\frac{\gamma(s_j - s_i) + \gamma(s_j - s_k) - \gamma(s_i - s_k)}{2\sqrt{\gamma(s_j - s_i)\,\gamma(s_j - s_k)}}, \qquad i, k \in J \setminus \{j\}.$$

When $J = \{j_1, j_2\}$ for $j_1, j_2 \in \{1, \ldots, d\}$, expression (3.3) simplifies to $\ell_J = 2\,\Phi\left(\sqrt{\gamma(s_{j_1} - s_{j_2})/2}\right)$.

We simulate 300 random samples of size $n = 1000$ from the Brown–Resnick process on a 3×4 unit distance grid using the R package SpatialExtremes (Ribatet

2015). To arrive at a more realistic estimation problem, we perturb the samples thus obtained with additive noise, i.e., if $Y_i = (Y_{i1}, \dots, Y_{id})$ is an observation from the Brown–Resnick process, then we set $X_{ij} = Y_{ij} + |\epsilon_{ij}|$ for $i = 1, \dots, n$ and $j = 1, \dots, d$, where ϵ_{ij} are independent $\mathcal{N}(0, 1/4)$ random variables.

We estimate the parameters $(\alpha, \rho) = (1, 1)$ using the extremal coefficients estimator based on the subset of $\mathcal{Q}_2$ in Eq. 3.1 consisting of pairs of neighbouring locations, i.e., locations that are at most a distance $\sqrt{2}$ apart. This leads to $q = 29$ pairs. Including pairs of locations that are further away tends to drastically increase the bias (Einmahl et al. 2016).

The upper panels of Fig. 2 show the bias, standard deviation and RMSE for three estimators: the estimator based on the empirical tail dependence function with $\Omega(\theta) = \Sigma(\theta)^{-1}$ (solid lines), the estimator based on the bias-corrected tail dependence function with $\Omega(\theta) = \Sigma(\theta)^{-1}$ (dotted lines), and the pairwise M-estimator from Einmahl et al. (2016) (dashed lines). We see that for the estimation of the shape parameter $\alpha = 1$ it is better to use one of the estimators based on the empirical stable tail dependence function, whereas for the scale parameter $\rho = 1$ the bias-corrected estimator performs better.

To show the feasibility of the estimation procedure in high dimensions, we simulate 300 samples of size $n = 1000$ from the perturbed Brown–Resnick process on a 10×15 unit-distance grid ($d = 150$), using again $(\alpha, \rho) = (1, 1)$ and selecting pairs of neighbouring locations only, yielding $q = 527$ pairs in total. The bottom panels of Fig. 2 show the bias, standard deviation and RMSE for the estimator based on the empirical tail dependence function with $\Omega(\theta) = I_q$ (solid lines), the estimator based on the bias-corrected tail dependence function with $\Omega(\theta) = I_q$ (dotted lines), and the pairwise M-estimator from Einmahl et al. (2016) (dashed lines). Compared to $d = 12$ above, the estimation of α has improved whereas the estimation quality of ρ stays roughly the same.

3.3 Max-linear models on directed acyclic graphs

A max-linear or max-factor model has stable tail dependence function

$$\ell(x) = \sum_{t=1}^{r} \max_{j=1,\dots,d} b_{jt} x_j, \qquad x \in [0, \infty)^d, \tag{3.4}$$

where the factor loadings b_{jt} are non-negative constants such that $\sum_{t=1}^{r} b_{jt} = 1$ for every $j \in \{1, \dots, d\}$ and all column sums of the $d \times r$ matrix $B := (b_{jt})_{j,t}$ are positive (Einmahl et al. 2012). An example of a random vector $Y = (Y_1, \dots, Y_d)$ that has tail dependence function (3.4) is $Y_j = \max_{t=1,\dots,r} b_{jt} Z_t$ for $j \in \{1, \dots, d\}$, where $Z_1, \dots, Z_r$ are independent unit Fréchet variables. The random variables Y_j are then unit Fréchet as well.

Since the rows of B sum up to one, it has only $d \times (r - 1)$ free elements. Further structure may be added to the coefficient matrix B, leading to parametric models whose parameter dimension is lower than $d \times (r - 1)$; see below. Even then, the map L in Eq. 2.6 induced by restricting the points c_m to be of the form e_J in (2.3) is

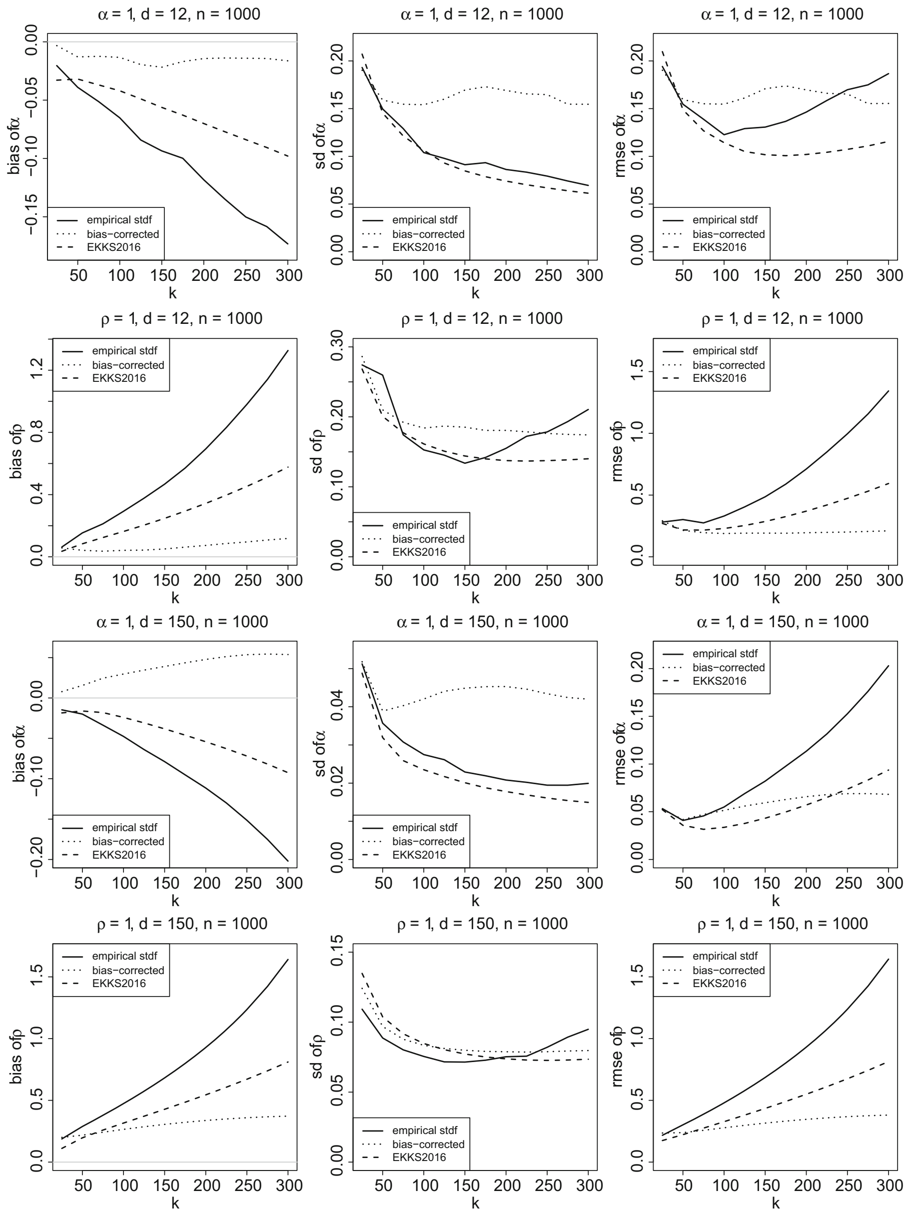


Fig. 2 Brown–Resnick process: bias, standard deviation and RMSE for the estimators in $d = 12$ (upper panels) and $d = 150$ (lower panels); 300 samples of size $n = 1000$

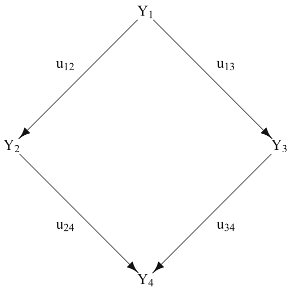
typically not one-to-one. Therefore, we need more general choices of the points c_m in the definition of the estimator.

In Gissibl and Klüppelberg (2017), a link is established between max-linear models and structural equation models, from which graphical models based on directed acyclic graphs (DAGs) can be constructed. A recursive max-linear model is defined via

$$Y_j = \max_{k \in \mathrm{pa}(j)} u_{kj} Y_k \vee u_j Z_j, \qquad j = 1, \ldots, d,$$

where $\mathrm{pa}(j) \subset \{1, \ldots, d\}$ denotes the set of parents of node j in the graph, $u_{kj} > 0$ for all $k \in \mathrm{pa}(j)$, and $u_j > 0$ for all $j \in \{1, \ldots, d\}$. We let $Z_1, \ldots, Z_d$ be independent unit Fréchet random variables. A recursive max-linear model can then be written as a max-linear model with parameters determined by the paths of the corresponding graph.

We focus on the four-dimensional model corresponding to the following directed acyclic graph (Gissibl and Klüppelberg 2017, Example 2.1):



$$\begin{aligned}
Y_1 &= u_1 Z_1, \\
Y_2 &= u_{12} Y_1 \vee u_2 Z_2 = u_{12} u_1 Z_1 \vee u_2 Z_2, \\
Y_3 &= u_{13} Y_1 \vee u_3 Z_3 = u_{13} u_1 Z_1 \vee u_3 Z_3, \\
Y_4 &= u_{24} Y_2 \vee u_{34} Y_3 \vee u_4 Z_4 \\
&= (u_{24} u_{12} u_1 \vee u_{34} u_{13} u_1) Z_1 \vee u_{24} u_2 Z_2 \vee u_{34} u_3 Z_3 \vee u_4 Z_4.
\end{aligned}$$

If we require $Y_1, \ldots, Y_4$ to be unit Fréchet, the matrix of factor loadings becomes

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ u_{12} & u_2 & 0 & 0 \\ u_{13} & 0 & u_3 & 0 \\ u_{12} u_{24} \vee u_{13} u_{34} & u_2 u_{24} & u_3 u_{34} & u_4 \end{pmatrix},$$

where the diagonal elements u_j for $j \in \{2, 3, 4\}$ are such that the row sums are equal to one. The parameter vector is then given by $\theta = (u_{12}, u_{13}, u_{24}, u_{34})$.

We conduct a simulation study based on 300 samples of size $n = 1000$ from the four-dimensional model with stable tail dependence function (3.4) and B as above, with parameter vector $\theta = (0.3, 0.8, 0.4, 0.55)$. As before, we put $X_{ij} = Y_{ij} + |\epsilon_{ij}|$, with $(Y_{i1}, \ldots, Y_{id})$ as above and ϵ_{ij} independent $\mathcal{N}(0, 1/4)$ random variables. The estimators are based on the $q = 72$ points c_m on the grid $\{0, 1/2, 1\}^4$ having at least two positive coordinates.

Figure 3 shows the RMSE for the estimator based on the empirical tail dependence function with $\Omega(\theta) = \Sigma(\theta)^{-1}$ (solid lines), the estimator based on the bias-corrected tail dependence function with $\Omega(\theta) = \Sigma(\theta)^{-1}$ (dotted lines) and the pairwise M-estimator from Einmahl et al. (2016) (dashed lines). The difference between the pairwise M-estimator and our estimators based on the empirical tail dependence function is negligible. The estimators based on the empirical tail dependence function perform better than the ones based on the bias-corrected version, especially for the parameters u_{13} and u_{24}.

Remark 2 For the weight matrix, we actually defined $\Omega(\theta)$ as $(\Sigma(\theta) + cI_q)^{-1}$ for some small $c > 0$. The reason for applying such a Tikhonov correction is that some

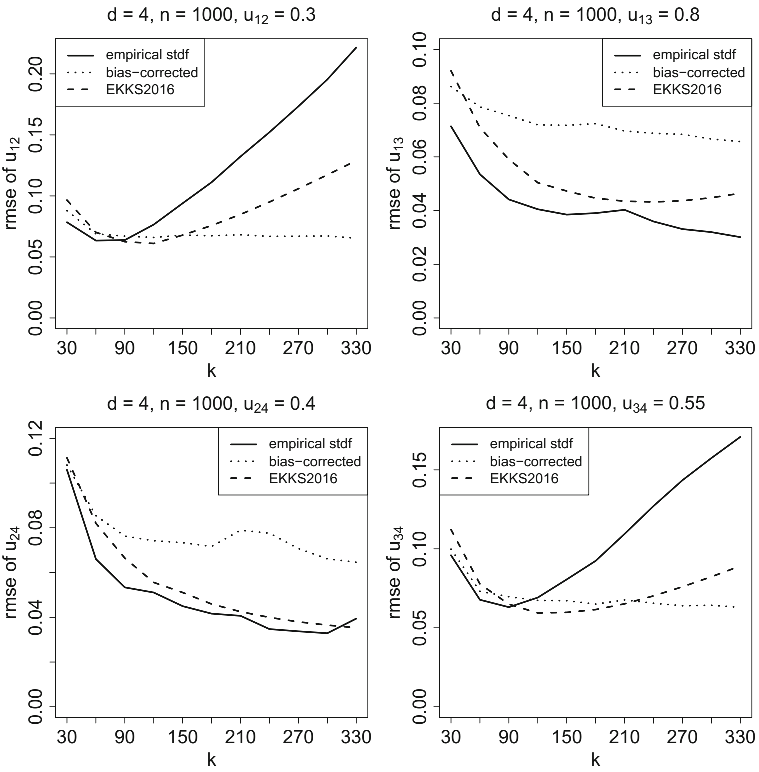


Fig. 3 Max-linear structural equation model based on a directed acylic graph: RMSE for the estimators; 300 samples of size $n = 1000$

eigenvalues of $\Sigma(\theta)$ are (near) zero, which can in turn be due to the fact that for max-linear models such as here, $\ell(c_m; \theta)$ may hit its lower bound $\max(c_{m,1}, \ldots, c_{m,d})$ for some $m \in \{1, \ldots, q\}$.

3.4 Goodness-of-fit test

We compare the performance of the goodness-of-fit test presented in Corollary 3 to the three goodness-of-fit test statistics κ_n, ω_n^2, and A_n^2 proposed in Can et al. (2015, page 18). In the simulation study there, the observed rejection frequencies are reported at the 5% significance level under null and alternative hypotheses for two bivariate models for ℓ; a bivariate logistic model with $\theta \in (0, 1)$ and

$$\ell(x_1, x_2; \psi) = (1 - \psi)(x_1 + x_2) + \psi\sqrt{x_1^2 + x_2^2}, \qquad \psi \in [0, 1], \tag{3.5}$$

i.e., a mixture between the logistic model and tail independence. For both models, they generate 300 samples of size $n = 1500$ from a "null hypothesis" distribution function, for which the model is correct, and 100 samples of $n = 1500$ from an "alternative hypothesis" distribution function, for which the model is incorrect. These distribution functions are described in equations (32), (33), (35), and (36) of Can et al. (2015). We take $c_m \in \{(1/2, 1/2), (1/2, 1), (1, 1/2), (1, 1)\}$, $m = 1, \ldots, 4$, and $k = 200$.

Table 1 shows the observed fractions of Type I errors under the null hypotheses and the observed fraction of rejections under the alternative hypotheses. The results for κ_n, ω_n^2, and A_n^2 are taken from Can et al. (2015, Table 1). We see that our goodness-of-fit test performs comparably to the test statistics in Can et al. (2015).

It should be noted that the tests are of very different nature. The three test statistics in Can et al. (2015) are functionals of a transformed empirical process and are therefore of omnibus-type. The results in there are based on the full max-domain of attraction condition on F and the procedure is computationally complicated and therefore difficult to apply in dimensions (much) higher than two. The present test only performs comparisons at q points and avoids integration. Therefore it is computationally much easier to apply in dimension $d > 2$.

To illustrate the power of our test in more detail, we suppose that under the null hypothesis ℓ is from a bivariate logistic model and we generate samples from the asymmetric logistic model with stable tail dependence function

$$\ell(x_1, x_2; \theta, \phi) = (1-\phi)x_2 + \left(x_1^{1/\theta} + (\phi x_2)^{1/\theta}\right)^\theta, \qquad \theta, \phi \in [0, 1].$$

Figure 4 shows the power of our test as a function of ϕ for $\theta = 0.5$. If ϕ decreases, the power becomes large. Note that $\phi = 0$ yields a point on the boundary of the parameter space under the null hypothesis.

4 Tail dependence in European stock markets

We analyze data from the EURO STOXX 50 Index, which represents the performance of the largest 50 companies among 19 different "supersectors" within the 12 main Eurozone countries. Since Germany (DE) and France (FR) together form 68% of the index, we will focus on these two countries only. Every company belongs to a supersector, of which there are 19 in total. We select two of them as an illustration: chemicals and insurance. We study the following five stocks: Bayer (DE, chemicals), BASF (DE, chemicals), Allianz (DE, insurance), Airliquide (FR, chemicals), and Axa (FR, insurance), and we take the weekly negative log-returns of the stock prices

Table 1 Observed rejection frequencies at the 5% significance level under null and alternative hypotheses

	Null		Alternative	
	logistic	mixture	logistic	mixture
κ_n	19/300	9/300	92/100	97/100
ω_n^2	11/300	13/300	90/100	97/100
A_n^2	17/300	18/300	95/100	100/100
$kf_{n,k}(\widehat{\theta}_{n,k})$	16/300	14/300	100/100	82/100

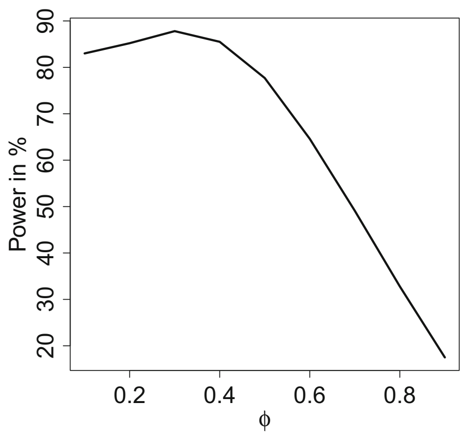


Fig. 4 Power for the asymmetric model for $k = 200$; 1000 samples of size $n = 1500$

of these companies from Yahoo Finance[1] for the period January 2002 to November 2015, leading to a sample of size $n = 711$. This dataset and the functions used to analyse it are available in the R package tailDepFun (Kiriliouk 2016).

We fit a structural equation model based on the directed acyclic graph given in Fig. 5. The nodes DE and FR are represented by their national stock market indices, the DAX and the CAC40, respectively, and the nodes chemicals and insurance are represented by corresponding sub-indices of the EURO STOXX 50 Index. Note that this is a model for the tail dependence function only, i.e., we only assume that the joint distribution of the negative log-returns has tail dependence function ℓ as in Eq. 3.4 with coefficient matrix B given in Table 2. We have $d = 10$ and the parameter vector is given by $\theta = (u_{12}, u_{13}, u_{14}, u_{15}, u_{26}, u_{46}, u_{27}, u_{47}, u_{38}, u_{48}, u_{39}, u_{59}, u_{2,10}, u_{5,10})$.

We perform the goodness-of-fit test described in Corollary 4, based on the $q = 1140$ points c_m in the grid $\{0, 1/2, 1\}^{10}$ having either two or three non-zero coordinates. We take $\Omega(\theta) = I_q$, $k = 40$, and we choose s such that $\nu_s > 0.1$, leading in this case to $s = 11$. The value of the test statistic is 5.28; the 95% quantile of a χ^2_{11} distribution is 19.68, so that the tail dependence model is not rejected.

The resulting parameter estimates are pictured at the edges of Fig. 5, where the relative width of each edge is proportional to its parameter value. The standard errors are given in parentheses. We note that, except for Allianz, the influence of the stock market indices DAX and CAC40 is (much) stronger than the influence of the sector indices chemicals and insurance.

5 Discussion

We have not addressed the number and choices of the points $c_1, \ldots, c_q$. The more points, the lower the asymptotic variance of the estimator. However, because q determines the dimension of Σ and thus of the weight matrix, choosing many points may

[1] http://finance.yahoo.com/

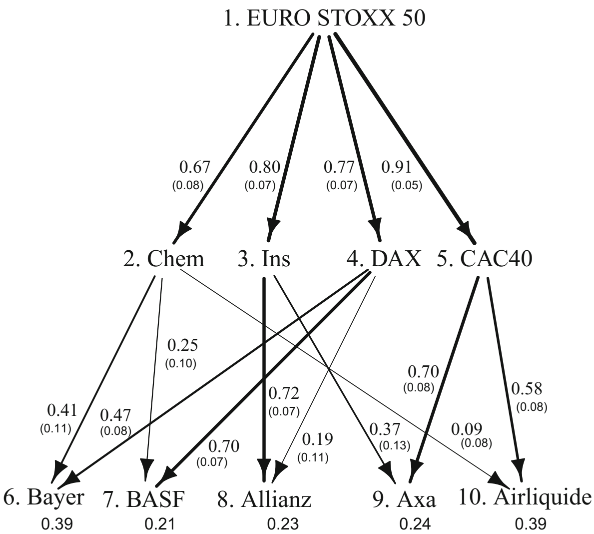


Fig. 5 European stock market data: directed acyclic graph with 14 parameters, whose estimates are shown near the corresponding edges. The relative width of each edge is proportional to its parameter value. The bottom row shows the estimated diagonal elements $u_6, \ldots, u_{10}$ of the matrix B in Table 2

cause the inverse of Σ to become numerically unstable. Moreover, for spatial models and the extremal coefficients estimator,one sees that using points c_m involving locations far away from another tends to increase the bias. Overall, the optimal number of points and their values depend highly on the choice of the parametric model. In the max-linear model, the choice of the points c_m is even more important since it influences the identifiability of the parameter vector; see Appendix B.

Table 2 European stock market data: coefficient matrix of the max-linear model stemming from the directed acyclic graph in Fig. 5

$$B = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
u_{12} & u_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
u_{13} & 0 & u_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
u_{14} & 0 & 0 & u_4 & 0 & 0 & 0 & 0 & 0 & 0 \\
u_{15} & 0 & 0 & 0 & u_5 & 0 & 0 & 0 & 0 & 0 \\
u_{12}u_{26} \vee u_{14}u_{46} & u_2u_{26} & 0 & u_4u_{46} & 0 & u_6 & 0 & 0 & 0 & 0 \\
u_{12}u_{27} \vee u_{14}u_{47} & u_2u_{27} & 0 & u_4u_{47} & 0 & 0 & u_7 & 0 & 0 & 0 \\
u_{13}u_{38} \vee u_{14}u_{48} & 0 & u_3u_{38} & u_4u_{48} & 0 & 0 & 0 & u_8 & 0 & 0 \\
u_{13}u_{39} \vee u_{15}u_{59} & 0 & u_3u_{39} & 0 & u_5u_{59} & 0 & 0 & 0 & u_9 & 0 \\
u_{12}u_{2,10} \vee u_{15}u_{5,10} & u_2u_{2,10} & 0 & 0 & u_5u_{5,10} & 0 & 0 & 0 & 0 & u_{10}
\end{pmatrix}$$

The diagonal elements u_i, for $i = 2; ... ; 10$ are such that the rows sum up to one. European stock market data: coefficient matrix...

The effect of the optimal weight matrix also depends on the choice of the model and the dimension d of the problem. The higher the dimension (and thus the more points c_m), the smaller the effect of the optimal weight matrix on the quality of our estimator. Moreover, the continuous updating procedure is relatively slow when q is large.

Finally, although our approach allows us to construct hypothesis tests, not all cases of interest are covered by our theory. For instance, the Smith model is a submodel of the Brown–Resnick process when $\alpha = 2$, or in max-linear models one could be interested in testing whether a factor loading is zero. These type of hypotheses concern parameters on the boundary of the parameter space and are the subject of current research by the second-named author.

Acknowledgments The research by A. Kiriliouk was funded by a FRIA grant of the "Fonds de la Recherche Scientifique – FNRS" (Belgium). J. Segers gratefully acknowledges funding by contract "Projet d'Actions de Recherche Concertées" No. 12/17-045 of the "Communauté française de Belgique" and by IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy).

Appendix A: Proofs

Proof of Theorem 1 This proof follows the same lines as the one of (Einmahl et al. 2016, Proof of Theorem 1). Let $\varepsilon_0 > 0$ be such that the closed ball $B_{\varepsilon_0}(\theta_0) = \{\theta : \|\theta - \theta_0\| \le \varepsilon_0\}$ is a subset of Θ; such an ε_0 exists since θ_0 is an interior point of Θ. Fix $\varepsilon > 0$ such that $0 < \varepsilon \le \varepsilon_0$. Let, more precisely than in Eq. 2.9, $\widehat{\Theta}_{n,k}$ be the *set* of minimizers of the right-hand side of Eq. 2.9. We show first that

$$\mathbb{P}[\widehat{\Theta}_{n,k} \neq \emptyset \text{ and } \widehat{\Theta}_{n,k} \subset B_\varepsilon(\theta_0)] \to 1, \qquad n \to \infty. \tag{3.6}$$

Because L is a homeomorphism, there exists $\delta > 0$ such that for $\theta \in \Theta$, $\|L(\theta) - L(\theta_0)\| \le \delta$ implies $\|\theta - \theta_0\| \le \varepsilon$. Equivalently, for every $\theta \in \Theta$ such that $\|\theta - \theta_0\| > \varepsilon$ we have $\|L(\theta) - L(\theta_0)\| > \delta$. Define the event

$$A_n = \left\{ \|L(\theta_0) - \widehat{L}_{n.k}\| < \frac{\delta\sqrt{\lambda_1}}{(1+\sqrt{\lambda_1})\max(1, \sqrt{\lambda_q(\theta_0)})} \right\}.$$

If $\theta \in \Theta$ is such that $\|\theta - \theta_0\| > \varepsilon$, then on the event A_n, we have

$$\begin{aligned}
\|D_{n,k}(\theta)\|_{\Omega(\theta)} &\ge \sqrt{\lambda_1(\theta)}\|D_{n,k}(\theta)\| \\
&\ge \sqrt{\lambda_1}\|L(\theta_0) - L(\theta) - \big(L(\theta_0) - \widehat{L}_{n,k}\big)\| \\
&\ge \sqrt{\lambda_1}\left(\|L(\theta_0) - L(\theta)\| - \|L(\theta_0) - \widehat{L}_{n,k}\|\right) \\
&> \sqrt{\lambda_1}\left(\delta - \frac{\delta\sqrt{\lambda_1}}{1+\sqrt{\lambda_1}}\right) = \frac{\delta\sqrt{\lambda_1}}{1+\sqrt{\lambda_1}}.
\end{aligned}$$

It follows that on A_n,

$$\begin{aligned}
\inf_{\theta:\|\theta-\theta_0\|>\varepsilon} \|D_{n,k}(\theta)\|_{\Omega(\theta)} &\ge \frac{\delta\sqrt{\lambda_1}}{1+\sqrt{\lambda_1}} > \sqrt{\lambda_q(\theta_0)}\|L(\theta_0) - \widehat{L}_{n,k}\| \\
&\ge \|L(\theta_0) - \widehat{L}_{n,k}\|_{\Omega(\theta_0)} \ge \inf_{\theta:\|\theta-\theta_0\|\le\varepsilon} \|L(\theta) - \widehat{L}_{n,k}\|_{\Omega(\theta)}.
\end{aligned}$$

The infimum on the right-hand side is actually a minimum since L is continuous and $B_\varepsilon(\theta_0)$ is compact. Hence on A_n the set $\widehat{\Theta}_{n,k}$ is non-empty and $\widehat{\Theta}_{n,k} \subset B_\varepsilon(\theta_0)$. To show Eq. 3.6, it remains to prove that $\mathbb{P}[A_n] \to 1$ as $n \to \infty$, but this follows from Eq. 2.11.

Next we will prove that, with probability tending to one, $\widehat{\Theta}_{n,k}$ has exactly one element, i.e., the function $f_{n,k}$ has a unique minimizer. To do so, we will show that there exists $\varepsilon_1 \in (0, \varepsilon_0]$ such that, with probability tending to one, the Hessian of $f_{n,k}$ is positive definite on $B_{\varepsilon_1}(\theta_0)$ and thus $f_{n,k}$ is strictly convex on $B_{\varepsilon_1}(\theta_0)$. In combination with Eq. 3.6 for $\varepsilon \in (0, \varepsilon_1]$, this will yield the desired conclusion.

For $\theta \in \Theta$, define the symmetric $p \times p$ matrix $\mathcal{H}(\theta; \theta_0)$ by

$$\begin{aligned}(\mathcal{H}(\theta;\theta_0))_{i,j} := {} & 2\left(\frac{\partial L(\theta)}{\partial\theta_j}\right)^T \Omega(\theta)\left(\frac{\partial L(\theta)}{\partial\theta_i}\right) - 2\left(\frac{\partial^2 L(\theta)}{\partial\theta_j\partial\theta_i}\right)^T \\ & \times\Omega(\theta)\big(L(\theta_0)-L(\theta)\big) - 2\left(\frac{\partial L(\theta)}{\partial\theta_i}\right)^T \frac{\partial\Omega(\theta)}{\partial\theta_j}\big(L(\theta_0)-L(\theta)\big) \\ & -2\left(\frac{\partial L(\theta)}{\partial\theta_j}\right)^T \frac{\partial\Omega(\theta)}{\partial\theta_i}\big(L(\theta_0)-L(\theta)\big) \\ & +\big(L(\theta_0)-L(\theta)\big)^T \frac{\partial^2\Omega(\theta)}{\partial\theta_j\partial\theta_i}\big(L(\theta_0)-L(\theta)\big),\end{aligned}$$

for $i, j \in \{1, \ldots, p\}$. The map $\theta \mapsto \mathcal{H}(\theta; \theta_0)$ is continuous and

$$\mathcal{H}(\theta_0;\theta_0) = 2\,\dot{L}(\theta_0)^T\,\Omega(\theta_0)\,\dot{L}(\theta_0), \tag{3.7}$$

is a positive definite matrix. This $p \times p$ matrix is non-singular, since the $q \times q$ matrix $\Omega(\theta_0)$ is non-singular and the $q \times p$ matrix $\dot{L}(\theta_0)$ has rank p (recall $q \ge p$). Let $\|\cdot\|$ denote the spectral norm of a matrix. From Weyl's perturbation theorem (Jiang 2010, page 145), there exists an $\eta > 0$ such that every symmetric matrix $A \in \mathbb{R}^{p\times p}$ with $\|A - \mathcal{H}(\theta_0; \theta_0)\| \le \eta$ has positive eigenvalues and is therefore positive definite. Let $\varepsilon_1 \in (0, \varepsilon_0]$ be sufficiently small such that the second-order partial derivatives of L and Ω are continuous on $B_{\varepsilon_1}(\theta_0)$ and such that $\|\mathcal{H}(\theta; \theta_0) - \mathcal{H}(\theta_0; \theta_0)\| \le \eta/2$ for all $\theta \in B_{\varepsilon_1}(\theta_0)$.

Let $\mathcal{H}_{n,k,\Omega}(\theta) \in \mathbb{R}^{p\times p}$ denote the Hessian matrix of $f_{n,k}$. Its (i, j)-th element is

$$\begin{aligned}\big(\mathcal{H}_{n,k,\Omega}(\theta)\big)_{ij} &= \frac{\partial^2}{\partial\theta_j\partial\theta_i}\left[D_{n,k}(\theta)^T\,\Omega(\theta)\,D_{n,k}(\theta)\right] \\ &= \frac{\partial}{\partial\theta_j}\left[-2D_{n,k}(\theta)^T\,\Omega(\theta)\frac{\partial L(\theta)}{\partial\theta_i} + D_{n,k}(\theta)^T\frac{\partial\Omega(\theta)}{\partial\theta_i}D_{n,k}(\theta)\right] \\ &= 2\left(\frac{\partial L(\theta)}{\partial\theta_j}\right)^T \Omega(\theta)\left(\frac{\partial L(\theta)}{\partial\theta_i}\right) - 2\left(\frac{\partial^2 L(\theta)}{\partial\theta_j\partial\theta_i}\right)^T \Omega(\theta)\,D_{n,k}(\theta) \\ &\quad -2\left(\frac{\partial L(\theta)}{\partial\theta_i}\right)^T \frac{\partial\Omega(\theta)}{\partial\theta_j}D_{n,k}(\theta) - 2\left(\frac{\partial L(\theta)}{\partial\theta_j}\right)^T \frac{\partial\Omega(\theta)}{\partial\theta_i}D_{n,k}(\theta) \\ &\quad +D_{n,k}(\theta)^T\frac{\partial^2\Omega(\theta)}{\partial\theta_j\partial\theta_i}D_{n,k}(\theta).\end{aligned}$$

Since $D_{n,k}(\theta) = \widehat{L}_{n,k} - L(\theta)$ and since $\widehat{L}_{n,k}$ converges in probability to $L(\theta_0)$, we obtain

$$\sup_{\theta \in B_{\varepsilon_1}(\theta_0)} \|\mathcal{H}_{n,k,\Omega}(\theta) - \mathcal{H}(\theta;\theta_0)\| \xrightarrow{p} 0, \qquad n \to \infty. \tag{3.8}$$

By the triangle inequality, it follows that

$$\Pr\Bigg[\sup_{\theta \in B_{\varepsilon_1}(\theta_0)} \|\mathcal{H}_{n,k,\Omega}(\theta) - \mathcal{H}(\theta_0;\theta_0)\| \le \eta\Bigg] \to 1, \qquad n \to \infty. \tag{3.9}$$

In view of our choice for η, this implies that, with probability tending to one, $\mathcal{H}_{n,k,\Omega}(\theta)$ is positive definite for all $\theta \in B_{\varepsilon_1}(\theta_0)$, as required. □

Proof of Theorem 2 Let $\nabla f_{n,k}(\theta)$, a $1 \times q$ vector, be the gradient of $f_{n,k}$ at θ. By Eq. 2.12, we have

$$\begin{aligned}\sqrt{k}\,\nabla f_{n,k}(\theta_0) &= -2\sqrt{k}\,D_{n,k}(\theta_0)^T\,\Omega(\theta_0)\,\dot{L}(\theta_0) + \sqrt{k}D_{n,k}(\theta_0)^T \\ &\quad \times \big(\nabla\Omega(\theta)|_{\theta=\theta_0}\big)D_{n,k}(\theta_0) \\ &= -2\sqrt{k}\,D_{n,k}(\theta_0)^T\,\Omega(\theta_0)\,\dot{L}(\theta_0) + o_P(1), \quad \text{as } n \to \infty. \end{aligned} \tag{3.10}$$

Since $\widehat{\theta}_{n,k}$ is a minimizer of $f_{n,k}$, we have $\nabla f_{n,k}(\widehat{\theta}_{n,k}) = 0$. An application of the mean value theorem to the function $t \mapsto \nabla f_{n,k}\big(\theta_0 + t(\widehat{\theta}_{n,k} - \theta_0)\big)$ at $t = 0$ and $t = 1$ yields

$$0 = \nabla f_{n,k}(\widehat{\theta}_{n,k})^T = \nabla f_{n,k}(\theta_0)^T + \mathcal{H}_{n,k,\Omega}(\widetilde{\theta}_{n,k})\,(\widehat{\theta}_{n,k} - \theta_0), \tag{3.11}$$

where $\widetilde{\theta}_{n,k}$ is a random vector on the segment connecting θ_0 and $\widehat{\theta}_{n,k}$ and $\mathcal{H}_{n,k,\Omega}$ is the Hessian matrix of $f_{n,k}$ as in the proof of Theorem 1. Since $\widehat{\theta}_{n,k} \xrightarrow{p} \theta_0$, we have $\widetilde{\theta}_{n,k} \xrightarrow{p} \theta_0$ as $n \to \infty$ too. By Eqs. 3.8 and 3.7 and continuity of $\theta \mapsto \mathcal{H}(\theta;\theta_0)$, it then follows that

$$\mathcal{H}_{n,k,\Omega}(\widetilde{\theta}_{n,k}) \xrightarrow{p} \mathcal{H}(\theta_0;\theta_0) = 2\dot{L}(\theta_0)^T\,\Omega(\theta_0)\,\dot{L}(\theta_0), \qquad \text{as } n \to \infty. \tag{3.12}$$

Since $\mathcal{H}(\theta_0;\theta_0)$ is non-singular, the matrix $\mathcal{H}_{n,k,\Omega}(\widetilde{\theta}_{n,k})$ is non-singular with probability tending to one as well. Combine Eqs. 3.10, 3.11 and 3.12 to see that

$$\begin{aligned}\sqrt{n}\big(\widehat{\theta}_{n,k} - \theta_0\big) &= -\mathcal{H}_{n,k,\Omega}(\widetilde{\theta}_{n,k})^{-1}\,\sqrt{k}\,\nabla f_{n,k}(\theta_0)^T + o_p(1) \\ &= \big(\dot{L}(\theta_0)^T\Omega(\theta_0)\dot{L}(\theta_0)\big)^{-1}\,\dot{L}(\theta_0)^T\,\Omega(\theta_0)\,\sqrt{k}\,D_{n,k}(\theta_0) \\ &\quad + o_p(1), \qquad \text{as } n \to \infty.\end{aligned}$$

Convergence in distribution to the stated normal distribution follows from Eq. 2.12 and Slutsky's lemma. □

Proof of Corollary 2 Since $D_{n,k}(\theta) = \widehat{L}_{n,k} - L(\theta)$, we have

$$\sqrt{k}\,D_{n,k}(\widehat{\theta}_{n,k}) = \sqrt{k}\,D_{n,k}(\theta_0) - \sqrt{k}\big(L(\widehat{\theta}_{n,k}) - L(\theta_0)\big).$$

By Eq. 2.13 and the delta method, we have

$$\begin{aligned}\sqrt{k}\big(L(\widehat{\theta}_{n,k}) - L(\theta_0)\big) &= \dot{L}\,\sqrt{k}(\widehat{\theta}_{n,k} - \theta_0) + o_p(1) \\ &= \dot{L}\,(\dot{L}^T\Omega\dot{L})^{-1}\dot{L}^T\Omega\,\sqrt{k}\,D_{n,k}(\theta_0) + o_p(1) \\ &= P(\theta_0)\,\sqrt{k}\,D_{n,k}(\theta_0) + o_p(1), \qquad \text{as } n \to \infty,\end{aligned}$$

where $\dot{L}$ and Ω are evaluated at θ_0. Combination of the two previous displays yields

$$\sqrt{k}\, D_{n,k}(\widehat{\theta}_{n,k}) = (I_q - P(\theta_0))\,\sqrt{k}\, D_{n,k}(\theta_0) + o_p(1), \qquad \text{as } n \to \infty.$$

By Eq. 2.12 and Slutsky's lemma, we arrive at Eq. 2.16, as required.

The $q \times q$ matrix P has rank p since the $q \times p$ matrix $\dot{L}$ has rank p and the $q \times q$ matrix Ω is non-singular. Since $P^2 = P$, it follows that $\operatorname{rank}(I_q - P) = \operatorname{rank}(I_q) - \operatorname{rank}(P) = q - p$. □

Proof of Corollary 3 Eq. 2.12 can be written as

$$Z_{n,k} := \sqrt{k}\, D_{n,k}(\theta_0) \xrightarrow{d} Z \sim \mathcal{N}_q\big(0, \Sigma(\theta_0)\big), \qquad \text{as } n \to \infty.$$

In view of Eq. 2.16 and $\Omega(\theta) = \Sigma(\theta)^{-1}$, we find, by Slutsky's lemma and the continuous mapping theorem,

$$\begin{aligned} k\, f_{n,k}(\widehat{\theta}_{n,k}) &= k\, D_{n,k}(\widehat{\theta}_{n,k})^T\, \Sigma(\widehat{\theta}_{n,k})^{-1}\, D_{n,k}(\widehat{\theta}_{n,k}) \\ &= Z_{n,k}^T\, (I_q - P(\theta_0))^T\, \Sigma(\widehat{\theta}_{n,k})^{-1}\, (I_q - P(\theta_0))\, Z_{n,k} + o_p(1) \\ &\xrightarrow{d} Z^T\, (I_q - P(\theta_0))^T\, \Sigma(\theta_0)^{-1}\, (I_q - P(\theta_0))\, Z, \qquad \text{as } n \to \infty; \end{aligned}$$

here $P = \dot{L}\, (\dot{L}^T \Sigma^{-1} \dot{L})^{-1}\, \dot{L}^T \Sigma^{-1}$, with $\dot{L}$ and Σ evaluated at θ_0.

It remains to identify the distribution of the limit random variable. The random vector Z is equal in distribution to $\Sigma^{1/2} Y$, where $Y \sim \mathcal{N}_q(0, I_q)$ and where $\Sigma^{1/2}$ is a symmetric square root of Σ. Straightforward calculation yields

$$Z^T (I_q - P)^T\, \Sigma^{-1}\, (I_q - P)\, Z \stackrel{\mathrm{d}}{=} Y^T (I_q - B) Y$$

where $B = \Sigma^{-1/2} \dot{L}\, (\dot{L}^T \Sigma^{-1} \dot{L})^{-1}\, \dot{L}^T \Sigma^{-1/2}$. It is easily checked that B is a projection matrix ($B = B^T = B^2$). Moreover, B has rank p. It follows that $I_q - B$ is a projection matrix too and that it has rank $q - p$. The distribution of the limit random variable now follows by standard properties of quadratic forms of normal random vectors. □

Proof of Corollary 4 Let $Z \sim \mathcal{N}_q(0, \Sigma(\theta_0))$, which by Eq. 2.12 is the limit in distribution of $\sqrt{k}\, D_{n,k}(\theta_0)$. By Eq. 2.16 and the continuous mapping theorem, we have, as $n \to \infty$,

$$k\, D_{n,k}(\widehat{\theta}_{n,k})^T\, A(\widehat{\theta}_{n,k})\, D_{n,k}(\widehat{\theta}_{n,k}) \xrightarrow{d} Z^T\, (I_q - P(\theta_0))^T\, A(\theta_0)\, (I_q - P(\theta_0))\, Z. \quad (3.13)$$

We can represent $(I_q - P)Z$ as $V D^{1/2} Y$, with $Y \sim \mathcal{N}_q(0, I_q)$. The limiting random variable in Eq. 3.13 is then given by

$$Y^T D^{1/2} V^T\, V_s D_s^{-1} V_s^T V D^{1/2} Y.$$

Since V is an orthogonal matrix, this expression simplifies to $\sum_{j=1}^{s} Y_j^2$, which has the stated χ_s^2 distribution. □

Proof of Remark 1 Inspection of the proofs of Corollaries 3 and 4 shows that the difference between the two test statistics converges in distribution to the random variable $Z^T R(\theta_0) Z$, where Z is a certain q-variate normal random vector and where

$$R(\theta_0) = \left(I_q - P(\theta_0)\right)^T \left(\Sigma(\theta_0)^{-1} - A(\theta_0)\right) \left(I_q - P(\theta_0)\right).$$

The matrix $R(\theta_0)$ can be shown to be equal to zero, proving the claim of the remark. To see why $R(\theta_0)$ is zero, note first that, suppressing θ_0 and writing $Q = I_q - P$, we have $Q^2 = Q$ and $\Sigma Q^T = Q\Sigma = Q\Sigma Q^T$. Recall the eigenvalue equation $Q\Sigma Q^T v_j = \nu_j v_j$ for $j = 1, \dots, q$. Note that $\nu_j > 0$ if $j \le s$ and $\nu_j = 0$ if $j \ge s+1$. The eigenvalue equation implies that $Qv_j = v_j$ for $j \le s$ while $Q\Sigma v_j = 0$ for $j \ge s + 1$. Since the vectors $v_1, \dots, v_q$ are orthogonal, we find that the vectors $v_1, \dots, v_s, \Sigma v_{s+1}, \dots, \Sigma v_q$ are linearly independent. It then suffices to show that $Rv_j = 0$ for all $j \le s$ and $R\Sigma v_j = 0$ for all $j \ge s + 1$. The first property follows from the fact that $\Sigma^{-1} v_j = \nu_j^{-1} Q^T v_j$ and $Av_j = \nu_j^{-1} v_j$ for $j \le s$ (use the eigenvalue equation again), while the second property follows from $Q\Sigma v_j = 0$ for $j \ge s + 1$. □

Appendix B: Checking the conditions of the main theorems

In the main theorems, the following conditions are imposed:

1. The map $L : \theta \mapsto (\ell(c_m; \theta))_{m=1}^q$ is a homeomorphism from Θ to $L(\Theta)$.
2. L is twice continuously differentiable on a neighbourhood of θ_0 and $\dot{L}(\theta_0)$ is of full rank (i.e., of rank p).
3. $\Omega : \Theta \to \mathbb{R}^{q\times q}$ is twice continuously differentiable on a neighbourhood of θ_0.

We verify these conditions on the parametric models in the simulation study.

Logistic model

For convenience, we exclude $\theta = 0$ and $\theta = 1$. For $J \subset \{1, \dots, d\}$ with $|J| \ge 2$, we have $\ell(e_J; \theta) = |J|^\theta$. Assume we include e_J for all $J \subset \{1, \dots, d\}$ with $|J| = a$ for some fixed $a \in \{2, \dots, d\}$. Then $L^T(\theta) = (a^\theta, \dots, a^\theta)$.

1. Continuity of L is clear. L is also one-to-one since $\theta = \log(a^\theta)/\log a$, and the map $z \mapsto \log z/\log a$ is continuous, so that L is a homeomorphism.
2. Clearly, L is twice continuously differentiable. Further, $\dot{L}^T(\theta) = (a^\theta \log a, \dots, a^\theta \log a)$ and $a^\theta \log a > 0$, so that $\dot{L}(\theta)$ has rank 1.
3. As the model is exchangeable, setting Ω equal to the identity matrix is already optimal, so there is nothing to prove.

Brown–Resnick model

Given d known points $s_1, \dots, s_d \in \mathbb{R}^2$, consider the stable tail dependence function $\ell(\,\cdot\,; \theta)$ of the Brown–Resnick model with parameter $\theta = (\alpha, \rho) \in (0, 2] \times (0, \infty)$. For pairs $j, k \in \{1, \dots, d\}$, $j \ne k$, we have

$$\ell(e_{\{j,k\}}; \theta) = 2\,\Phi(a_{jk}/2) \quad \text{for} a_{jk} = \{2\gamma(s_j - s_k)\}^{1/2} = 2^{1/2}\|s_j - s_k\|^{\alpha/2}/\rho^{\alpha/2}, \tag{3.14}$$

where Φ denotes the standard normal distribution function.

We consider points s_j on a finite, rectangular, unit-distance grid and we set $c_m = e_{J_m}$ for all possible pairs $J_m \subset \{1, \ldots, d\}$ such that the locations s_{j_1} and s_{j_2} with $J_m = \{j_1, j_2\}$ are neighbours (horizontally, vertically, or diagonally). The number, q, of such pairs depends on the grid size; however, the pairwise extremal coefficient (3.14) only depends on the distance between the two locations s_j and s_k. For neighbouring locations on a unit-size rectangular grid, there are thus only two distinct values to consider, $\|s_j - s_k\| \in \{1, \sqrt{2}\}$.

For L, we can thus reduce the analysis to the one of the function

$$L(\theta) = \left(2\,\Phi(2^{-1/2}\rho^{-\alpha/2}), 2\,\Phi(2^{-1/2+\alpha/4}\rho^{-\alpha/2})\right).$$

The partial derivatives are

$$\begin{aligned}
\frac{\partial L_1(\theta)}{\partial\alpha} &= -2^{-1/2}\varphi(2^{-1/2}\rho^{-\alpha/2})\,\rho^{-\alpha/2}\log(\rho),\\
\frac{\partial L_1(\theta)}{\partial\rho} &= -2^{-1/2}\alpha\,\varphi(2^{-1/2}\rho^{-\alpha/2})\,\rho^{-\alpha/2-1},\\
\frac{\partial L_2(\theta)}{\partial\alpha} &= \varphi(2^{-1/2+\alpha/4}\rho^{-\alpha/2})\,2^{-1/2+\alpha/4}\rho^{-\alpha/2}\{2^{-1}\log(2) - \log(\rho)\},\\
\frac{\partial L_2(\theta)}{\partial\rho} &= -2^{-1/2+\alpha/4}\alpha\,\varphi(2^{-1/2+\alpha/4}\rho^{-\alpha/2})\,\rho^{-\alpha/2-1},
\end{aligned}$$

with φ the standard normal density function. The determinant of the Jacobian matrix is

$$\det\dot{L}(\theta) = \alpha(\log 2)\rho^{-\alpha-1}2^{-2+\alpha/4}\varphi(2^{-1/2}\rho^{-\alpha/2})\,\varphi(2^{-1/2+\alpha/4}\rho^{-\alpha/2}),$$

which is always positive. By the inverse function theorem, the function L is a diffeomorphism in the neighbourhood of θ_0. Whether L is also continuously invertible globally is a more difficult question. The two components of L are decreasing in ρ, but the dependence on α is not monotone and depends on ρ.

For the empirical stable tail dependence function, the entries of $\Sigma(\theta)$ are clearly smooth as a function of θ. We do not have an analytical proof that $\Sigma(\theta)$ is invertible, but numerical experiments indicate it is. In that case, the matrix $\Omega(\theta) = \Sigma(\theta)^{-1}$ satisfies the conditions. Otherwise, we can define $\Omega(\theta) = \{\Sigma(\theta) + \lambda I_q\}^{-1}$ for some constant $\lambda > 0$; see also the paragraph on max-linear graphical models.

Max-linear graphical model

Consider the four-dimensional max-linear model induced by the diamond-shaped directed acyclic graph in the simulation study. Its coefficient matrix $B = B(\theta) \in \mathbb{R}^{4\times 4}$ is defined by

$$B = (b_{jt})_{j,t=1}^4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \theta_1 & 1-\theta_1 & 0 & 0 \\ \theta_2 & 0 & 1-\theta_2 & 0 \\ \theta_1\theta_3 \vee \theta_2\theta_4 & (1-\theta_1)\theta_3 & (1-\theta_2)\theta_4 & b_{44} \end{pmatrix},$$

with $b_{44} = b_{44}(\theta)$ such that the sum of the fourth row is one. The stable tail dependence function is

$$\begin{aligned}\ell(x;\theta) &= x_1 \vee \theta_1 x_2 \vee \theta_2 x_3 \vee (\theta_1\theta_3 \vee \theta_2\theta_4)x_4 \\ &\quad +(1-\theta_1)x_2 \vee (1-\theta_1)\theta_3 x_4 \\ &\quad +(1-\theta_2)x_3 \vee (1-\theta_2)\theta_4 x_4 \\ &\quad +b_{44}x_4. \end{aligned} \tag{3.15}$$

The parameter set is $\Theta = (0,1)^4$, although later on, we will have to limit the parameter set to $\Theta = \{\theta \in (0,1)^4 : \theta_1\theta_3 \neq \theta_2\theta_4\}$.

We consider points $x = c_m = e_{J_m}$ for J_m equal to one of the following $q = 5$ subsets of $\{1,\ldots,4\}$:

$$J_1 = \{1,2\}, \quad J_2 = \{1,3\}, \quad J_3 = \{1,2,4\}, \quad J_4 = \{1,3,4\}, \quad J_5 = \{1,2,3,4\}. \tag{3.16}$$

Clearly, the map $\theta \mapsto L(\theta) = (\ell(c_m;\theta))_{m=1}^q$ is continuous. Next we show that it is one-to-one and that the inverse map from $L(\Theta)$ to Θ is continuous too. We have

$$\ell(e_{\{1,2\}};\theta) = 1 + (1-\theta_1), \quad \ell(e_{\{1,3\}};\theta) = 1 + (1-\theta_2),$$

from which we can identify θ_1 and θ_2. Next,

$$\ell(e_{\{1,2,3,4\}};\theta) = 1 + (1-\theta_1) + (1-\theta_2) + b_{44},$$

from which we can recover b_{44}. Finally,

$$\begin{aligned}\ell(e_{\{1,2,4\}};\theta) &= 1 + (1-\theta_1) + (1-\theta_2)\theta_4 + b_{44}, \\ \ell(e_{\{1,3,4\}};\theta) &= 1 + (1-\theta_1)\theta_3 + (1-\theta_2) + b_{44},\end{aligned}$$

from which we can solve θ_4 and θ_3, respectively. We find that $L : (0,1)^4 \to L((0,1)^4)$ is one-to-one and that its inverse function is continuous.

If θ_0 is such that $\theta_1\theta_3 \neq \theta_2\theta_4$, then $b_{44} = b_{44}(\theta)$ is a polynomial function of θ for θ in a neighbourhood of θ_0. Writing $\dot{b}_{44,m} = \partial b_{44}/\partial\theta_m$, the Jacobian of L is

$$\begin{aligned}\dot{L}(\theta) &= \left(\frac{\partial \ell(c_m;\theta)}{\partial\theta_r}\right)_{\substack{m=1,\ldots,5\\ r=1,\ldots,4}} \\ &= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ -1+\dot{b}_{44,1} & -\theta_4+\dot{b}_{44,2} & \dot{b}_{44,3} & (1-\theta_2)+\dot{b}_{44,4} \\ -\theta_3+\dot{b}_{44,1} & -1+\dot{b}_{44,2} & (1-\theta_1)+\dot{b}_{44,3} & \dot{b}_{44,4} \\ -1+\dot{b}_{44,1} & -1++\dot{b}_{44,2} & \dot{b}_{44,3} & \dot{b}_{44,4} \end{pmatrix}.\end{aligned}$$

The determinant of the upper 4×4 submatrix is

$$\begin{aligned}\det \dot{L}_{1:4,1:4}(\theta) &= \dot{b}_{44,3}\dot{b}_{44,4} - (1-\theta_1+\dot{b}_{44,3})(1-\theta_2+\dot{b}_{44,4}) \\ &= -(1-\theta_1)(1-\theta_2) - (1-\theta_1)\dot{b}_{44,4} - (1-\theta_2)\dot{b}_{44,3}.\end{aligned}$$

We need to show that this is nonzero. There are two cases: $\theta_1\theta_3 > \theta_2\theta_4$ or $\theta_1\theta_3 < \theta_2\theta_4$. In the first case,

$$b_{44} = 1 - \theta_1\theta_3 - (1-\theta_1)\theta_3 - (1-\theta_2)\theta_4 = 1 - \theta_3 - (1-\theta_2)\theta_4$$

whence $\dot{b}_{44,3} = -1$ and $\dot{b}_{44,4} = -(1-\theta_2)$, and thus

$$\begin{aligned}\det \dot{L}_{1:4,1:4}(\theta) &= -(1-\theta_1)(1-\theta_2) - (1-\theta_1)\{-(1-\theta_2)\} - (1-\theta_2)(-1)\\ &= 1-\theta_2 > 0,\end{aligned}$$

as required. The case $\theta_1\theta_3 < \theta_2\theta_4$ is similar. We find that on $\{\theta \in (0,1)^4 : \theta_1\theta_3 \neq \theta_2\theta_4\}$, the map L is twice continuously differentiable and its Jacobian is of full rank.

For the weight matrix, we set $\Omega(\theta) = \{\Sigma(\theta) + \lambda I_q\}^{-1}$, where $\Sigma(\theta)$ is the asymptotic covariance matrix of the empirical stable tail dependence function evaluated at the points c_m and where $\lambda > 0$ is a constant. The entries of the matrix $\Sigma(\theta)$ are a polynomial expression in $\ell(x;\theta)$ and its partial derivatives with respect to $x_1, \ldots, x_d$ for a finite set of points x. The matrix $\Sigma(\theta)$ is symmetric positive semidefinite so that the matrix $\Sigma(\theta) + \lambda I_q$ is symmetric and positive definite with all eigenvalues larger than or equal to λ. As a consequence, $\Sigma(\theta) + \lambda I_q$ is invertible for all θ. Matrix inversion being a smooth operation, smoothness properties of $\Omega(\theta)$ follow from those of $\Sigma(\theta) + \lambda I_q$ and thus of $\Sigma(\theta)$. In view of the maxima in the expression for $\ell(x;\theta)$ in Eq. 3.15, choosing x of the form e_J requires excluding certain values of θ from the parameter set. Modulo this restriction, we find that $\Omega(\theta)$ is twice continuously differentiable on a neighbourhood of θ_0.

In the simulation study, we considered many more points c_m than the five points $e_{J_1}, \ldots, e_{J_5}$ with J_m as in Eq. 3.16. The stated properties of L and Ω remain true, with the exception that even more parameter values θ need to be excluded to ensure smoothness of L and Ω in a neighbourhood of θ_0. Since the set of points thus excluded is a Lebesgue null set, this restriction should (hopefully) pose no problems in practice.

Max-linear model: a counterexample

Consider the three-dimensional factor model

$$\begin{cases} X_1 = Z_1, \\ X_2 = \alpha Z_1 \vee (1-\alpha)Z_2, \\ X_3 = \beta Z_1 \vee (1-\beta)\theta Z_2 \vee (1-\beta)(1-\theta)Z_3, \end{cases}$$

where $\alpha, \beta \in (0,1)$ are known and where $\theta \in (0,1)$ is the only parameter. The stable tail dependence function is given by

$$\ell(x;\theta) = x_1 \vee \alpha x_2 \vee \beta x_3 + (1-\alpha)x_2 \vee (1-\beta)\theta x_3 + (1-\beta)(1-\theta)x_3.$$

If $(1-\beta)\theta x_3 \geq (1-\alpha)x_2$, then the expression simplifies to

$$\ell(x;\theta) = x_1 \vee \alpha x_2 \vee \beta x_3 + (1-\beta)x_3,$$

and does no longer depend on θ. If $x_3 = 0$, then $\ell(x;\theta)$ does not depend on θ either. To identify θ from $\ell(x;\theta)$, we therefore need to choose x in such a way that $x_3 > 0$ and $(1-\beta)\theta x_3 < (1-\alpha)x_2$. Since $\theta \in (0,1)$, it is sufficient to choose x such that $x_3/x_2 < (1-\alpha)/(1-\beta)$.

However, if α is a parameter too, then the bound $(1-\alpha)/(1-\beta)$ approaches 0 as α approaches 1. Hence, in the two-parameter model where α and θ are unknown, we cannot identify all $(\alpha, \theta) \in (0, 1)^2$ from the values of ℓ in a pre-specified, finite set of q points x.

References

Abadir, K.M., Magnus, J.R.: Matrix Algebra, vol. 1. Cambridge University Press, Cambridge (2005)

Beirlant, J., Goegebeur, Y., Segers, J.: Statistics of Extremes: Theory and Applications. Wiley, New Jersey (2004)

Beirlant, J., Escobar-Bach, M., Goegebeur, Y., Guillou, A.: Bias-corrected estimation of the stable tail dependence function. J. Multivar. Anal. **143**(1), 453–466 (2016)

Bücher, A., Segers, J.: Extreme value copula estimation based on block maxima of a multivariate stationary time series. Extremes **17**(3), 495–528 (2014)

Can, S.U., Einmahl, J.H.J., Khmaladze, E.V., Laeven, R.J.A., et al.: Asymptotically distribution-free goodness-of-fit testing for tail copulas. Ann. Stat. **43**(2), 878–902 (2015)

Coles, S.G., Tawn, J.A.: Modelling extreme multivariate events. J. Royal Stat. Soc. Ser. B (Stat. Methodol.) **53**(2), 377–392 (1991)

Davison, A.C., Padoan, S.A., Ribatet, M.: Statistical modeling of spatial extremes. Stat. Sci. **27**(2), 161–186 (2012)

de Haan, L.: A spectral representation for max-stable processes. Ann. Probab. **12**(4), 1194–1204 (1984)

de Haan, L., Ferreira, A.: Extreme Value Theory: An Introduction. Springer-Verlag Inc, German (2006)

de Haan, L., Pereira, T.T.: Spatial extremes: Models for the stationary case. Ann. Stat. **34**(1), 146–168 (2006)

Drees, H., Huang, X.: Best attainable rates of convergence for estimators of the stable tail dependence function. J. Multivar. Anal. **64**(1), 25–47 (1998)

Einmahl, J.H.J., Krajina, A., Segers, J.: An M-estimator for tail dependence in arbitrary dimensions. Ann. Stat. **40**(3), 1764–1793 (2012)

Einmahl, J.H.J., Kiriliouk, A., Krajina, A., Segers, J.: An M-estimator of spatial tail dependence. J. Royal Stat. Soc. Ser. B (Stat. Methodol.) **78**(1), 275–298 (2016)

Fougères, A.L., De Haan, L., Mercadier, C.: Bias correction in multivariate extremes. Ann. Stat. **43**(2), 903–934 (2015)

Fougères, A.L., Mercadier, C., Nolan, J.: Estimating Semi-Parametric Models for Multivariate Extreme Value Data. Working paper (2016)

Gissibl, N., Klüppelberg, C.: Max-Linear Models on Directed Acyclic Graphs. To be published in Bernoulli, available at arXiv:1512.07522 (2017)

Gumbel, E.J.: Bivariate exponential distributions. J. Amer. Stat. Assoc. **55**(292), 698–707 (1960)

Hansen, L.P., Heaton, J., Yaron, A.: Finite-sample properties of some alternative GMM estimators. J. Bus. Econ. Stat. **14**(3), 262–280 (1996)

Hofert, M., Kojadinovic, I., Maechler, M., Yan, J.: Copula: multivariate dependence with copulas. R package version 0.999-13 (2015)

Huang, X.: Statistics of Bivariate Extreme Values. PhD thesis, Tinbergen Institute Research Series, Amsterdam (1992)

Huser, R., Davison, A.: Composite likelihood estimation for the Brown-Resnick process. Biometrika **100**(2), 511–518 (2013)

Huser, R., Davison, A.: Space–time modelling of extreme events. J. Royal Stat. Soc. Ser. B (Stat. Methodol.) **76**(2), 439–461 (2014)

Huser, R., Davison, A.C., Genton, M.G.: Likelihood estimators for multivariate extremes. Extremes **19**(1), 79–103 (2016)

Jiang, J.: Large Sample Techniques for Statistics. Springer (2010)

Kabluchko, Z., Schlather, M., De Haan, L.: Stationary max-stable fields associated to negative definite functions. Ann. Probab. **37**(5), 2042–2065 (2009)

Kiriliouk, A.: tailDepFun: Minimum Distance Estimation of Tail Dependence Models. R package version 1.0.0. (2016)

Molchanov, I.: Convex geometry of max-stable distributions. Extremes **11**(3), 235–259 (2008)

Oesting, M., Schlather, M., Friedrichs, P.: Statistical post-processing of forecasts for extremes using bivariate Brown-Resnick processes with an application to wind gusts. Extremes **20**(2), 309–332 (2015)

R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (2015)

Ressel, P.: Homogeneous distributions–and a spectral representation of classical mean values and stable tail dependence functions. J. Multivar. Anal. **117**, 246–256 (2013)

Ribatet, M.: SpatialExtremes: Modelling Spatial Extremes. http://CRAN.R-project.org/package=SpatialExtremes, r package version 2.0–2 (2015)

Schlather, M., Tawn, J.: A dependence measure for multivariate and spatial extreme values: Properties and inference. Biometrika **90**(1), 139–156 (2003)

Smith, R.L.: Max-stable processes and spatial extremes, unpublished manuscript (1990)

Wadsworth, J.L., Tawn, J.A.: Efficient inference for spatial extreme-value processes associated to log-gaussian random functions. Biometrika **101**(1), 1–15 (2014)

Wang, Y., Stoev, S.A.: Conditional sampling for spectrally discrete max-stable random fields. Adv. Appl. Prob. **43**(2), 461–483 (2011)