UCL Université catholique de Louvain

INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

ISBA

DISCUSSION PAPER

2017/27

Hypothesis testing for tail dependence parameters on the boundary of the parameter space with application to generalized max-linear models

A. Kiriliouk

Hypothesis testing for tail dependence parameters on the boundary of the parameter space with application to generalized max-linear models

Anna Kiriliouk
Université catholique de Louvain
Institut de Statistique, Biostatistique et Sciences Actuarielles
Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.
E-mail: anna.kiriliouk@uclouvain.be

Abstract

Modelling multivariate tail dependence is one of the key challenges in extreme-value theory. The max-linear model is a parametric tail dependence model which is dense in the class of multivariate extreme-value models. Being non-differentiable, it cannot be estimated using likelihood-based methods, so that minimum distance estimation forms a valuable alternative. Currently, estimation is limited to the set-up where the number of factors and/or the structure of the model is defined a priori, because answering these questions necessitates estimation and testing at the boundary of the parameter space. The main goal of this paper is to propose hypothesis tests for tail dependence parameters that, under the null hypothesis, are on the boundary of the alternative hypothesis. We give the asymptotic distribution of the weighted least squares estimator proposed in Einmahl, Kiriliouk and Segers (2017, to be published in *Extremes*) when the true parameter is on the boundary of the parameter space, and we propose two test statistics whose asymptotic distribution is easily computable. An extensive simulation study evaluates the performance of the test statistics, which are then applied to the stock market prices of two NYSE companies.

Keywords: hypothesis testing; max-linear model; multivariate extremes; stable tail dependence function; tail dependence.

1 Introduction

Extreme-value theory is the branch of statistics concerned with the characterization of extreme events. These occur in a large variety of fields, such as hydrology, meteorology, finance and insurance, but also in settings like air pollution or athletic records. The focus can be on either block maxima (for instance, the yearly maximal log-returns on a financial stock) or on threshold exceedances (eg, all coastal water heights exceeding some high level). In the univariate case, the limiting distribution of suitably normalized maxima and threshold exceedances can be characterized entirely using the generalized extreme-value (Fisher and Tippett, 1928; Gnedenko, 1943) and the generalized Pareto distribution (Balkema and De Haan, 1974; Pickands III, 1975) respectively. However, many extreme events are inherently multivariate, and an important challenge is to model the tail dependence between two or more random variables of interest. This can be done conveniently using the stable tail dependence function (Huang, 1992; Drees and Huang, 1998), or, equivalently, the exponent

measure function (de Haan and Resnick, 1977; Coles and Tawn, 1991). Contrary to the univariate setting, the family of limiting distributions for multivariate maxima or threshold exceedances is infinite-dimensional.

Multivariate extremes are usually characterized using parametric models, because the curse of dimensionality makes nonparametric inference unwieldly in high dimensions. These parametric tail dependence models, Gumbel (1960); Smith (1990); Tawn (1990); Coles and Tawn (1991); Kabluchko et al. (2009) among others, can be estimated using likelihood-based methods (Davison et al., 2012; Huser and Genton, 2016) or minimum distance estimation (Einmahl et al., 2012, 2016, 2017).

Max-linear models form a conceptually simple but useful class of models, since they are dense in the class of multivariate extreme-value models. A max-linear model is constructed using a set of independent factors and a parameter vector representing the impact of each factor. Every component can be interpreted as the maximum shock among a set of factors. Recent work on max-linear models is numerous. In Wang and Stoev (2011), an algorithm is prposed for conditional sampling from max-linear models with the aim of doing prediction. In Gissibl and Klüppelberg (2015) and Gissibl et al. (2017), a recursive max-linear model defined on a directed acyclic graph is considered. Models handling asymptotic independence, based on bivariate max-linear structures, are proposed in Kereszturi and Tawn (2017). An application where max-linear models are used to model a latent factor structure for financial returns can be found in Cui and Zhang (2016). In Falk et al. (2015), a generalized max-linear process is proposed whose factor variables do not need to be independent, meant to do (spatial) prediction by interpolation. Finally, the Marshall–Olkin copula (Embrechts et al., 2003; Segers, 2012) and the Tawn–Molchanov model (Yuen and Stoev, 2014; Strokorb et al., 2015) are both special cases of the max-linear model.

Because of their non-differentiability, max-linear models cannot be estimated by standard likelihood methods. Estimation has been considered in Einmahl et al. (2012), Yuen and Stoev (2014) and Einmahl et al. (2017). Currently, it is limited to the set-up where one decides on the number of factors a priori. However, it is of interest to be able to test how many factors are necessary to model a dataset, or to test if a more specific model (eg, Marshall–Olkin) is sufficient. Although a hypothesis testing framework is described in Einmahl et al. (2016), these results are only valid in the interior of the parameter space, whereas we are interested in testing whether one or more factor loadings are equal to zero.

The main goal of this paper is to propose hypothesis tests for tail dependence parameters that, under the null hypothesis, are on the boundary of the alternative hypothesis. After having introduced the basics of extreme-value theory and the weighted least squares estimator of Einmahl et al. (2017) in Section 2, Section 3 establishes asymptotic normality for the weighted least squares estimator when the true parameter vector lies on the boundary of the parameter space. Next, we describe in which cases the asymptotic distribution can be simplified and we propose a deviance- and a Wald-type test statistic whose asymptotic distributions are easily computable when the dimension of the parameter on the boundary is not too large. These results are obtained as a special case of more general results in Andrews (1999), Andrews (2001) and Andrews (2002), where much higher-level assumptions are made. In Section 4, we introduce the generalized max-linear model that was proposed in Falk et al. (2015) and perform extensive simulations to evaluate the performance of the test statistics. Finally, Section 5 illustrates our methods on the stock market prices of two NYSE companies and Section 6 discusses the limitations and prospects of the testing procedure. Technical definitions and proofs are deferred to the appendix.

2 Background

2.1 Multivariate extreme-value theory

Let $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})$, $i \in \{1, \ldots, n\}$, be random vectors in $\mathbb{R}^d$ with a common cumulative distribution function F and marginal cumulative distribution functions $F_1, \ldots, F_d$. Let $\boldsymbol{M}_n := (M_{n,1}, \ldots, M_{nd})$ with $M_{nj} := \max(X_{1j}, \ldots, X_{nj})$ for $j = 1, \ldots, d$. We say that F is in the *max-domain of attraction* of an *extreme-value distribution* G if there exist sequences of normalizing constants $\boldsymbol{a}_n = (a_{n1}, \ldots, a_{nd}) > 0$ and $\boldsymbol{b}_n = (b_{n1}, \ldots, b_{nd}) \in \mathbb{R}^d$ such that

$$\mathbb{P}\bigg[\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \le \boldsymbol{x}\bigg] = F^n(\boldsymbol{a}_n \boldsymbol{x} + \boldsymbol{b}_n) \xrightarrow{\mathrm{d}} G(\boldsymbol{x}), \qquad \text{as } n \to \infty. \tag{2.1}$$

The margins, $G_1, \ldots, G_d$, of G are univariate extreme-value distributions and the function G is determined by

$$G(\boldsymbol{x}) = \exp\{-\ell(-\log G_1(x_1), \ldots, -\log G_d(x_d))\},$$

where $\ell : [0, \infty)^d \to [0, \infty)$ is called the *stable tail dependence function*,

$$\ell(\boldsymbol{x}) := \lim_{t \downarrow 0} t^{-1}\, \mathbb{P}[1 - F_1(X_{11}) \le t x_1 \text{ or } \ldots \text{ or } 1 - F_d(X_{1d}) \le t x_d]. \tag{2.2}$$

Existence of the limit in (2.2) 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 G.

The function ℓ is convex, homogeneous of order one, and satisfies $\max(x_1, \ldots, x_d) \le \ell(\boldsymbol{x}) \le x_1 + \cdots + x_d$ for all $x \in [0, \infty)^d$. If $d = 2$, these properties characterize the class of all d-variate stable tail dependence functions, but not if $d \ge 3$ (Molchanov, 2008; Ressel, 2013). Because the class of stable tail dependence functions is infinite-dimensional, one usually considers parametric models for ℓ. The following model construction tool is proposed in Falk et al. (2010, Chapter 4) and Segers (2012), among others.

Let R be a unit Fréchet random variable, i.e., $\mathbb{P}[R \le r] = \exp(-1/r)$ for $r > 0$ and let $\boldsymbol{V} = (V_1, \ldots, V_d) \in [0, \infty)^d$ be a vector with $0 < \mathbb{E}[V_j] < \infty$ for $j \in \{1, \ldots, d\}$. If $\boldsymbol{X} = (X_1, \ldots, X_d) = R\boldsymbol{V}$, then $\boldsymbol{X}$ is in the max-domain of attraction of

$$G(\boldsymbol{x}) = \exp\{-\mathbb{E}\left[\max(V_1/x_1, \ldots, V_d/x_d)\right]\},$$

and we say that $\boldsymbol{X}$ is generated by $\boldsymbol{V}$. Here, G has unit Fréchet margins if $\mathbb{E}[V_j] = 1$ for $j \in \{1, \ldots, d\}$. The corresponding stable tail dependence function is

$$\ell_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{E}\left[\max_{j=1,\ldots,d} \frac{x_j V_j}{\mathbb{E}[V_j]}\right], \qquad \boldsymbol{x} \in [0, \infty)^d. \tag{2.3}$$

We will often drop the subscript of ℓ when it is clear which variable is of interest. For more background on multivariate extreme value theory, see Beirlant et al. (2004), de Haan and Ferreira (2006) or Falk et al. (2010).

2.2 Nonparametric estimation of the stable tail dependence function

Let $k = k_n \in (0, n]$ be an intermediate sequence, that is,

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

A straightforward estimator of ℓ is obtained by replacing $\mathbb{P}$ and $F_1, \ldots, F_d$ in (2.2) by the (modified) empirical distribution functions and replacing t by k/n, yielding

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

This estimator, the *empirical stable tail dependence function*, was introduced in slightly different form for $d = 2$ in Huang (1992) and studied further in Drees and Huang (1998). Using $n + 1/2$ rather than n allows for better finite-sample properties (Einmahl et al., 2012).

Another nonparametric estimator is the *beta stable tail dependence function*, which was very recently proposed in Kiriliouk et al. (2017). Contrary to the empirical stable tail dependence function, the beta tail dependence function is a smooth estimator which leads to some improvement in its finite-sample behavior. Its name stems from the fact that it is based on the empirical beta copula (Segers et al., 2017). We define

$$\widetilde{\ell}_{n,k}^{\beta}(\boldsymbol{x}) = \frac{n}{k}\left\{1 - \mathbb{C}_n^{\beta}\left(1 - \frac{k\boldsymbol{x}}{n}\right)\right\}, \quad \text{where} \quad \mathbb{C}_n^{\beta}(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} F_{n,R_{ij}^n}(u_j),$$

and for $r \in \{1, \ldots, n\}$, $F_{n,r}(u) = \sum_{s=1}^{r} \binom{n}{s} u^s (1-u)^{n-s}$ is the cumulative distribution function of a $\mathrm{Beta}(r, n + 1 - r)$ random variable.

A drawback of $\widetilde{\ell}_{n,k}$ or $\widetilde{\ell}_{n,k}^{\beta}$ might be their possibly growing bias as k increases. Recently, two bias-corrected estimators have been proposed (Beirlant et al., 2016; Fougères et al., 2015). One of these has already been used in Einmahl et al. (2017) as the initial estimator in the weighted least squares procedure.

2.3 Weighted least squares estimator

Henceforth we assume that ℓ belongs to a parametric family $\{\ell(\cdot\,;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ with $\Theta \subset \mathbb{R}^p$. A semi-parametric estimation method for $\boldsymbol{\theta}_0$, the true parameter vector, has been proposed in Einmahl et al. (2017).

Let $\widehat{\ell}_{n,k}$ denote an initial estimator of ℓ based on $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$; an obvious candidate is one of the nonparametric estimators introduced in Section 2.2. Let $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q \in [0, \infty)^d$, with $\boldsymbol{c}_m = (c_{m1}, \ldots, c_{md})$ for $m = 1, \ldots, q$, be q points in which we will evaluate ℓ and $\widehat{\ell}_{n,k}$. Consider for $\boldsymbol{\theta} \in \Theta$ the $q \times 1$ column vectors

$$\widehat{\boldsymbol{L}}_{n,k} := \left(\widehat{\ell}_{n,k}(\boldsymbol{c}_m)\right)_{m=1}^{q}, \qquad L(\boldsymbol{\theta}) := \left(\ell(\boldsymbol{c}_m;\boldsymbol{\theta})\right)_{m=1}^{q}, \qquad D_{n,k}(\boldsymbol{\theta}) := \widehat{\boldsymbol{L}}_{n,k} - L(\boldsymbol{\theta}),$$

The points $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q$ need to be chosen in such a way that the map $L : \Theta \to \mathbb{R}^q$ is one-to-one, i.e., $\boldsymbol{\theta}$ is identifiable from $\ell(\boldsymbol{c}_1;\boldsymbol{\theta}), \ldots, \ell(\boldsymbol{c}_q;\boldsymbol{\theta})$. In particular, we will need to assume that $q \geq p$, where p is the dimension of the parameter space Θ.

For $\boldsymbol{\theta} \in \Theta$, let $\Omega(\boldsymbol{\theta})$ be a symmetric, positive definite $q \times q$ matrix and define

$$f_{n,k}(\boldsymbol{\theta}) := D_{n,k}^T(\boldsymbol{\theta})\, \Omega(\boldsymbol{\theta})\, D_{n,k}(\boldsymbol{\theta}).$$

The *continuous updating weighted least squares estimator* for $\boldsymbol{\theta}_0$ is defined as

$$\widehat{\boldsymbol{\theta}}_{n,k} := \operatorname*{arg\,min}_{\boldsymbol{\theta}\in\Theta} f_{n,k}(\boldsymbol{\theta}) = \operatorname*{arg\,min}_{\boldsymbol{\theta}\in\Theta} \left\{D_{n,k}(\boldsymbol{\theta})^T\, \Omega(\boldsymbol{\theta})\, D_{n,k}(\boldsymbol{\theta})\right\}. \tag{2.4}$$

In Section 4.2, we will study the performance of this estimator for $\Omega(\boldsymbol{\theta}) = I_q$, the $q \times q$ identity matrix. Expression (2.4) then simplifies to

$$\widehat{\boldsymbol{\theta}}_{n,k} = \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{m=1}^{q} \left(\widehat{\ell}_{n,k}(\boldsymbol{c}_m) - \ell(\boldsymbol{c}_m;\boldsymbol{\theta})\right)^2.$$

Einmahl et al. (2017) show the consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}_{n,k}$ under the assumption that $\boldsymbol{\theta}_0$ is in the interior of Θ. In Section 3 we give the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_{n,k}$ without this restriction.

3 Inference on tail dependence parameters on the boundary of the parameter space

The asymptotic results presented in this section are a special case of Andrews (1999), where the asymptotic distribution of a general extremum estimator is established when one or more parameters lie on the boundary of the parameter space. The methodology to obtain the asymptotic distribution of $\sqrt{k}(\widehat{\boldsymbol{\theta}}_{n,k} - \boldsymbol{\theta}_0)$ can be summarized as follows: first, one shows that as $n \to \infty$, the criterion function $f_{n,k}(\boldsymbol{\theta})$ is equal to a quadratic function $q_{n,k}(\sqrt{k}(\boldsymbol{\theta} - \boldsymbol{\theta}_0))$ plus a term that does not depend on $\boldsymbol{\theta}$. If $\widehat{\boldsymbol{\theta}}_{n,k}$ is consistent, then its asymptotic distribution is only affected by the part of Θ close to $\boldsymbol{\theta}_0$; equivalently, we are only concerned with $\Theta - \boldsymbol{\theta}_0$ near the origin. If $\Theta - \boldsymbol{\theta}_0$ can be approximated near the origin by a convex cone Λ, one can show that minimizing $f_{n,k}(\boldsymbol{\theta})$ over $\boldsymbol{\theta} \in \Theta$ is asymptotically equivalent to minimizing $q_{n,k}(\boldsymbol{\lambda})$ over $\boldsymbol{\lambda} \in \Lambda$. Finally, $\sqrt{k}(\widehat{\boldsymbol{\theta}}_{n,k} - \boldsymbol{\theta}_0)$ converges in distribution to the argument minimizing the limit of $q_{n,k}(\boldsymbol{\lambda})$ as $n \to \infty$.

We set up some notations that will be needed in this section. Let $B_\varepsilon(\boldsymbol{\theta})$ denote an open ball centered at $\boldsymbol{\theta}$ with radius ε and let $C_\varepsilon(\boldsymbol{\theta})$ denote an open cube centered at $\boldsymbol{\theta}$ with sides of length 2ε. Let $\mathrm{cl}(\Theta)$ denote the closure of Θ. A set $\Gamma \subset \mathbb{R}^p$ is said to be *locally equal* to a set $\Lambda \subset \mathbb{R}^p$ if $\Gamma \cap B_\varepsilon(\mathbf{0}) = \Lambda \cap B_\varepsilon(\mathbf{0})$ for some $\varepsilon > 0$. A set $\Lambda \subset \mathbb{R}^p$ is a *cone* if $\boldsymbol{\lambda} \in \Lambda$ implies $a\boldsymbol{\lambda} \in \Lambda$ for all $a > 0$.

3.1 Estimation, consistency and asymptotic normality

When $\boldsymbol{\theta}_0$ is on the boundary of Θ, the map L is not defined and thus not differentiable on a neighbourhood of $\boldsymbol{\theta}_0$. We will need the following assumption.

(A1) Θ includes a set Θ^+ such that $\Theta^+ - \boldsymbol{\theta}_0$ equals the intersection of a union of orthants and an open cube $C_\varepsilon(\mathbf{0})$ for some $\varepsilon > 0$ and $\Theta \cap B_{\varepsilon_1}(\boldsymbol{\theta}_0) \subset \Theta^+$ for some $\varepsilon_1 > 0$.

If $\Theta - \boldsymbol{\theta}_0$ happens to be locally equal to a union of orthants, we can set $\Theta^+ = \Theta \cap C_\varepsilon(\boldsymbol{\theta}_0)$. This is the case for the models considered in Section 4. We will assume existence of the so-called *left/right (l/r) partial derivatives* on Θ^+; a formal definition is given in the appendix. The shape of Θ^+ is such that these can always be defined.

Write $\dot{L} := (\partial/\partial\boldsymbol{\theta})L(\boldsymbol{\theta}) \in \mathbb{R}^{q\times p}$ for $\boldsymbol{\theta} \in \Theta^+$, where $\dot{L}$ denotes the matrix of l/r partial derivatives. Let $\lambda_1(\boldsymbol{\theta}) > 0$ denote the smallest eigenvalue of $\Omega(\boldsymbol{\theta})$.

Theorem 3.1 (Existence, uniqueness and consistency)**.** *Let $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q \in [0,\infty)^d$ be $q \geq p$ points such that the map $L : \theta \mapsto (\ell(\boldsymbol{c}_m;\boldsymbol{\theta}))_{m=1}^q$ is a homeomorphism. Let $\boldsymbol{\theta}_0 \in \mathrm{cl}(\Theta)$ and assume that (A1) holds, that each element of $L(\boldsymbol{\theta})$ has continuous l/r partial derivatives of*

order two on Θ^+*, that* $\dot{L}(\boldsymbol{\theta}_0)$ *is of full rank, that* $\Omega : \Theta \to \mathbb{R}^{q\times q}$ *has continuous l/r partial derivatives on* Θ^+ *and that* $\inf_{\boldsymbol{\theta}\in\Theta} \lambda_1(\boldsymbol{\theta}) > 0$ *. Finally assume, for* $m = 1, \ldots, q$*,*

$$\widehat{\ell}_{n,k}(\boldsymbol{c}_m) \xrightarrow{\text{p}} \ell(\boldsymbol{c}_m; \boldsymbol{\theta}_0), \qquad \text{as } n \to \infty.$$

Then with probability tending to one, the minimizer $\widehat{\boldsymbol{\theta}}_{n,k}$ *in* (2.4) *exists and is unique. Moreover,*

$$\widehat{\boldsymbol{\theta}}_{n,k} \xrightarrow{\text{p}} \boldsymbol{\theta}_0, \qquad \text{as } n \to \infty.$$

We omit the proof of this theorem since it is directly obtained by replacing $B_\varepsilon(\boldsymbol{\theta}_0)$ by Θ^+ in the proof of Einmahl et al. (2017, Theorem 1).

When $f_{n,k}$ is not defined on a neighbourhood of $\boldsymbol{\theta}_0$, but there exists a set Θ^+ which satisfies (A1), a Taylor expansion of $f_{n,k}(\boldsymbol{\theta})$ around $f_{n,k}(\boldsymbol{\theta}_0)$ holds (Andrews, 1999, Theorem 6). For each $\boldsymbol{\theta} \in \Theta^+$, we have

$$f_{n,k}(\boldsymbol{\theta}) = f_{n,k}(\boldsymbol{\theta}_0) + Df_{n,k}(\boldsymbol{\theta}_0)^T(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T D^2 f_{n,k}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)/2 + R_{n,k}(\boldsymbol{\theta}),$$

where $Df_{n,k}$ and $D^2 f_{n,k}$ are based on l/r partial derivatives and $R_{n,k}(\boldsymbol{\theta})$ is the remainder term. We suppress dependence of Ω, $\dot{L}$ and $\boldsymbol{D}_{n,k}$ on $\boldsymbol{\theta}_0$ for ease of notation. From Einmahl et al. (2017, Proof of Theorems 3.2.1 and 3.2.2) we know that

$$Df_{n,k}(\boldsymbol{\theta}_0) = -2D_{n,k}^T \Omega \dot{L} + o_p(1), \qquad D^2 f_{n,k}(\boldsymbol{\theta}_0) = 2\dot{L}^T \Omega \dot{L} + o_p(1),$$

so the quadratic expansion above is equal to

$$f_{n,k}(\boldsymbol{\theta}) = f_{n,k}(\boldsymbol{\theta}_0) - 2D_{n,k}^T \Omega \dot{L}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \dot{L}^T \Omega \dot{L}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + R_{n,k}(\boldsymbol{\theta}), \tag{3.1}$$

Define

$$J := \dot{L}^T \Omega \dot{L} \in \mathbb{R}^{p\times p}, \qquad \boldsymbol{Y}_{n,k} := \sqrt{k} J^{-1} \dot{L}^T \Omega D_{n,k} \in \mathbb{R}^p,$$

and

$$q_{n,k}(\boldsymbol{\lambda}) = (\boldsymbol{\lambda} - \boldsymbol{Y}_{n,k})^T J (\boldsymbol{\lambda} - \boldsymbol{Y}_{n,k}).$$

Then expression (3.1) can be written as

$$f_{n,k}(\boldsymbol{\theta}) = f_{n,k}(\boldsymbol{\theta}_0) - k^{-1} \boldsymbol{Y}_{n,k}^T J \boldsymbol{Y}_{n,k} + k^{-1} q_{n,k}\left(\sqrt{k}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right) + R_{n,k}(\boldsymbol{\theta}).$$

We see that $f_{n,k}(\boldsymbol{\theta})$ is equal to a quadratic function plus a term that does not depend on $\boldsymbol{\theta}$ plus $R_{n,k}(\boldsymbol{\theta})$. In Andrews (2002, Lemma 3), it is shown that under the assumptions of Theorem 3.1 and (A3) below, the remainder term $R_{n,k}(\boldsymbol{\theta})$ is sufficiently small so that minimizing the quadratic approximation to $f_{n,k}(\boldsymbol{\theta})$ is equivalent to minimizing $q_{n,k}\left(\sqrt{k}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right)$.

Assume that

(A2) $\Theta - \boldsymbol{\theta}_0$ is locally equal to a convex cone $\Lambda \subset \mathbb{R}^p$.

Then

$$\inf_{\boldsymbol{\theta}\in\Theta} q_{n,k}\left(\sqrt{k}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right) = \inf_{\boldsymbol{\lambda}\in\Lambda} q_{n,k}(\boldsymbol{\lambda}) + o_p(1),$$

for some cone Λ. Finally, assume that

(A3) $\sqrt{k}\, D_{n,k}(\boldsymbol{\theta}_0) \xrightarrow{\text{d}} \boldsymbol{D} \sim \mathcal{N}_q(0, \Sigma(\boldsymbol{\theta}_0))$ as $n \to \infty$, for some covariance matrix $\Sigma(\boldsymbol{\theta}_0)$.

Assumption (A3) holds for the nonparametric estimators presented in Section 2.2 under some general regularity conditions. The matrix Σ can be expressed in terms of ℓ and its first-order partial derivatives. We have

$$\boldsymbol{Y}_{n,k} \xrightarrow{\mathrm{d}} \boldsymbol{Y} := J^{-1}\dot{L}^T\Omega\boldsymbol{D}, \quad q_{n,k}(\boldsymbol{\lambda}) \xrightarrow{\mathrm{d}} q(\boldsymbol{\lambda}) := (\boldsymbol{\lambda}-\boldsymbol{Y})^T J(\boldsymbol{\lambda}-\boldsymbol{Y}) \quad \text{for all } \boldsymbol{\lambda}\in\mathbb{R}^p.$$

Theorem 3.2 (Asymptotic normality)**.** *If all of the above assumptions hold, then*

$$\sqrt{k}(\widehat{\boldsymbol{\theta}}_{n,k}-\boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} \widehat{\boldsymbol{\lambda}} = \operatorname*{arg\,min}_{\boldsymbol{\lambda}\in\Lambda} q(\boldsymbol{\lambda}), \qquad \text{as } n\to\infty.$$

Here $\widehat{\boldsymbol{\lambda}}$ can be interpreted as the projection of $\boldsymbol{Y}$ on Λ with respect to the norm $\|y\|_J := y^T Jy$. This theorem is a special case of Andrews (1999, Theorem 3). In the appendix, we verify that our assumptions are sufficient. Note that if Θ includes a neighbourhood of $\boldsymbol{\theta}_0$, then $\Lambda=\mathbb{R}^p$ and we find the same distribution as in Einmahl et al. (2017, Theorem 2) since $\widehat{\boldsymbol{\lambda}}=\boldsymbol{Y}$.

3.2 Simplifying the asymptotic distribution

The goal of this section is to simplify the asymptotic distribution of $\sqrt{k}(\widehat{\boldsymbol{\theta}}_{n,k}-\boldsymbol{\theta}_0)$ and to give conditions under which (part of) $\widehat{\boldsymbol{\lambda}}$ has a closed-form expression. We start by partitioning the vector $\boldsymbol{\theta}_0\in\mathbb{R}^p$ into two subvectors, $\boldsymbol{\beta}_0\in\mathbb{R}^c$ and $\boldsymbol{\delta}_0\in\mathbb{R}^{p-c}$ for $c\in\{1,\dots,p\}$, where $\boldsymbol{\delta}_0$ consists of all parameters that are not on the boundary of the parameter space. Let $\boldsymbol{G} := \dot{L}^T\Omega\boldsymbol{D}$, i.e., $\boldsymbol{Y}=J^{-1}\boldsymbol{G}$. We partition $\boldsymbol{\theta}$, $\boldsymbol{Y}$, J, $\boldsymbol{G}$ and $\boldsymbol{\lambda}$ accordingly:

$$\boldsymbol{\theta}=\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\delta}\end{pmatrix}, \quad \boldsymbol{Y}=\begin{pmatrix}\boldsymbol{Y_\beta}\\ \boldsymbol{Y_\delta}\end{pmatrix}, \quad J=\begin{pmatrix}J_\beta & J_{\beta\delta}\\ J_{\delta\beta} & J_\delta\end{pmatrix}, \quad \boldsymbol{G}=\begin{pmatrix}\boldsymbol{G_\beta}\\ \boldsymbol{G_\delta}\end{pmatrix}, \quad \boldsymbol{\lambda}=\begin{pmatrix}\boldsymbol{\lambda}_\beta\\ \boldsymbol{\lambda}_\delta\end{pmatrix}.$$

Let I_c denote the $c\times c$ identity matrix. For $H := (I_c : \mathbf{0})\in\mathbb{R}^{c\times p}$, define

$$q_\beta(\boldsymbol{\lambda}_\beta) := (\boldsymbol{\lambda}_\beta-\boldsymbol{Y}_\beta)^T\left(HJ^{-1}H^T\right)^{-1}(\boldsymbol{\lambda}_\beta-\boldsymbol{Y}_\beta).$$

Suppose that

(A4) The cone Λ of assumption (A2) is equal to the product set $\Lambda_{\boldsymbol{\beta}}\times\mathbb{R}^{p-c}$, where $\Lambda_{\boldsymbol{\beta}}\subset\mathbb{R}^c$ is a cone.

Corollary 3.3. *If all of the above assumptions hold, then*

$$\sqrt{k}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \xrightarrow{\mathrm{d}} \widehat{\boldsymbol{\lambda}}_\beta := \operatorname*{arg\,min}_{\boldsymbol{\lambda}_\beta\in\Lambda_\beta} q_\beta(\boldsymbol{\lambda}_\beta), \quad \sqrt{k}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}_0) \xrightarrow{\mathrm{d}} J_\delta^{-1}\boldsymbol{G}_\delta - J_\delta^{-1}J_{\delta\beta}\widehat{\boldsymbol{\lambda}}_\beta$$

When Λ_β is defined by equality and/or inequality constrains, a closed-form expression for $\widehat{\boldsymbol{\lambda}}_\beta$ can be computed; we give some examples that will be used in Section 4. For a more formal solution, see Andrews (1999, Theorem 5).

Example 3.1 ($c=1$)**.** Suppose that $\Theta=[0,1]^p$. If $\beta_0=0$, then $\Lambda_\beta=[0,\infty)$ and $\widehat{\lambda}_\beta=\max(Y_\beta,0)$. If $\beta_0=1$, then $\Lambda_\beta=(-\infty,0]$ and $\widehat{\lambda}_\beta=\min(Y_\beta,0)$.

Example 3.2 ($c = 2$)**.** Suppose that $\Theta = [0,1]^p$. If $\boldsymbol{\beta}_0 = (1,1)$, then $\Lambda_\beta = (-\infty, 0]^2$. Let $\rho := HJ^{-1}H^T$. Then

$$\begin{aligned}\widehat{\boldsymbol{\lambda}}_\beta := &\mathbb{1}\left\{Y_{\beta 1} < 0, Y_{\beta 2} < 0\right\} \boldsymbol{Y}_\beta \\ &+ \mathbb{1}\left\{Y_{\beta 1} - \rho_{21} Y_{\beta 2} < 0, Y_{\beta 2} \geq 0\right\} (Y_{\beta 1} - \rho_{21} Y_{\beta 2}, 0)^T \\ &+ \mathbb{1}\left\{Y_{\beta 1} \geq 0, Y_{\beta 2} - \rho_{12} Y_{\beta 1} < 0\right\} (0, Y_{\beta 2} - \rho_{12} Y_{\beta 1})^T.\end{aligned}$$

If $\boldsymbol{\beta}_0 = (0,0)$, $\boldsymbol{\beta}_0 = (0,1)$ or $\boldsymbol{\beta}_0 = (1,0)$, similar expressions can be obtained by reversing the first and/or the second inequalities in each of the indicator functions above.

3.3 Hypothesis testing for tail dependence parameters on the boundary of the parameter space

We are interested in constructing hypothesis tests for parameter values that, under the null hypothesis, are on the boundary of the alternative hypothesis. We propose two test statistics, whose asymptotic distribution follows from the results in Andrews (2001).

Assume that there are no nuisance parameters on the boundary, i.e., all components of $\boldsymbol{\theta}_0$ that lie on the boundary are part of the null hypothesis. We are interested in testing

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_* \quad \text{vs} \quad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_* \qquad \boldsymbol{\beta}_* \in \mathbb{R}^c.$$

Let $\Theta_0 := \{\boldsymbol{\theta} \in \Theta : \boldsymbol{\theta} = (\boldsymbol{\beta}_*, \boldsymbol{\delta}) \text{ for some } \boldsymbol{\delta} \in \mathbb{R}^{p-c}\}$ denote the restricted parameter space. Assume that

(A5) For all $\boldsymbol{\theta} \in \Theta_0$, Θ is a product set with respect to $(\boldsymbol{\beta}, \boldsymbol{\delta})$ local to $\boldsymbol{\theta}$. That is, for all $\boldsymbol{\theta} \in \Theta_0$,

$$\Theta \cap B_\varepsilon(\boldsymbol{\theta}) = (\mathcal{B} \times \Delta) \cap B_\varepsilon(\boldsymbol{\theta}) \quad \text{for some } \mathcal{B} \subset \mathbb{R}^c, \Delta \subset \mathbb{R}^{p-c} \text{ and } \varepsilon > 0.$$

Define $\widehat{\boldsymbol{\theta}}^{(0)}_{n,k} := \arg\min_{\boldsymbol{\theta} \in \Theta_0} f_{n,k}(\boldsymbol{\theta})$. A deviance test statistic can be defined as

$$T^{(1)}_{n,k} := k\left(f_{n,k}(\widehat{\boldsymbol{\theta}}^{(0)}_{n,k}) - f_{n,k}(\widehat{\boldsymbol{\theta}}_{n,k})\right).$$

Corollary 3.4. *Suppose $\boldsymbol{\theta}_0 \in \Theta_0$ and all previous assumptions hold. Then*

$$T^{(1)}_{n,k} \xrightarrow{\text{d}} \widehat{\boldsymbol{\lambda}}^T_\beta \left(HJ^{-1}H^T\right)^{-1} \widehat{\boldsymbol{\lambda}}_\beta.$$

Recall that $\boldsymbol{G} = \dot{L}^T \Omega \boldsymbol{D}$ and let $\mathcal{J}$ denote its covariance matrix, i.e., $\boldsymbol{G} \sim \mathcal{N}_p(\boldsymbol{0}, \mathcal{J})$ with $\mathcal{J} = \dot{L}^T \Omega \Sigma \Omega \dot{L} \in \mathbb{R}^{p \times p}$. Note that $\mathcal{J} = J$ if and only if $\Omega = \Sigma^{-1}$. If $\Lambda_\beta = \mathbb{R}^c$ (no parameters on the boundary) and $\mathcal{J} = J$, then

$$T^{(1)}_{n,k} \xrightarrow{\text{d}} \boldsymbol{Y}^T_\beta \left(HJ^{-1}H^T\right)^{-1} \boldsymbol{Y}_\beta \sim \chi^2_c,$$

where χ^2_c denotes a chi-squared random variable with c degrees of freedom.

A Wald-type test statistic can be based on the quadratic form in $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*$. Let $\widehat{V}^{-1}_n := (HJ^{-1}_n H^T)^{-1} \in \mathbb{R}^{c \times c}$ denote a weight matrix where $J_n := J(\widehat{\boldsymbol{\theta}}_{n,k})$. Define

$$T^{(2)}_{n,k} := k\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*\right)^T \widehat{V}^{-1}_n \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*\right).$$

Corollary 3.5. *Suppose $\boldsymbol{\theta}_0 \in \Theta_0$ and all of the above assumptions hold. Then*

$$T_{n,k}^{(2)} \xrightarrow{\mathrm{d}} \widehat{\boldsymbol{\lambda}}_\beta^T \left(HJ^{-1}H^T\right)^{-1} \widehat{\boldsymbol{\lambda}}_\beta.$$

Corollary 3.4 is a special case of Andrews (2001, Theorem 4c) and Corollary 3.5 is a special case of Andrews (2001, Theorem 6d); their proofs are direct and thus ommited.

4 Generalized max-linear models

4.1 Theory and definitions

Let $\boldsymbol{Z} = (Z_1, \ldots, Z_r)$ be a random vector whose margins are unit Fréchet, $\mathbb{P}[Z_s \leq z] = \exp(-1/z)$ for $z > 0$ and $s \in \{1, \ldots, r\}$. A *generalized max-linear* or *max-factor* model for $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ is obtained by setting

$$Y_j = \max_{s=1,\ldots,r} b_{js} Z_s, \qquad j \in \{1, \ldots, d\}, \tag{4.1}$$

where the factor loadings b_{js} are non-negative constants such that all column sums of the $d \times r$ matrix $B := (b_{js})_{j,s}$ are positive.

Lemma 4.1. *The model in* (4.1) *has stable tail dependence function*

$$\ell_{\boldsymbol{Y}}(\boldsymbol{y}) = \ell_{\boldsymbol{Z}}\left(\max_{j=1,\ldots,d} \frac{b_{j1} z_j}{\ell_{\boldsymbol{Z}}(\boldsymbol{b}_j)}, \ldots, \max_{j=1,\ldots,d} \frac{b_{jr} z_j}{\ell_{\boldsymbol{Z}}(\boldsymbol{b}_j)}\right), \tag{4.2}$$

where $\boldsymbol{b}_j = (b_{j1}, \ldots, b_{jr})$. If $\ell_{\boldsymbol{Z}}(\boldsymbol{b}_j) = 1$ for every $j \in \{1, \ldots, d\}$, then $\boldsymbol{Y}$ has unit Fréchet margins.

See also Falk et al. (2015, Lemma 2.1). This means that we can choose any existing stable tail dependence function $\ell_{\boldsymbol{Z}}$ and plug it into (4.2) to get $\ell_{\boldsymbol{Y}}$.

Example 4.1 (Regular max-linear model)**.** If $\boldsymbol{Z}$ has independent components, i.e., $\ell_{\boldsymbol{Z}}(\boldsymbol{z}) = z_1 + \cdots + z_r$, then

$$\ell_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{s=1}^{r} \max_{j=1,\ldots,d} \frac{b_{js} y_j}{\sum_{t=1}^{r} b_{jt}} = \sum_{s=1}^{r} \max_{j=1,\ldots,d} a_{js} y_j,$$

for $a_{js} := b_{js} / \sum_{t=1}^{r} b_{jt}$, which is exactly what is obtained in Einmahl et al. (2012, Lemma 6.1) by direct computation.

Example 4.2 (Logistic max-linear model)**.** Let $\ell_{\boldsymbol{Z}}$ be the stable tail dependence function of the well-known *logistic model*,

$$\ell_{\boldsymbol{Z}}(\boldsymbol{z}) = \left(z_1^{1/\alpha} + \cdots + z_r^{1/\alpha}\right)^{\alpha}, \qquad \alpha \in (0, 1].$$

Set $\sum_{s=1}^{r} b_{js}^{1/\alpha} = 1$ for $j \in \{1, \ldots, d\}$ so that $\boldsymbol{Y}$ has unit Fréchet margins. The stable tail dependence function is

$$\ell_{\boldsymbol{Y}}(\boldsymbol{y}) = \left(\sum_{s=1}^{r} \left(\max_{j=1,\ldots,d} a_{js} y_j\right)^{1/\alpha}\right)^{\alpha}.$$

For $\alpha = 1$, this is the regular max-linear model. For $\alpha \downarrow 0$, $\ell_{\boldsymbol{Y}}(\boldsymbol{y}) = \max(y_1, \ldots, y_d)$.

If we require $\boldsymbol{Y}$ to have unit Fréchet margins, the last column of B is defined by the first $(r-1)$ columns and by α. The (generalized) max-linear model is identifiable when the constants $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q$ are well-chosen; see Einmahl et al. (2017, Appendix B). For $\boldsymbol{\beta} \subset \boldsymbol{\theta}$, we are interested in testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_*$ for $\boldsymbol{\beta}_* \in \{0, 1\}^c$.

4.2 Simulation studies

4.2.1 Testing the strucutre of a max-linear model

We consider models with $r = 2$ factors in dimension $d = 3$, i.e., with $\boldsymbol{\theta} = (b_1, b_2, b_3, \alpha)$ and parameter matrix

$$B = \begin{pmatrix} b_1 & (1 - b_1^{1/\alpha})^\alpha \\ b_2 & (1 - b_2^{1/\alpha})^\alpha \\ b_3 & (1 - b_3^{1/\alpha})^\alpha \end{pmatrix}.$$

The following four scenarios are considered:

Model 1 Let $\boldsymbol{\theta}_0 = (1, 0.725, 0.35, 1)$. We estimate a regular max-linear model and we test $H_0 : b_1 = 1$. Assumption (A4) holds with $\Lambda = (-\infty, 0] \times \mathbb{R}^2$.

Model 2 Let $\boldsymbol{\theta}_0 = (1, 0.725, 0, 1)$. We estimate a regular max-linear model and we test $H_0 : (b_1, b_3) = (1, 0)$. Assumption (A4) holds with $\Lambda = (-\infty, 0] \times [0, \infty) \times \mathbb{R}$.

Model 3 Let $\boldsymbol{\theta}_0 = (1, 0.725, 0.35, 0.6)$. We estimate a logistic max-linear model and we test $H_0 : b_1 = 1$. Assumption (A4) holds with $\Lambda = (-\infty, 0] \times \mathbb{R}^3$.

Model 4 Let $\boldsymbol{\theta}_0 = (1, 0.725, 0.35, 1)$. We estimate a logistic max-linear model and we test $H_0 : (b_1, \alpha) = (1, 1)$. Assumption (A4) holds with $\Lambda = (-\infty, 0]^2 \times \mathbb{R}^2$.

We start by assessing the quality of our estimator in these four scenarios. To this end, we simulate 500 samples of size $n = 5000$ from these four models and we estimate the parameters for $k \in \{25, 50, \ldots, 300\}$. We use the empirical and the beta stable tail dependence function as initial estimators. The points $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q$ are chosen such that $\boldsymbol{c}_m \in \{0, 0.1, \ldots, 1\}^p$, so that $q = 1300$. We take $\Omega = I_q$, the $q \times q$ identity matrix for the weight matrix; we do not consider an optimal weight matrix here because inverting Σ is often hindered by numerical problems and the influence on the quality of estimation is negligible; see Einmahl et al. (2017).

Figure 1 shows the root mean squared error (RMSE) of the parameter estimates in the four max-linear models based on the empirical stable tail dependence function (solid lines) and the beta stable tail dependence function (dashed lines). We see that the empirical and beta stable tail dependence functions behave similarly, although the beta estimator performs slightly better (the empirical estimator outperforms the beta estimator for model 4, $\alpha = 1$ only). The gap between the two estimators increases as the sample size n decreases, especially for low values of k (plots not shown).

Next, we study the performance of the two test statistics introduced in Section 3.3. Table 1 shows the empirical level of the test statistics $T_{n,k}^{(1)}$ and $T_{n,k}^{(2)}$ for $k \in \{25, 50\}$ based on the empirical and the beta stable tail dependence function, using a significance level of 0.05. On average, the empirical level of the test is lower when using the beta stable tail dependence function. When testing $H_0 : b_1 = 1$ in model 3, we see that using the test statistic $T_{n,k}^{(1)}$ we never reject the null hypothesis. This illustrates a major drawback of the deviance test statistic, namely the fact that it allows for compensation: when b_1 is far from one, the values of $f_{n,k}(\widehat{\boldsymbol{\theta}}_{n,k}^{(0)})$ and $f_{n,k}(\widehat{\boldsymbol{\theta}}_{n,k})$ can still be close to each other by changing the value of α. Moreover, $T_{n,k}^{(1)}$ is numerically slower when there are many parameters, because one has to calculate $\widehat{\boldsymbol{\theta}}_{n,k}^{(0)}$.

To assess the power of the two test statistics, we consider the following models, similar to those used above:

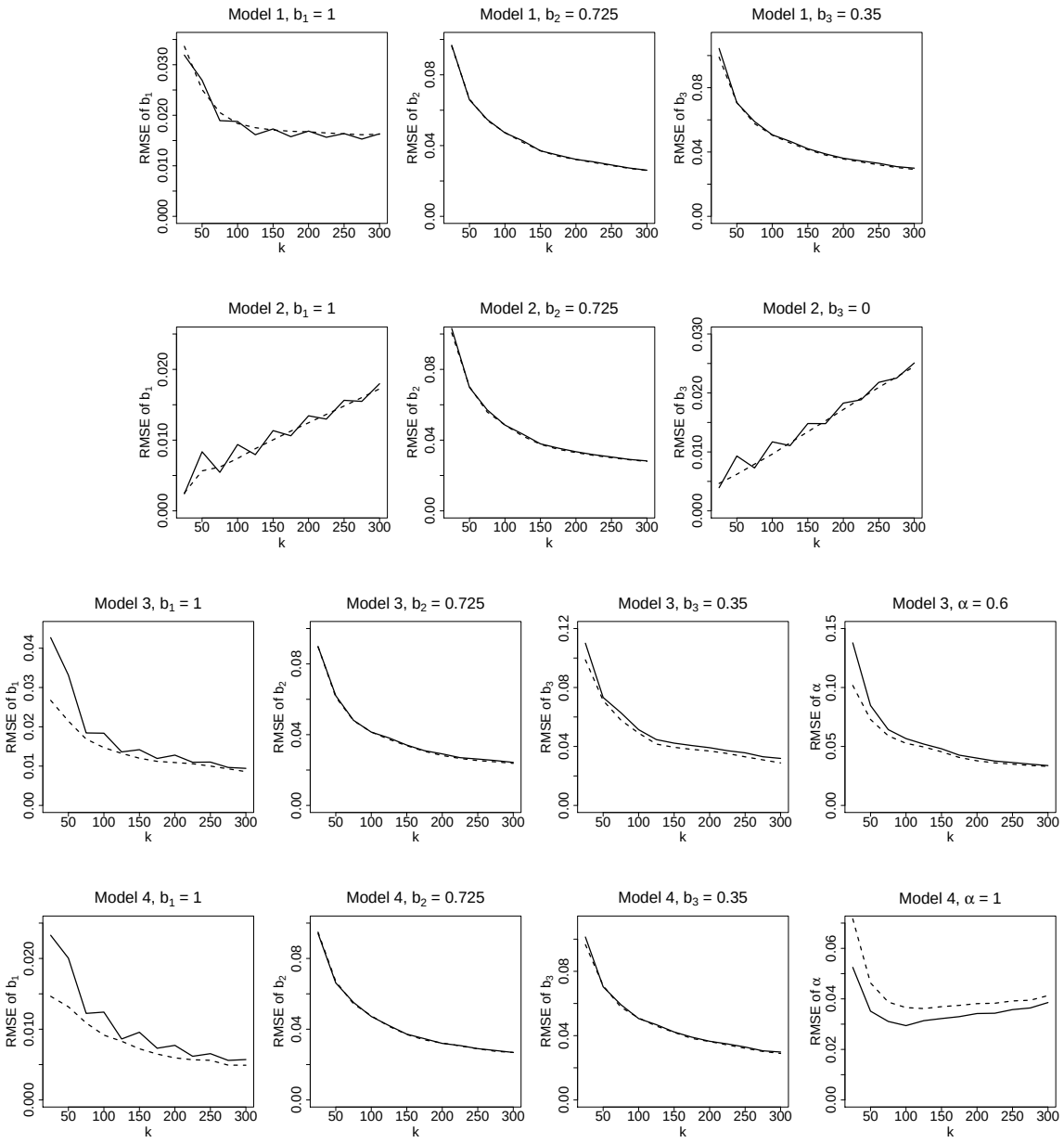

Figure 1: RMSE for models 1–4 based on the empirical stable tail dependence function (solid lines), and the beta stable tail dependence function (dashed lines); 500 samples of size $n = 5000$.

	$T_{n,k}^{(1)}$				$T_{n,k}^{(2)}$			
	$k = 25$		$k = 50$		$k = 25$		$k = 50$	
	emp	beta	emp	beta	emp	beta	emp	beta
M1	3.4	2.6	5.4	3.4	3.6	3.4	5.4	4.2
M2	1.0	1.2	18.0	6.0	1.0	1.4	18.2	6.6
M3	0.0	0.0	0.0	0.0	3.6	1.6	8.8	3.4
M4	3.8	2.6	6.6	4.4	1.8	2.0	2.8	2.2

Table 1: Empirical level of the test statistics $T_{n,k}^{(1)}$ and $T_{n,k}^{(2)}$ for models 1–4 (M1–M4) for $n = 5000$ based on the empirical and the beta stable tail dependence function for a significance level of 0.05.

Model 5 Let $\boldsymbol{\theta} = (b_1, 0.725, 0.35, 1)$ for $b_1 \in \{0.9, 0.95\}$. We estimate a regular max-linear model and we test $H_0 : b_1 = 1$.

Model 6 Let $\boldsymbol{\theta} = (b_1, 0.725, b_3, 1)$ for $b_1 \in \{0.9, 0.95\}$ and $b_3 \in \{0.05, 0.1\}$. We estimate a regular max-linear model and we test $H_0 : (a_1, a_3) = (1, 0)$.

Model 7 Let $\boldsymbol{\theta} = (b_1, 0.725, 0.35, 0.6)$ for $b_1 \in \{0.9, 0.95\}$. We estimate a logistic max-linear model and we test $H_0 : b_1 = 1$.

Model 8 Let $\boldsymbol{\theta} = (b_1, 0.725, 0.35, \alpha)$ for $b_1 \in \{0.9, 0.95\}$ and $\alpha \in \{0.8, 0.9\}$. We estimate a logistic max-linear model and we test $H_0 : (b_1, \alpha) = (1, 1)$.

Table 2 shows the empirical power of the test statistics $T_{n,k}^{(1)}$ and $T_{n,k}^{(2)}$ calculated using the empirical and the beta stable tail dependence function. The empirical power of the test is usually higher when using the beta stable tail dependence function, except for model 7. We see again that $T_{n,k}^{(1)}$ does not perform well when compensation between parameter values is possible: the power of the tests for model 3 are near zero.

4.2.2 Testing the number of factors in a max-linear model

Consider a d-dimensional regular max-linear model with r factors. We would like to study the capability of our test statistics to correctly retrieve the true number of factors, i.e., for $s \in \{1, \ldots, r\}$, we wish to test the hypothesis $H_0 : (b_{1s}, \ldots, b_{js}) = \mathbf{0}$. However, the max-linear model is only defined for parameter values with $\sum_{j=1}^{d} b_{js} > 0$ for all $s \in \{1, \ldots, d\}$; when this condition is not met, there are zeroes on the diagonal of $\Sigma(\boldsymbol{\theta}_0)$ and thus zeroes on the diagonal of $M(\boldsymbol{\theta}_0)$, making computation of $\widehat{\boldsymbol{\lambda}}_\beta$ impossible. We solve this problem by computing the asymptotic distribution of the test statistics using $\widehat{\boldsymbol{\theta}}_{n,k}$ rather than $\boldsymbol{\theta}_0$.

Consider a two-dimensional model with parameter vector $\boldsymbol{\theta}_0 = (b_{11}, b_{21}, b_{12}, b_{22}) = (0, 0, 0.825, 0.625)$, i.e., the first column of B is zero and the model has effectively two factors. We estimate a model with three factors and we test whether a two-factor model suffices, i.e., $H_0 : (b_{11}, b_{21}) = (0, 0)$. We simulate 500 samples of size $n = 5000$ from this model and we estimate the parameters for $k \in \{25, 50, 75, 100, 125\}$ using the empirical stable tail dependence function as the initial estimator. The beta stable tail dependence function is not suited for this type of problem because of its smoothness: it tends to include more factors than necessary. We de not consider the deviance test statistic because of the compensation discussed

		$T_{n,k}^{(1)}$				$T_{n,k}^{(2)}$			
		$k=25$		$k=50$		$k=25$		$k=50$	
		emp	beta	emp	beta	emp	beta	emp	beta
M5	$b_1 = 0.9$	50.4	50.4	84.8	83.0	50.8	56.8	85.2	84.6
M5	$b_1 = 0.95$	21.8	21.0	44.2	41.0	23.4	23.6	44.0	44.8
M6	$(b_1, b_3) = (0.9, 0.1)$	97.2	99.0	100.0	100.0	97.2	99.0	100.0	100.0
M6	$(b_1, b_3) = (0.9, 0.05)$	92.6	95.0	99.8	99.8	92.6	95.0	99.8	99.8
M6	$(b_1, b_3) = (0.95, 0.1)$	91.2	94.8	100.0	99.8	91.2	94.6	100.0	99.8
M6	$(b_1, b_3) = (0.95, 0.05)$	79.0	83.2	98.2	97.2	78.8	83.6	98.2	97.2
M7	$b_1 = 0.9$	0.2	0.0	5.6	0.2	37.6	8.0	72.2	50.6
M7	$b_1 = 0.95$	0.0	0.0	0.8	0.0	15.4	4.2	37.8	21.6
M8	$(b_1, \alpha) = (0.9, 0.8)$	88.6	93.6	99.8	100.0	86.4	92.4	99.6	99.8
M8	$(b_1, \alpha) = (0.9, 0.9)$	75.4	79.0	97.0	97.8	68.0	72.4	93.8	93.6
M8	$(b_1, \alpha) = (0.95, 0.8)$	72.0	79.4	97.2	98.0	69.2	77.2	93.6	95.4
M8	$(b_1, \alpha) = (0.95, 0.9)$	44.2	45.2	80.6	80.4	33.4	37.8	66.8	64.2

Table 2: Empirical power of the test statistics $T_{n,k}^{(1)}$ and $T_{n,k}^{(2)}$ for models 5–8 (M5–M8) for $n = 5000$ based on the empirical and the beta stable tail dependence function for a significance level of 0.05.

earlier; based on $T_{n,k}^{(1)}$, we would never reject the null hypothesis. The points $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q$ are chosen such that

$$\boldsymbol{c}_m \in \{0, 0.005, 0.01, 0.1, 0.2, \ldots, 0.8, 0.9, 0.99, 0.995, 1\}^2, \qquad m \in \{1, \ldots, q\},$$

so that $q = 196$; we have chosen more values close to the boundaries to avoid identifiability problems. Table 3 (upper row) shows the empirical level of the Wald test statistic $T_{n,k}^{(2)}$ using a significance level of 0.05. To assess the power of the two test statistics, we consider the three models with $\boldsymbol{\theta}_0 = (0.1, 0, 0.825, 0.625)$, $\boldsymbol{\theta}_0 = (0, 0.1, 0.825, 0.625)$ and $\boldsymbol{\theta}_0 = (0.1, 0.1, 0.825, 0.625)$ respectively. Table 3 (second to fourth row) shows the empirical power of the test statistic $T_{n,k}^{(2)}$. It is surprising to see that the power is lower in the fourth row than in the second or third. This might be due to the fact that for $\boldsymbol{\theta}_0 = (0.1, 0.1, 0.825, 0.625)$, there are two columns with values near zero.

	$k=25$	$k=50$	$k=75$	$k=100$
$\boldsymbol{\theta}_0 = (0, 0, 0.175, 0.375)$	9.7	6.5	7.1	5.8
$\boldsymbol{\theta}_0 = (0.1, 0, 0.175, 0.375)$	39.6	64.4	66.6	76.6
$\boldsymbol{\theta}_0 = (0, 0.1, 0.175, 0.375)$	31.2	52.8	50.6	60.8
$\boldsymbol{\theta}_0 = (0.1, 0.1, 0.175, 0.375)$	26.2	34.2	35.2	42.6

Table 3: Empirical level (top row) and power (second to fourth row) of the test statistic $T_{n,k}^{(2)}$ for $n = 5000$ based on the empirical stable tail dependence function for a significance level of 0.05.

5 Application to stock market data

Consider two large industrial goods companies, Honeywell International and 3M, both listed at the New York Stock Exchange. We take the weekly negative log-returns of the stock prices of these two companies from `https://finance.yahoo.com` for the period January 1970 to August 2017, leading to a sample of size $n = 2482$. Figure 2 shows the time series plots of log-returns and the dependence structure for the returns standardized to unit Pareto margins, plotted on the exponential scale. The time series plots look approximately stationary except for a few outliers for Honeywell.

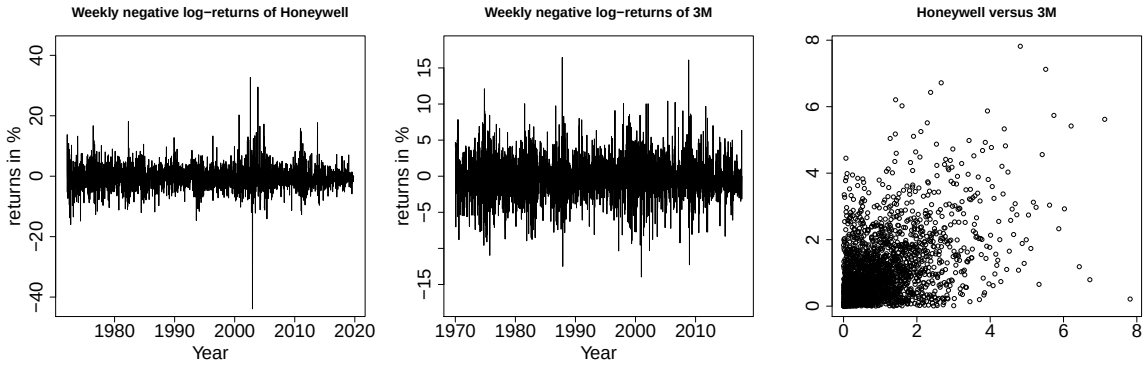

Figure 2: Time series plot of weekly negative log-returns of the stock prices of Honeywell (left) and 3M (middle); scatterplot of weekly negative log-returns of the stock prices of Honeywell versus 3M on the unit Pareto scale, plotted on the exponential scale (right).

Before fitting a max-linear model, we needs to decide how many factors we will use. Table 4 shows the result of fitting a four-factor model to our data based on the empirical stable tail dependence function. Values of $k \in \{25, 30, \ldots, 70, 75\}$ were considered: we chose $k = 60$ because parameter estimates are stable around this value. Standard errors were calculated using the asymptotic variance matrix M.

The results in Table 4 suggest to test $H_0 : (b_{14}, b_{24}) = (0, 0)$; the value of the test statistic $T^{(2)}_{n,k}$ is 1.76. Comapring with a critical value of 2.37 based on a significance level of 0.05, we find that we cannot reject the null hypothesis and thus we consider a model with $r = 3$ factors. The parameter estimates in Table 5 (left) suggest that an even simpeler model might suffice: if the hypothesis $H_0 : (b_1, b_{23}) = (0, 0)$ cannot be rejected, then we find that a bivariate Marshall–Olkin model suffices. The value of the test statistic is 1.63 and the critical value is 4.20. The final parameter estimates for the Marshall–Olkin model are given in Table 5 (right).

Honeywell	0.01 (0.02)	0.76 (0.06)	0.12 (0.06)	0.10 (0.06)
3M	0.60 (0.11)	0.01 (0.02)	0.29 (0.13)	0.10 (0.06)

Table 4: Parameter matrix for a bivariate max-linear model with $r = 4$ factors for $k = 60$; standard errors are in parentheses.

	$r = 3$ factors			Marshall–Olkin		
GE	0.03 (0.02)	0.78 (0.06)	0.19 (0.06)	–	0.71 (0.01)	0.27 (0.01)
3M	0.70 (0.08)	0.02 (0.03)	0.28 (0.09)	0.73 (0.03)	–	0.29 (0.03)

Table 5: Parameter matrices for a max-linear model with $r = 3$ factors (left) and a bivariate Marshall–Olkin model (right) for $k = 60$; standard errors are in parentheses.

6 Discussion

We proposed two test statistics for parameters that, under the null hypothesis, are on the boundary of the alternative hypothesis. These test were applied to the (generalized) max-linear model, but the results are generic and could be applied to any multivariate extreme-value model where submodels appear at boundary values: for instance, the Smith model is a special case of the Brown–Resnick process when the shape parameter of the latter is on the boundary.

The test statistics are convenient because their asymptotic distribution has an explicit expression. In practice however, this distribution becomes cumbersome when testing for a higher dimensional ($c > 4$) parameter vector. A possible solution might be to consider a multiple testing procedure with a Bonferroni-type of correction.

A potential application of the theory in this paper can be found in recursive max-linear models, which are max-linear models defined on a directed acyclic graph (Gissibl and Klüppelberg, 2015). In Gissibl et al. (2017), a procedure to reconstruct the graph structure of a dataset is outlined. This procedure is based on testing which pairwise tail dependence coefficients are zero, i.e., it is based on testing whether $\ell_{uv}(1, 1) = 2$ for all pairs $u, v \in \{1, \dots, d\}$. This is equivalent to testing whether B is equal to a 2×2 identity matrix for each pair of variables, and could hence be done using our methodology.

A Appendix

Definition (*Left/right partial derivatives*)**.** Let f be a function whose support includes $\mathcal{X} \subset \mathbb{R}^p$ and let $\boldsymbol{a} \in \mathcal{X}$. Suppose that $\mathcal{X} - \boldsymbol{a}$ equals the intersection of a union of orthants and an open cube $C_\varepsilon(\mathbf{0})$ for some $\varepsilon > 0$, i.e., $\mathcal{X} - \boldsymbol{a}$ is locally equal to a union of orthants. The function f is said to have *left/right (l/r) partial derivatives* of order 1 on $\mathcal{X}$ if

1. it has partial derivatives at each interior point of $\mathcal{X}$;
2. it has partial derivatives at each boundary point of $\mathcal{X}$ with respect to coordinates that can be perturbed to the left and right;
3. it has left (right) partial derivatives at each boundary point of $\mathcal{X}$ with respect to coordinates that can be perturbed only to the left (right).

The shape of $\mathcal{X}$ is such that for all $\boldsymbol{x} \in \mathcal{X}$ and for all $i \in \{1, \dots, p\}$, it is possible to perturb x_i to the left, the right, or both and stay within $\mathcal{X}$. This means that it is always possible to define the left, right, or two-sided partial derivative of f with respect to x_i. We say that f has l/r partial derivatives of order k on $\mathcal{X}$ for $k \geq 2$ if f has l/r partial derivatives of order $k - 1$ on $\mathcal{X}$ and each of the latter has l/r partial derivatives on $\mathcal{X}$. When we say the f has continuous l/r partial derivatives, continuity is defined in terms of local perturbations within $\mathcal{X}$ only.

Proof of Theorem 3.2. This theorem is a special case of Andrews (1999, Theorem 3b). A closely related work is Andrews (2002), where the focus is on generalized method of moment estimators. This setting is closer to ours but uses a convergence rate of $\sqrt{n}$, whereas Andrews (1999) allows for more generality.

The quantities ℓ_T, B_T and R_T in Andrews (1999) correspond to $-(k/2)f_{n,k}$, $\sqrt{k}I_p$ and $-(k/2)R_{n,k}$ respectively in this paper. Theorem 3b in Andrews (1999) holds under the assumptions 1, 2*, 3*, 5*, and 6 of that paper. We show that these are implied by ours:

Assumption 1 The assumptions made in Theorem 3.1 imply assumption 1.

Assumption 2* Assumption GMM2 in Andrews (2002) implies Assumption 2*; this is proven in Andrews (2002, Lemma 3). We show that our assumption imply GMM2: assumption GMM2(a) holds because because $D_{n,k}(\boldsymbol{\theta})$ converges in probability to $L(\boldsymbol{\theta}_0) - L(\boldsymbol{\theta})$. Assumption GMM2(b) holds if assumption GMM$2^{2*}(b)$ holds, which in turn is implied by the assumptions of Theorem 3.1 and (A1). Assumption GMM2(c) holds since $\boldsymbol{D}(\boldsymbol{\theta}_0) = \mathbf{0}$. Assumption GMM2(d) holds since $D_{n,k}(\boldsymbol{\theta}) - \boldsymbol{D}(\boldsymbol{\theta}) - D_{n,k}(\boldsymbol{\theta}_0) = 0$. Finally, Assumption GMM2(e) holds automatically since our weight matrix isn't random.

Assumption 3* Assumption 3* holds because of (A3) and because J is non-random, symmetric and non-singular.

Assumption 5* Assumption 5* holds because of (A2) and because $B_T = \sqrt{k}I_p$ and $\sqrt{k} \to \infty$ as $n \to \infty$.

Assumption 6 Assumption (A2) implies assumption 6.

□

Proof of Corollary 3.3. This corollary is a special case of Andrews (1999, Corollary 1b), where no parameter ψ appears. Assumptions 1, 2*, 3*,5* and 6–8 in that paper are needed: in the proof of Theorem 3.2, we've already shown that our assumptions imply 1, 2*, 3*,5* and 6; assumptions 7 and 8 hold by our assumption (A4). □

Proof of Lemma 4.1. The vector $\boldsymbol{Z}$ is generated by some $\boldsymbol{V} = (V_1, \ldots, V_r) \in [0, \infty)^r$ with $\mathbb{E}[V_s] = 1$ for $s \in \{1, \ldots, r\}$, so that $\boldsymbol{Z} = R\boldsymbol{V}$ and

$$\ell_{\boldsymbol{Z}}(\boldsymbol{z}) = \mathbb{E}\left[\max(z_1 V_1, \ldots, z_r V_r)\right],$$

see (2.3). Because

$$Y_j = \max_{s=1,\ldots,r} b_{js} Z_s = R \max_{s=1,\ldots,r} b_{js} V_s,$$

the vector $\boldsymbol{Y}$ is generated by $\boldsymbol{W} = (W_1, \ldots, W_d)$ where $W_j := \max_{s=1,\ldots,r} b_{js} V_s$. The margins of $\boldsymbol{Y}$ are unit Fréchet if $\mathbb{E}[W_j] = \ell_{\boldsymbol{Z}}(\boldsymbol{b}_j) = 1$ for all $j \in \{1, \ldots, d\}$. Its stable tail dependence function is

$$\begin{aligned}
\ell_{\boldsymbol{Y}}(\boldsymbol{y}) &= \mathbb{E}\left[\max_{j=1,\ldots,d} \frac{y_j}{\mathbb{E}[W_j]} \max_{s=1,\ldots,r} b_{js} V_s\right] = \mathbb{E}\left[\max_{s=1,\ldots,r} V_s \max_{j=1,\ldots,d} \frac{b_{js} y_j}{\mathbb{E}[W_j]}\right] \\
&= \ell_{\boldsymbol{Z}}\left(\max_{j=1,\ldots,d} \frac{b_{j1} y_j}{\ell_{\boldsymbol{Z}}(\boldsymbol{b}_j)}, \ldots, \max_{j=1,\ldots,d} \frac{b_{jr} y_j}{\ell_{\boldsymbol{Z}}(\boldsymbol{b}_j)}\right)
\end{aligned}$$

□

Acknowledgements

The author gratefully acknowledges support from the Fonds de la Recherche Scientifique (FNRS) and the Projet d'Actions de Recherche Concertées No. 12/17-045 of the Communauté française de Belgique. She would also like to thank Johan Segers for helpful comments and discussion.

REFERENCES

Andrews, D. W. K. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67(6):1341–1383.

Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3):683–734.

Andrews, D. W. K. (2002). Generalized method of moments estimation when a parameter is on a boundary. *Journal of Business & Economic Statistics*, 20(4):530–544.

Balkema, A. A. and De Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5):792–804.

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

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

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 53(2):377–392.

Cui, Q. and Zhang, Z. (2016). Max-linear competing factor models. *Journal of Business & Economic Statistics*, (just-accepted):1–39.

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

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: an Introduction*. Springer-Verlag Inc.

de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift fÃijr Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 40(4):317–337.

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

Einmahl, J. H., Kiriliouk, A., Krajina, A., and Segers, J. (2016). An M-estimator of spatial tail dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):275–298.

Einmahl, J. H., Kiriliouk, A., and Segers, J. (2017). A continuous updating weighted least squares estimator of tail dependence in high dimensions. To be published in Extremes. Available at `http://arxiv.org/abs/1601.04826`.

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

Embrechts, P., Lindskog, F., and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. In Rachev, S., editor, *Handbook of Heavy Tailed Distributions in Finance*, chapter 8, pages 329–384. Elsevier.

Falk, M., Hofmann, M., and Zott, M. (2015). On generalized max-linear models and their statistical interpolation. *Journal of Applied Probability*, 52(3):736–751.

Falk, M., Hüsler, J., and Reiss, R.-D. (2010). *Laws of small numbers: extremes and rare events*. Springer Science & Business Media.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press.

Fougères, A.-L., de Haan, L., and Mercadier, C. (2015). Bias correction in multivariate extremes. *The Annals of Statistics*, 43(2):903–934.

Gissibl, N. and Klüppelberg, C. (2015). Max-linear models on directed acyclic graphs. Available at `http://arxiv.org/abs/1512.07522`.

Gissibl, N., Klüppelberg, C., and Otto, M. (2017). Tail dependence of recursive max-linear models with regularly varying noise variables. Available at `https://arxiv.org/abs/1701.07351`.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, 22(3):423–453.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292):698–707.

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

Huser, R. and Genton, M. G. (2016). Non-stationary dependence structures for spatial extremes. *Journal of Agricultural, Biological, and Environmental Statistics*, pages 1–22.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *Annals of Probability*, 37(5):2042–2065.

Kereszturi, M. and Tawn, J. (2017). Properties of extremal dependence models built on bivariate max-linearity. *Journal of Multivariate Analysis*, 155:52–71.

Kiriliouk, A., Segers, J., and Tafakori, L. (2017). An estimator of the stable tail dependence function based on the empirical beta copula. To be submitted.

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

Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Ressel, P. (2013). Homogeneous distributions — and a spectral representation of classical mean values and stable tail dependence functions. *Journal of Multivariate Analysis*, 117(1):246–256.

Segers, J. (2012). Max-stable models for multivariate extremes. *REVSTAT — Statistical Journal*, 10(1):61–92.

Segers, J., Sibuya, M., and Tsukahara, H. (2017). The empirical beta copula. *Journal of Multivariate Analysis*, 155:35–51.

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

Strokorb, K., Schlather, M., et al. (2015). An exceptional max-stable process fully parameterized by its extremal coefficients. *Bernoulli*, 21(1):276–302.

Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, 77(2):245–253.

Wang, Y. and Stoev, S. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability*, 43(2):461–483.

Yuen, R. and Stoev, S. (2014). Upper bounds on value-at-risk for the maximum portfolio loss. *Extremes*, 17(4):585–614.