Spectral Density Ratio Models for Multivariate Extremes

Miguel de Carvalho,   Anthony C. Davison

Institute of Mathematics

Ecole Polytechnique Fédérale de Lausanne

Swiss Federal Institute of Technology

This Draft: January 18, 2011

Abstract

The joint modelling of extremal events is of increasing attention. This paper develops a semiparametric model for the situation where several multivariate extremal distributions are linked through the action of a covariate on a baseline distrbution. Empirical likelihood inference and estimation for this spectral density ratio model are discussed, and an application is given to pairs of extreme temperatures at different sites, under forest cover and in the open.

Keywords: Atmospheric Temperature; Empirical Likelihood; Exponential Tilt; Forest Microclimates; Multivariate Extreme Values; Spectral Distribution; Semiparametric Modelling.

1 Introduction

As human society becomes more complex, it becomes more vulnerable to rare but catastrophic events, such as the oil spill in the Gulf of Mexico or the continuing turbulence in the financial markets. Assessment of the risks of such events involves the estimation of small probabilities and hence entails extrapolation into the tails of multivariate distributions, often beyond any existing data. The mathematical basis for this is the statistics of multivariate extremes, which is a topic of current interest from both theoretical and applied viewpoints. Published applications include air quality monitoring (Heffernan and Tawn, 2004), wave surge analysis (Ramos and Ledford, 2009), aviation

1

safety (Einmahl *et al.*, 2009), precipitation studies (Jarušková , 2009), and finance (Embrechts *et al.*, 2009). A central concept in multivariate extreme value theory is the so-called spectral distribution, which determines the degree of dependence of the extremes of different variables. The spectral distribution must satisfy certain marginal moment constraints, so it is awkward both to devise suitable models and to find estimators that obey the constraints. A variety of parametric models have been proposed, for small to moderate numbers of dimensions (Tawn, 1988; Coles and Tawn, 1991; Kotz and Nadarajah, 2000, Chapter 3), and others are currently under development. Boldi and Davison (2007) introduced a constrained mixture of Dirichlet distributions which is weakly dense in the class of all possible spectral measures, and Einmahl and Segers (2009) recently proposed a nonparametric estimator that imposes the marginal constraints using empirical likelihood (Owen, 1988, 2001).

However, all these approaches consider only a single spectral distribution H, whereas in some settings it may be necessary to consider a family of spectral measures, $H_0, \ldots, H_{K-1}$, to capture the effect of covariates on joint extremes. In this paper, we discuss the joint modelling of extremal events when data are gathered from several samples to each of which corresponds a vector of covariates. The work is motivated by data on air temperatures under the forest canopy and in a nearby open field at several sites around Switzerland (Renaud and Rebetez, 2009). Each site is characterized by a covariate x_κ that summarises features such as its altitude, the soil type and dominant tree species, and the scientific questions require not merely the fitting of distributions to the extreme temperatures, but also the quantification of how their joint distribution depends on the site features. This is of interest for ecological reasons such as the shelter offered by different types of forest during heat waves (Ferrez *et al.*, 2011) and for the design of wooded parks in urban areas, which can offer relief during heatwaves.

A simple approach to the modelling of such data would be to fit parametric distributions to the individual subpopulations, and to allow the parameters to depend on the covariates using some form of regression model. We take a different approach, for several reasons. First, the number of flexible parametric models is limited, so that a good fit to all the subpopulations is not assured by any single model; moreover the interpretation of single parameters in terms of the distributional shape may be unclear. Second, in our application the focus of interest is on the effects of the covariates, and if possible we prefer to avoid having to choose a particular parametric model. Third, just as there are efficiency gains from using the threshold exceedance rather than the block maximum

approach to extremal modelling (Coles, 2001), it seems less wasteful of data to attempt to combine the models for the individual subpopulations. Our strategy allows us to estimate each of the spectral distribution functions H_κ using all K samples. Beyond the obvious efficiency gains, this borrowing of strength also allows improved estimation for subpopulations whose samples are too small to be individually informative about their tails.

The key idea is to build a semiparametric spectral density ratio model, wherein the spectral distribution functions are unspecified but related through a known weight function, such as an exponential tilt. Density ratio models have been applied in scenarios including logistic discrimination (Qin and Zhang, 1997), kernel density estimation (Fokianos, 2004), case-control studies (Kedem *et al.*, 2009) and instrumental variable methods (Cheng *et al.*, 2009). The model proposed here differs from these because of the imposition of marginal moment conditions, the inclusion of covariates, and the setting of rare event modelling.

2 Spectral density ratio model

2.1 Models for multivariate extremes

We begin with D-dimensional extreme value distributions specified in the classical one-sample framework. Let $Z_1, Z_2, \ldots$ be independent and identically distributed vectors of continuous random variables on $\mathbb{R}^D$ whose distribution function F lies in the joint domain of attraction of an extreme value distribution G. Without loss of generality, suppose that F has unit Fréchet marginal distributions, i.e., $\exp(-1/z)$, for $z > 0$. Pickands' (1981) representation theorem asserts that the limiting distribution of the standardised maximum $M_n = n^{-1} \max\{Z_1, \ldots, Z_n\}$ may be written as $G(z) = \exp\{-V(z)\}$, where

$$V(z) = D \int_{S_D} \max\{w_1/z_1, \ldots, w_D/z_D\} \, \mathrm{d}H(w).$$

Here H represents the so-called spectral measure defined on the unit simplex in $\mathbb{R}^D$, i.e., $S_D = \big\{w \in \mathbb{R}_+^D : \sum_{i=1}^D w_i = 1, \ w = (w_1, \ldots, w_D)\big\}$. Further connections between the exponential measure V and the spectral distribution function H can be found in Coles and Tawn (1991) or Beirlant *et al.* (2004, §8.2). The distribution H determines the interaction between joint extremes, but must

satisfy the normalization and moment conditions

$$\int_{S_D} \mathrm{d}H(w) = 1, \quad \int_{S_D} w \mathrm{d}H(w) = D^{-1}\mathbf{1}_D, \tag{1}$$

where $\mathbf{1}_D$ is the D-vector of ones.

A pseudo-polar transformation is helpful in understanding H: if we map the joint extremal data $e_1, \ldots, e_D$, which have unit Fréchet marginal distributions, into the pseudo-angular coordinates $w_1 = e_1/r, \ldots, w_D = e_D/r$, with pseudo-radius $r = e_1 + \cdots + e_D$, then the directional part $w_1, \ldots, w_D$ has measure H. The limiting cases of independence and dependence in the two-dimensional case illustrate this: if extreme values of the two variables tend to occur individually then the mass of H is concentrated close to the limits 0 and 1, whereas if the extremes tend to occur together then H places most of its mass near $w = 1/2$. This pseudo-polar representation also arises in a point process characterization. As $n \to \infty$, the counting process of the rescaled standard Fréchet observations, $N_n = \{n^{-1}Z_i : i = 1, \ldots, n\}$, converges in distribution to a Poisson process whose intensity measure on $\mathbb{R}_+^D$ factorizes as

$$\Lambda(dz) = \frac{\mathrm{d}r}{r^2} \times D \ \mathrm{d}H(w), \quad r > 0, w \in S_D.$$

These pseudo-polar representations generalise to the K-sample case. Suppose that we have independent sets of observations $\{w_{\kappa 1}, \ldots, w_{\kappa n_\kappa}\}$ from K unknown spectral distributions $(H_0, \ldots, H_{K-1})$. The measures must satisfy the normalization and moment constraints

$$\begin{cases} \int_{S_D} \mathrm{d}H_0(w) = 1, & \int_{S_D} w \mathrm{d}H_0(w) = D^{-1}\mathbf{1}_D, \\ \int_{S_D} \mathrm{d}H_1(w) = 1, & \int_{S_D} w \mathrm{d}H_1(w) = D^{-1}\mathbf{1}_D, \\ \quad \vdots & \quad \vdots \\ \int_{S_D} \mathrm{d}H_{K-1}(w) = 1, & \int_{S_D} w \mathrm{d}H_{K-1}(w) = D^{-1}\mathbf{1}_D. \end{cases} \tag{2}$$

A similar point process representation can be given for the K-sample case by replacing the univariate point process with a multivariate point process. For each sample κ, we now suppose that $Z_{1\kappa}, Z_{2\kappa}, \ldots$ are sequences of independent and identically distributed vectors on $\mathbb{R}^D$, with a distribution function F_κ in the joint domain of attraction of G. The counting process for the K-sample problem may be denoted by $\mathcal{N}_{\mathbf{n}} = (N_{n_0}, \ldots, N_{n_{K-1}})$, where $\mathbf{n} = (n_0, \ldots, n_{K-1})$ and

$$N_\kappa = \{n_\kappa^{-1} Z_{i\kappa} : i = 1, \ldots, n_\kappa\}, \quad \kappa = 0, \ldots, K - 1.$$

The process $\mathcal{N}_{\mathbf{n}}$ converges to a multivariate Poisson process on $\mathbb{R}_+^{KD}$ with intensity process $\mathbf{\Lambda} = (\Lambda_0, \ldots, \Lambda_{K-1})$, as $\min\{n_0, \ldots, n_{K-1}\} \to \infty$, where each component of the intensity process factorizes along the radial and directional parts, respective to each sample, viz.:

$$\Lambda_\kappa(dz) = \frac{\mathrm{d}r_\kappa}{r_\kappa^2} \times D \ \mathrm{d}H_\kappa(w), \quad \kappa = 0, \ldots, K-1, \qquad r_\kappa > 0, w \in S_D.$$

The next section introduces the spectral density ratio model for modelling K-sample multivariate extremes. Just as the point process characterization for one-sample multivariate extremes avoids the wastefulness of data implied by block maximum strategies, our model enables more effective use of the available data in the K-sample framework.

2.2   Introducing the model

Later we focus our attention on the simplex S_D on which spectral measures are defined, but for now we let H_κ denote any distribution function. Our interest lies in the measures $(H_0, \ldots, H_{K-1})$, which are linked through a positive function g_κ with known functional form

$$\left\{ (H_0, \ldots, H_{K-1}) : \frac{\mathrm{d}H_\kappa(w)}{\mathrm{d}H_0(w)} = g(w, \gamma_\kappa), \text{ for some } g(w, \gamma_\kappa) > 0; \ \kappa = 0, \ldots, K-1 \right\}. \tag{3}$$

Here $\gamma = (\gamma_1, \ldots, \gamma_{K-1})^{\mathrm{T}}$ represents a q-vector of parameters, and we set $g(w, \gamma_0) = 1$ for identifiability. This specification is common to many models (Efron and Tibshirani, 1996; Qin and Zhang, 1997; Fokianos et al., 2001; Fokianos, 2004; Cheng et al., 2009). Under (3) the distributions H_κ are left unspecified but related through a known weight function. The measure H_0 acts as a reference from which the other $K-1$ measures are obtained, through a distortion controlled by g and γ_κ. Some examples are given below, where we take the two-sample case for ease of notation. Let $g(w; \alpha, \beta) = \exp\{\alpha + \beta c(w)\}$, where α is a scale parameter, β a is $p \times 1$ vector parameter, and $c(w)$ is a known distortion function, such as $c(w) = w$. We then obtain a model where the log-ratio of the densities is linear in the parameters, $\log\{\mathrm{d}H_1(w)/\mathrm{d}H_0(w)\} = \alpha + \beta w$. The logistic regression model can be derived by applying Bayes' theorem, and the multinomial logistic regression model arises for $p > 1$. A property of such models known as independence of irrelevant alternatives implies that the inference is independent of the baseline $\mathrm{d}H_0$ (Fokianos, 2004). This type of semiparametric approach provides gains in efficiency by estimating each density $\mathrm{d}H_\kappa$ on the basis of the entire sample (Gilbert et al., 1999; Fokianos, 2004; Kedem et al., 2009). Specification (3) can also be used to construct density estimates, by taking $g(w; \alpha, \beta) = \exp\{\alpha + \beta s(w)\}$, with $s(w)$ denoting a vector

of sufficient statistics. Efron and Tibshirani (1996) used this representation to estimate a density $\mathrm{d}H_1/\mathrm{d}w$ on the basis of a carrier density $\mathrm{d}H_0/\mathrm{d}w$ obtained by nonparametric kernel procedures, the idea being that $\mathrm{d}H_0$ should control local adaptation to the data, while the exponential term $\exp\{\alpha + \beta s(w)\}$ should capture global features.

Specification (3) also turns out to be natural for modelling the K-sample multivariate extreme value framework discussed in §2.1, with the constraints (2) restated as

$$
\begin{cases}
\int_{S_D} g(w, \gamma_0)\mathrm{d}H_0(w) = 1, & \int_{S_D} wg(w, \gamma_0)\mathrm{d}H_0(w) = D^{-1}\mathbf{1}_D, \\
\int_{S_D} g(w, \gamma_1)\mathrm{d}H_0(w) = 1, & \int_{S_D} wg(w, \gamma_1)\mathrm{d}H_0(w) = D^{-1}\mathbf{1}_D, \\
\quad\vdots & \quad\vdots \\
\int_{S_D} g(w, \gamma_{K-1})\mathrm{d}H_0(w) = 1, & \int_{S_D} wg(w, \gamma_{K-1})\mathrm{d}H_0(w) = D^{-1}\mathbf{1}_D.
\end{cases}
\tag{4}
$$

We suppress the dependence of α on β, and vice versa, but it should be noted in what follows that the normalization constraints in (4) force these parameters to be associated, and in particular if $p = 1$ they need to be perfectly correlated. Below we refer to the general semiparametric setting (3), subject to the normalization and marginal moment constraints (4), as the spectral density ratio model. We propose to fit it through empirical likelihood methods (Owen, 1988, 2001).

2.3 Estimation

Let $v = \{v_1, \ldots, v_n\}$ denote the combined sample $\{w_{01}, \ldots, w_{0n_0}, \ldots, w_{(K-1)1}, \ldots, w_{(K-1)n_{K-1}}\}$ from all K unknown spectral distributions H_κ. The likelihood of the K-sample multivariate extreme value problem under (3) is

$$
\mathscr{L}(\gamma, H_0) = \prod_{k=0}^{K-1} \prod_{j=1}^{n_k} \mathrm{d}H_k(w_{kj}) = \prod_{i=1}^{n} p_i \prod_{k=1}^{K-1} \prod_{j=1}^{n_k} g(w_{kj}, \gamma_k),
$$

where $p_i = \mathrm{d}H_0(v_i) = H_0(v_i^+) - H_0(v_i^-)$ denotes the size of the jump of the baseline spectral distribution function at the observed v_i.

We restrict our attention to the tilting function $g(w, \gamma_\kappa) = \exp\{\alpha_\kappa + \beta_\kappa c(w)\}$, where $\gamma_\kappa = (\alpha_\kappa, \beta_\kappa)$; for identifiability we set $\alpha_0 = \beta_0 = 0$. The loglikelihood is thus

$$
\ell(\gamma, H_0) = \sum_{i=1}^{n} \log(p_i) + \sum_{k=1}^{K-1} \sum_{j=1}^{n_k} \{\alpha_k + \beta_k c(w_{kj})\}.
\tag{5}
$$

6

Empirical likelihood estimation of the spectral density ratio model involves maximizing ℓ with respect to p_i, for a fixed γ, subject to the empirical versions of constraints (4), conveniently rewritten as

$$p_i \geq 0, \qquad v_i \in S_D$$

$$\sum_{i=1}^{n} p_i = 1, \qquad \sum_{i=1}^{n} p_i \left\{ v_i - D^{-1}\mathbf{1}_D \right\} = 0,$$

$$\sum_{i=1}^{n} p_i \{ g(v_i, \gamma_1) - 1 \} = 0, \qquad \sum_{i=1}^{n} p_i \{ v_i g(v_i, \gamma_1) - D^{-1}\mathbf{1}_D \} = 0, \qquad (6)$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$\sum_{i=1}^{n} p_i \{ g(v_i, \gamma_{K-1}) - 1 \} = 0, \qquad \sum_{i=1}^{n} p_i \{ v_i g(v_i, \gamma_{K-1}) - D^{-1}\mathbf{1}_D \} = 0.$$

Using an approach similar to that of Qin and Lawless (1994), it is shown in Appendix A that if we use Lagrange multiplier procedures to profile p_i with the normalization and marginal moment constraints (6), then the jump size for the baseline spectral distribution function can be written as

$$p_i = \frac{1}{n_0} \frac{1}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}} \mathcal{M}(v_i, \gamma)}, \qquad (7)$$

where $\rho_\kappa = n_\kappa / n_0$,

$$\mathcal{M}(v, \gamma) = \mathcal{G}(v, \gamma) \otimes (v - D^{-1}\mathbf{1}_D), \qquad (8)$$

and $\mathcal{G}(v, \gamma) = (1, g_1(v, \gamma_1), \ldots, g_{K-1}(v, \gamma_{K-1}))^{\mathrm{T}}$. Here and below, $\delta = (\delta_0, \ldots, \delta_{K-1})^{\mathrm{T}}$ denotes the Lagrange multipliers corresponding to the marginal moment constraints, which are determined through the conditions

$$\frac{1}{n_0} \sum_{i=1}^{n} \frac{v_i g(v_i, \gamma_\kappa) - D^{-1}\mathbf{1}_D}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}} \mathcal{M}(v_i, \gamma)} = 0, \quad \kappa = 0, \ldots, K - 1. \qquad (9)$$

Thus apart from a constant the profiled empirical loglikelihood for γ can be written as

$$\ell_{\mathrm{p}}(\gamma) = -\sum_{i=1}^{n} \log \left\{ 1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}} \mathcal{M}(v_i, \gamma) \right\} + \sum_{k=0}^{K-1} \sum_{j=0}^{n_k} \{ \alpha_k + \beta_k c(w_{kj}) \}, \qquad (10)$$

and so the semiparametric empirical likelihood estimator $\widehat{\gamma} = \arg\max_{\gamma} \ell_{\mathrm{p}}(\gamma)$ of the spectral density

ratio model can be obtained by combining (9) with the score equations

$$
\begin{aligned}
\frac{\partial \ell_{\mathrm{p}}}{\partial \alpha_\kappa} &= -\sum_{i=1}^{n} \frac{\rho_\kappa g(v_i, \gamma_\kappa) + \delta^{\mathrm{T}} M_{\alpha_\kappa}(v_i, \gamma)}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}} \mathcal{M}(v_i, \gamma)} + n_\kappa = 0, \\
\frac{\partial \ell_{\mathrm{p}}}{\partial \beta_\kappa} &= -\sum_{i=1}^{n} \frac{\rho_\kappa c(w_{\kappa j}) g(v_i, \gamma_\kappa) + \delta^{\mathrm{T}} M_{\beta_\kappa}(v_i, \gamma)}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}} \mathcal{M}(v_i, \gamma)} + \sum_{j=0}^{n_\kappa} c(w_{\kappa j}) = 0, \quad \kappa = 0, \dots, K-1.
\end{aligned}
\tag{11}
$$

Here, M_{α_κ} and M_{β_κ} represent the partial derivatives of the moment estimating function (8); for explict expressions see (22). The existence of the semiparametric likelihood estimates $\widehat{\gamma}$ is a corollary of Lemma 1 in Qin and Lawless (1994) which states that under suitable regularity conditions the semiparametric empirical likelihood estimator lies in the interior of the ball $||\gamma - \gamma_o|| \leq n^{-1/3}$, and thus is centred on the true value γ_o, as $n \to \infty$.

On using the estimates obtained from the score equations (11), the maximum likelihood estimator of the size of the jump of the baseline spectral density function turns out to be

$$
\widehat{p}_i = \frac{1}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \widehat{\gamma}_k) + \widehat{\delta}^{\mathrm{T}} \mathcal{M}(v_i, \widehat{\gamma})},
$$

so the maximum likelihood estimator of the baseline spectral distribution function is

$$
\widehat{H}_0(w) = \sum_{i=1}^{n} \widehat{p}_i I(v_i \leq w) = \frac{1}{n_0} \sum_{i=1}^{n} \frac{1}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \widehat{\gamma}_k) + \widehat{\delta}^{\mathrm{T}} \mathcal{M}(v_i, \widehat{\gamma})} I(v_i \leq w),
\tag{12}
$$

and for $\kappa = 1, \dots, K-1$ the other spectral spectral distribution functions are estimated as

$$
\widehat{H}_\kappa(w) = \sum_{i=1}^{n} \widehat{p}_i g(v_i, \widehat{\gamma}_k) I(v_i \leq w) = \frac{1}{n_0} \sum_{i=1}^{n} \frac{g(v_i, \widehat{\gamma}_k)}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \widehat{\gamma}_k) + \widehat{\delta}^{\mathrm{T}} \mathcal{M}(v_i, \widehat{\gamma})} I(v_i \leq w).
\tag{13}
$$

2.4 Inference

The following result states the asymptotic normality of the semiparametric empirical likelihood estimator and a version of Wilks' theorem for the spectral density ratio model.

Theorem 1 *Consider the following estimating equation representation of the profile empirical likelihood as a function of the baseline spectral distribution function*

$$
\sup \left\{ \prod_{i=1}^{n} p_i \prod_{k=1}^{K-1} \prod_{j=1}^{n_k} g(w_{kj}, \gamma_k) : p_i \geq 0, \ v_i \in S_D, \ \sum_{i=1}^{n} p_i \mathcal{E}(v_i, \gamma) = 0 \right\},
$$

where

$$
\mathcal{E}(v, \gamma) = \begin{pmatrix} \mathcal{G}(v, \gamma) - \mathbf{1}_K \\ \mathcal{G}(v, \gamma) \otimes (v - D^{-1} \mathbf{1}_D) \end{pmatrix}.
$$

Suppose that

i) $\partial\mathcal{E}(v,\gamma)/\partial\gamma$ and $\partial^2\mathcal{E}(v,\gamma)/\partial\gamma\partial\gamma^{\mathrm{T}}$ *are continuous over some neighbourhood of the true parameter* γ_o;

ii) $||\partial\mathcal{E}(v,\gamma)/\partial\gamma||$, $||\partial^2\mathcal{E}(v,\gamma)/\partial\gamma\partial\gamma^{\mathrm{T}}||$ *and* $||\mathcal{E}(v,\gamma)||^3$ *are bounded by some function* $B(v)$ *integrable with respect to the baseline spectral measure* H_0 *over this neighbourhood;*

iii) $E\{\partial\mathcal{E}(v,\gamma)/\partial\gamma\}$ *has rank* q; *and*

iv) $E\{\mathcal{E}(v,\gamma)\big(\mathcal{E}(v,\gamma)\big)^{\mathrm{T}}\}$ *is positive definite.*

Then

1. *the semiparametric empirical likelihood estimator,* $\widehat{\gamma} = \arg\max_\gamma \ell_{\mathrm{p}}(\gamma)$, *satisfies*

$$n^{1/2}(\widehat{\gamma} - \gamma_o) \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, \Sigma),$$

   where $\overset{\mathcal{L}}{\longrightarrow}$ *denotes convergence in distribution and* Σ *is defined in* (23); *and*

2. *if* $\gamma^0 = \big((a_1, b_1), \ldots, (a_{K-1}, b_{K-1})\big)$ *with* $a_k, b_k \neq 0$, *for* $k = 1, \ldots, K-1$, *then under* $\mathrm{H}_0 : \gamma = \gamma^0$, *we have*

$$2\left\{\ell_{\mathrm{p}}(\widehat{\gamma}) - \ell_{\mathrm{p}}(\gamma^0)\right\} \overset{\mathcal{L}}{\longrightarrow} \chi^2_{q/2+1}.$$

The following corollary characterizes the limiting behavior of the empirical likelihood ratio for other hypotheses.

Corollary 2 *Under the regularity conditions of Theorem 1 the following large sample results hold.*

1. *Let* $\gamma = (\gamma_{\mathrm{I}}, \gamma_{\mathrm{II}})$ *where* γ_{I} *and* γ_{II} *are* $m\times 1$ *and* $(q-m)\times 1$ *vectors. Let* $\gamma_{\mathrm{I}}^0 = \big((a_1, b_1), \ldots, (a_m, b_m)\big)$ *with* $a_k, b_k \neq 0$, *for* $k = 1, \ldots, m$; *under* $\mathrm{H}_0 : \gamma_{\mathrm{I}} = \gamma_{\mathrm{I}}^0$ *we have*

$$2\left\{\ell_{\mathrm{p}}(\widehat{\gamma}_{\mathrm{I}}, \widehat{\gamma}_{\mathrm{II}}) - \ell_{\mathrm{p}}(\gamma_{\mathrm{I}}^0, \widehat{\gamma}_{\mathrm{II}}^0)\right\} \overset{\mathcal{L}}{\longrightarrow} \chi^2_{m/2+1},$$

   where $\widehat{\gamma}_{\mathrm{II}}^0$ *maximizes* $\ell_{\mathrm{p}}(\gamma_{\mathrm{I}}^0, \gamma_{\mathrm{II}})$ *with respect to* γ_{II}.

2. *Let* $\gamma = (\gamma_{\mathrm{I}}, \gamma_{\mathrm{II}})$ *where* γ_{I} *and* γ_{II} *are* $m\times 1$ *and* $(q-m)\times 1$ *vectors. Let* $\gamma_{\mathrm{I}}^0 = \big((a_1, b_1), \ldots, (a_m, b_m)\big)$ *with* $a_k, b_k = 0$, *for* $k = 1, \ldots, m$; *under* $\mathrm{H}_0 : \gamma_{\mathrm{I}} = \gamma_{\mathrm{I}}^0$ *we have*

$$2\left\{\ell_{\mathrm{p}}(\widehat{\gamma}_{\mathrm{I}}, \widehat{\gamma}_{\mathrm{II}}) - \ell_{\mathrm{p}}(\gamma_{\mathrm{I}}^0, \widehat{\gamma}_{\mathrm{II}}^0)\right\} \overset{\mathcal{L}}{\longrightarrow} \chi^2_{m/2},$$

   where $\widehat{\gamma}_{\mathrm{II}}^0$ *maximizes* $\ell_{\mathrm{p}}(\gamma_{\mathrm{I}}^0, \gamma_{\mathrm{II}})$ *with respect to* γ_{II}.

A benchmark to verify whether the specification of the weight function g is adequate can be based on the one-sample empirical likelihood estimate $\ddot{H}_\kappa$ (Einmahl and Segers, 2009). Following Qin and Zhang (1997), we suggest using the statistic

$$\Delta_\kappa = n^{1/2}||\widehat{H}_\kappa - \ddot{H}_\kappa||_\infty = n^{1/2} \sup_{0 < w < 1} |\widehat{H}_\kappa(w) - \ddot{H}_\kappa(w)|, \quad \kappa = 0, \ldots, K - 1, \tag{14}$$

a p-value for which can be estimated using the bootstrap, as follows. Generate bootstrap samples $\{w_{\kappa 1}^{\star(b)}, \ldots, w_{\kappa n_\kappa}^{\star(b)}\}$ from $g(v, \widehat{\gamma}_\kappa)\mathrm{d}\widehat{H}_0$, and use the combined bootstrap sample $v^{\star(b)} = \{v_1^{\star(b)}, \ldots, v_n^{\star(b)}\}$ to compute

$$\Delta_\kappa^{\star(b)} = n^{1/2} \sup\{|\widehat{H}_\kappa(w) - \ddot{H}_\kappa(w)| : w \in v^{\star(b)}\}, \quad \kappa = 0, \ldots, K - 1, \; b = 1, \ldots, B. \tag{15}$$

We then compare the B bootstrapped statistics $\Delta^{\star(b)}$, with the values of (14) obtained with the observed combined sample v, i.e., $\Delta_\kappa^{\mathrm{obs}} = \sup\{|\widehat{H}_\kappa(w) - \ddot{H}_\kappa(w)| : w \in v\}$; the p-value for the goodness of fit test is then

$$\widehat{P}_\kappa = B^{-1} \sum_{b=1}^{B} I(\Delta_\kappa^{\star(b)} \geq \Delta_\kappa^{\mathrm{obs}}), \quad \kappa = 0, \ldots, K - 1.$$

2.5   Covariates

To avoid complex model selection problems, we introduce covariate modelling only in a second stage. If covariates had been introduced earlier, it could be troublesome to identify whether poor fit of a model was due to the tilting function or to the covariate structure. Hence, after empirical likelihood inference for the tilt parameters has been conducted we study their association with p covariates of interest stored in a $(K \times p)$ design matrix $X = (x_0^{\mathrm{T}}, \ldots, x_{K-1}^{\mathrm{T}})^{\mathrm{T}}$. Motivated by the asymptotic normality of the semiparametric empirical likelihood estimator $\widehat{\gamma}$, we use a simple regression model wherein only $\widehat{\beta} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_{K-1})^{\mathrm{T}} \sim G$ is considered, so that

$$E(\widehat{\beta} \mid X) = X\theta. \tag{16}$$

Using the sample data $\{(x_k^{\mathrm{T}}, \widehat{\beta}_k)\}_{k=0}^{K-1}$, we propose to assess the effect of each of the p covariates on the extremal dependence by conducting inference over $\theta = (\theta_1, \ldots, \theta_p)^{\mathrm{T}}$. Consider the auxiliary variable $\mathcal{Z}_\kappa(\theta) = x_\kappa^{\mathrm{T}}(\widehat{\beta}_\kappa - x_\kappa^{\mathrm{T}}\theta)$ and let $p_\kappa = \mathrm{d}G(\widehat{\beta}_i)$. The profile empirical likelihood of θ is

$$\ell_{\mathrm{p}}(\theta) = \sup\left\{\prod_{k=0}^{K-1} p_k : p_k \geq 0, \; \sum_{k=0}^{K-1} p_k = 1, \; \sum_{k=0}^{K-1} p_k \mathcal{Z}_k(\theta) = 0\right\}.$$

10

We choose model (16) for its simplicity but more general models could have been used to describe the degree of association between extremal dependence, as measured by the tilt parameters, and the covariates of interest. Inference for the maximum empirical likelihood estimator $\widetilde{\theta} = \arg\max_\theta \ell_{\mathrm{p}}(\theta)$ can then be performed using a Wilks' theorem, which suggests a χ^2 calibration with coverage error of order $O(n^{-1})$ (Owen, 1991), though for small samples bootstrap calibration is preferable. If similar reasoning is applied to $\widehat{\alpha} = (\widehat{\alpha}_0, \dots, \widehat{\alpha}_{K-1})^{\mathrm{T}}$ we obtain the fitted parameters $\widetilde{\gamma} = (\widetilde{\alpha}, \widetilde{\beta})$.

3   Simulation study

3.1   A K-sample bivariate spectral density ratio model

To model the spectral density corresponding to each subpopulation, we consider a family of symmetric Beta distributions,

$$\mathrm{d}H_\kappa(w) = \frac{1}{B(\phi_\kappa)} w^{\phi_\kappa - 1}(1 - w)^{\phi_\kappa - 1}\mathrm{d}w, \quad \phi_\kappa > 0, \kappa = 0, \dots, K - 1, \tag{17}$$

where $B(\phi) = \int_0^1 \{u(1-u)\}^{\phi-1}\mathrm{d}u$. The mean for each spectral density equals $1/2$, so that constraints (4) are satisfied. Using the distortion function $c(w) = \log\{w(1 - w)\}$ we can rewrite (17) as

$$\begin{cases} \mathrm{d}H_\kappa(w) & = \exp\{a_\kappa + b_\kappa c(w)\}\mathrm{d}w, \\ (a_\kappa, b_\kappa) & = (-\log B(\phi_\kappa), \phi_\kappa - 1), \quad \kappa = 0, \dots, K - 1. \end{cases}$$

Hence, making use of (17), we obtain the following spectral density representation of the K-sample bivariate extreme value beta model,

$$\frac{\mathrm{d}H_\kappa(w)}{\mathrm{d}H_0(w)} = \exp\left\{\alpha_\kappa + \beta_\kappa c(w)\right\}, \tag{18}$$

where the tilt parameters are

$$(\alpha_\kappa, \beta_\kappa) = (\log\left\{B(\phi_0)/B(\phi_\kappa)\right\}, \phi_\kappa - \phi_0), \quad \kappa = 0, \dots, K - 1. \tag{19}$$

A consequence of (19) is that $(\alpha_0, \beta_0) = (0, 0)$, so this parametrization of the spectral density ratio model (18) is identifiable. This model is closed: tilting always produces a symmetric beta distribution.

11

3.2 Numerical exercise

We now report computational experience with the model described in §3.1. We use 1000 simulated data sets with $K = 3$ and $n = 180$, such that $n_0 = n_1 = n_2 = 60$, chosen for comparability with the data in §4. Here we consider $\phi_0 = 0.5$, $\phi_1 = 1$ and $\phi_2 = 2$, but our conclusions also hold for other parameter values. Given the computational cost of obtaining full optimisation estimates, in practice we have found it best to use a two-step strategy wherein one maximizes the unconstrained outer objective function (5) with respect to γ, and minimizes the inner dual problem with respect to the nuisance parameter δ. For large n the inner dual problem not only has much lower dimensionality than the corresponding primal problem, but it also has the advantage of being subject to a set of linear constraints that can be removed by using a pseudo-logarithmic function (Owen, 2001, p. 62). For a representation of the inner dual optimisation problem and further computational remarks, see Appendix B. A similar two-step estimation strategy is used by Chauduri *et al.* (2007), and a more general discussion on two-step nearly optimal estimators which asymptotically behave like M-estimators can be found in van der Vaart (1998, p. 71).

Table 1 shows a summary of the estimates; the true values of the tilting parameters were computed using (19). The estimates of the exponential tilt parameters and the Lagrange multipliers are on average close to their true values, but the latter are much more variable. Table 1 also shows that the estimates from the two-step method are on average close to the average full optimisation estimates. QQ-plots of the estimates, not shown here, confirm the approximate normality of the tilt parameters suggested by Theorem 1. To illustrate the asymptotic results obtained in the previous section in each these simulations we tested the null hypothesis that the baseline sample and sample 1 have the same form of extremal dependence as the baseline sample, i.e., $H_0 : \alpha_1 = \beta_1 = 0$. The rejection rate for comparing the empirical likelihood ratio statistic with the 95% quantile of a χ_1^2 distribution was 78.7%, suggesting that a bootstrap approach would be preferable.

Figure 1 compares semiparametric empirical likelihood estimates of the spectral measures with those based on the empirical likelihood estimate of Einmahl and Segers (2009), for a run with $(\phi_0, \phi_1, \phi_2) = (0.1, 1, 5)$ and $(n_0, n_1, n_2) = (25, 50, 25)$. This is representative of a situation where one of the samples has more extremes, from which the others can borrow strength.
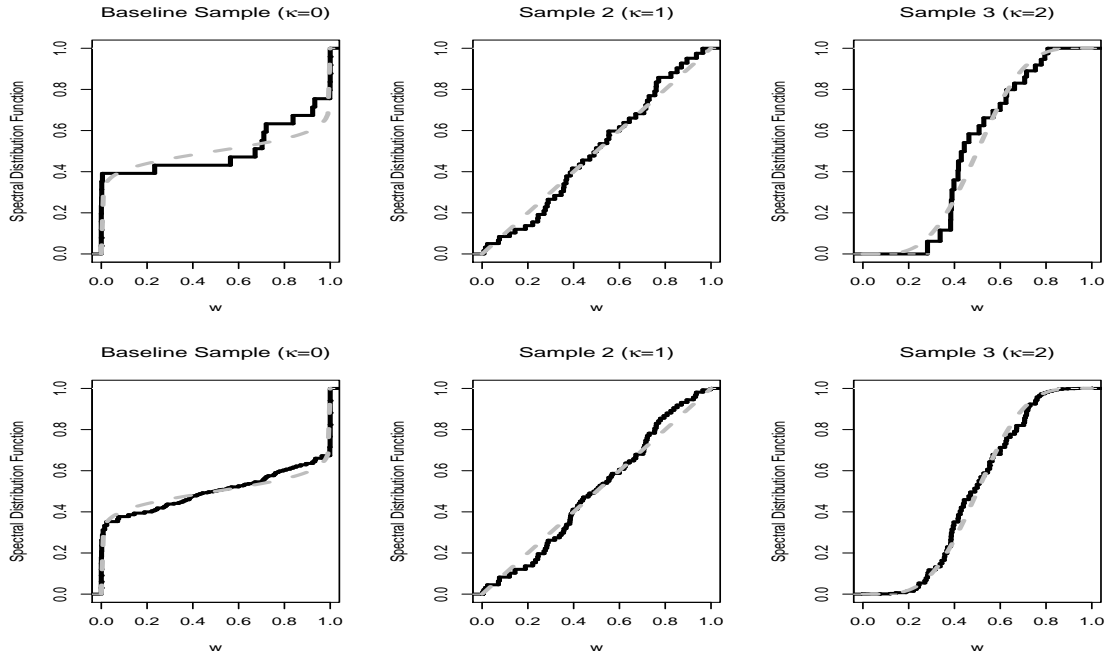
Figure 1: Comparison of the spectral distribution functions estimated by empirical likelihood (above) and using the spectral density ratio model (below); the dashed lines represent the true distribution functions.

Table 1: Estimates of the tilting parameters and the Lagrange multipliers, based on 1000 simulations; RMSE stands for root mean square error.

Outputs for the tilting parameters	κ		
	0	1	2
True values	$(0,0)$	$(1.14, 0.50)$	$(2.94, 1.50)$
Average full optimisation estimates	$(0,0)$	$(1.19, 0.53)$	$(2.96, 1.52)$
RMSE	0	$(0.52, 0.24)$	$(0.81, 0.45)$
Average two-step estimates	$(0,0)$	$(1.21, 0.53)$	$(3.08, 1.59)$
RMSE	0	$(0.46, 0.21)$	$(0.76, 0.43)$
Outputs for the Lagrange multipliers	κ		
	0	1	2
True values	0	0	0
Average full optimisation estimate	0.13	-0.39	0.23
RMSE	6.08	9.10	6.79
Average two-step estimates	0.10	-0.27	0.13
RMSE	5.92	9.02	6.68

4 Extreme temperature analysis case study

In this section we describe an application to modelling the dependence between extreme air temperatures under the forest canopy and in a nearby open field at 14 sites. We took the data from the Long-term Forest Ecosystem Research database maintained by LWF (Langfristige Waldökosystem-Forschung). Further information may be found at

```
http://www.wsl.ch/forschung/forschungsunits/wald
```

Temperatures are measured two meters above ground in circular metal shelters, and have been collected at every ten minutes since 1997 (Ferrez *et al.*, 2011). We take daily maxima of the residual series that result from removal of the annual cycle in both location and scale. After thresholding the residuals of each sample at its 98% quantile, we reduce the initial 38,923 observations to $n = 785$ pairs of residuals, one of the pair being under the forest canopy and the other being in the open. We treat the pairs as mutually independent, and investigate how the dependence within them depends on characteristics of the 14 samples, focusing here on the effect of slope.

The estimates of the tilting parameters when our model is applied using the distortion function $c(w) = \log\{w(1-w)\}$ are given in Table 2; the baseline sample is Beatenberg. A negative estimated tilt parameter corresponds to lower extremal dependence than in the baseline sample. For example, since $(\widehat{\alpha}_1, \widehat{\beta}_1) = (-1.26, -0.72)$, Bettlachstock is interpreted as having weaker dependence of extremes during heat waves than Beatenberg, though the standard errors suggest that this is not a significant effect. All forests with negative $\widehat{\alpha}_\kappa$ and $\widehat{\beta}_\kappa$ display weaker dependence of extremes than the baseline. The estimated spectral distributions are plotted in Figure 2. The more positive the estimated tilt parameters, the more the distribution concentrates at $1/2$. Two forests with much larger estimates, Isole and Jussy, correspond to more 'S-shaped' spectral distributions. To evaluate the fit of the spectral density ratio model, 1000 bootstrap samples were generated and used to compute (15). All p-values suggest that the spectral density ratio model is appropriate for the data; the smallest p-value is 0.17 for Schänis.

A plot of tilting parameters against slope in Figure 3 suggests that, with the exception of Isone and Jussy, there may be a mild quadratic effect of slope on extremal dependence; this was also suggested by Ferrez *et al.* (2011), based on a parametric analysis. We assess the evidence for this

15

Table 2: Characteristics of sites, estimated tilting parameters and standard errors. Beatenberg is the baseline station

Location	No.	Slope	Estimates		Asymptotic S.E.	
κ	n_κ	(%)	$\widehat{\alpha}_\kappa$	$\widehat{\beta}_\kappa$	$\widehat{\Sigma}_{\widehat{\alpha}_\kappa}$	$\widehat{\Sigma}_{\widehat{\beta}_\kappa}$
Beatenberg	57	33	–	–	–	–
Bettlachstock	54	66	-1.26	-0.72	0.75	0.42
Celerina	53	34	-1.94	-1.08	0.70	0.39
Chironico	45	35	-1.41	-0.80	0.76	0.42
Isone	45	58	8.78	5.78	2.35	1.59
Jussy	62	3	5.07	3.26	1.48	0.97
Lausanne	63	7	0.07	0.04	0.83	0.49
National Park	59	11	1.47	0.91	1.00	0.61
Neunkirch	42	58	-2.00	-1.11	0.71	0.39
Novaggio	65	68	-0.26	-0.15	0.80	0.46
Othmarsingen	57	27	-1.65	-0.93	0.71	0.40
Schänis	58	60	-1.49	-0.85	0.72	0.40
Visp	64	80	0.11	0.06	0.83	0.49
Vordemwald	61	14	-2.26	-1.23	0.68	0.38

Beatenberg: $(\alpha_0, \beta_0) = (0,0)$

Spectral Distribution Function

w
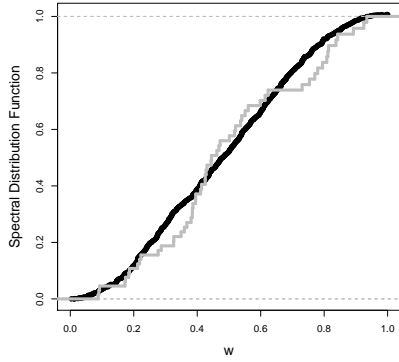
Bettlachstock: $(\alpha_1, \beta_1) = (-1.26, -0.72)$
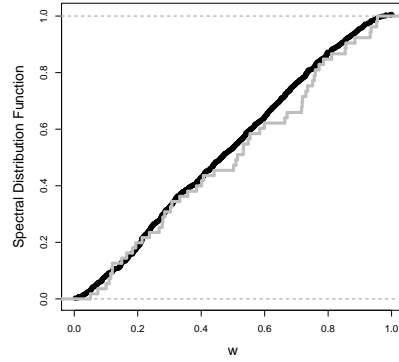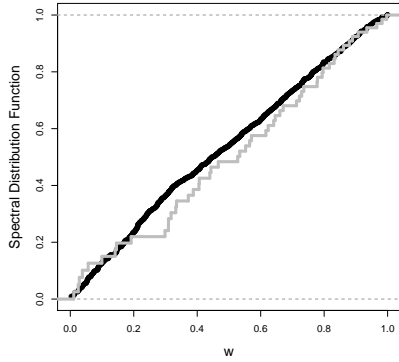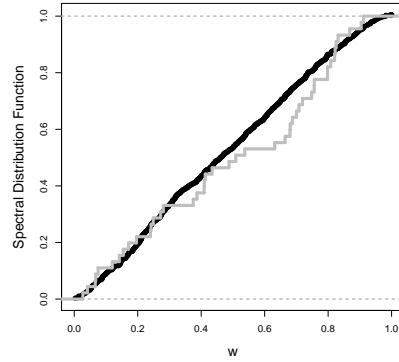
Spectral Distribution Function

w

Celerina: $(\alpha_2, \beta_2) = (-1.94, -1.08)$

Spectral Distribution Function

w

Chironico: $(\alpha_3, \beta_3) = (-1.41, -0.80)$

Spectral Distribution Function

w

Isone: $(\alpha_4, \beta_4) = (8.78, 5.78)$

Spectral Distribution Function

w

Jussy: $(\alpha_5, \beta_5) = (5.07, 3.26)$

Spectral Distribution Function

w

Lausanne: $(\alpha_6, \beta_6) = (0.07, 0.04)$

Spectral Distribution Function

$\hat{P}_6 = 1.00$

w

National Park: $(\alpha_7, \beta_7) = (1.47, 0.91)$
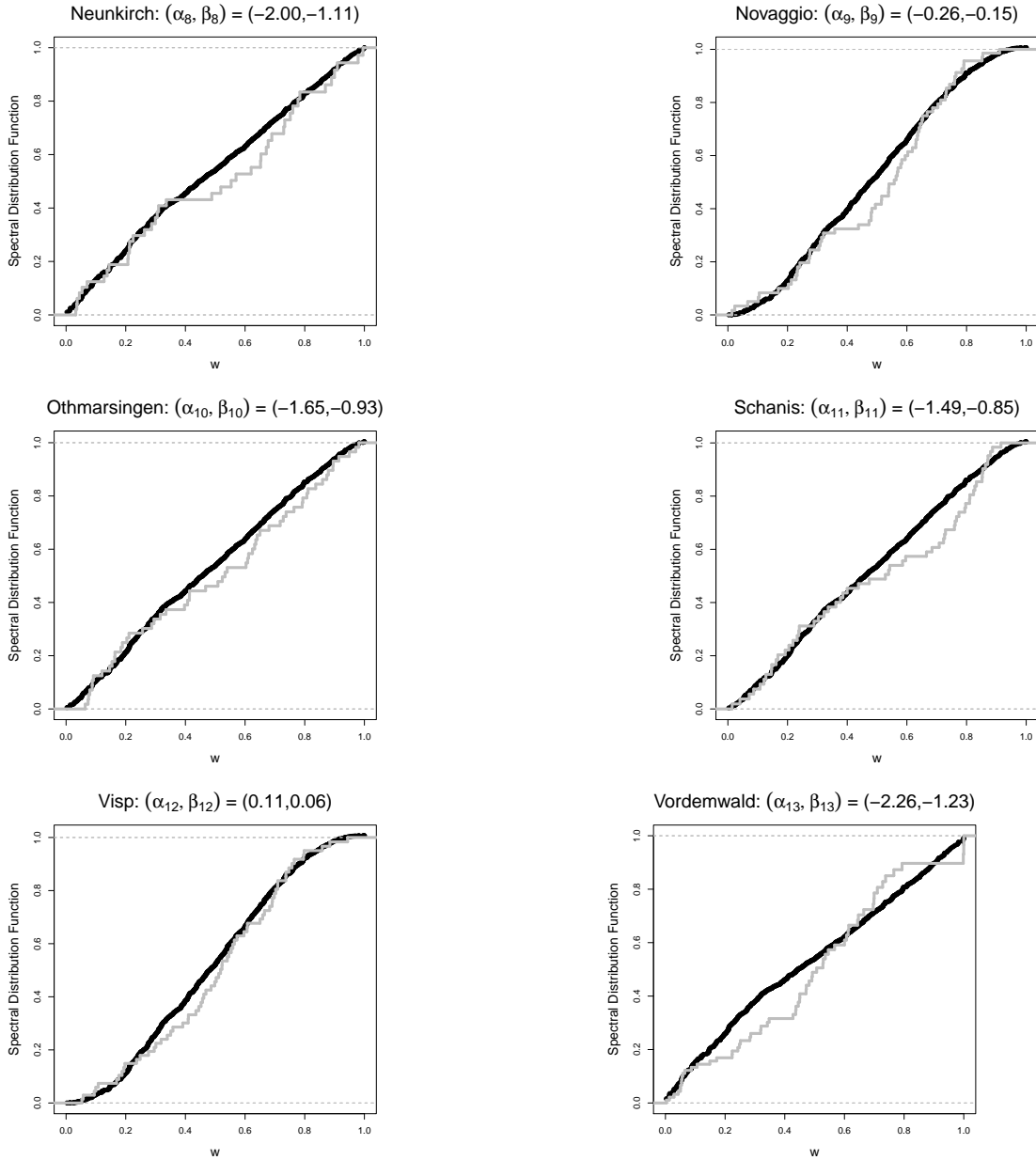
Spectral Distribution Function

w

17
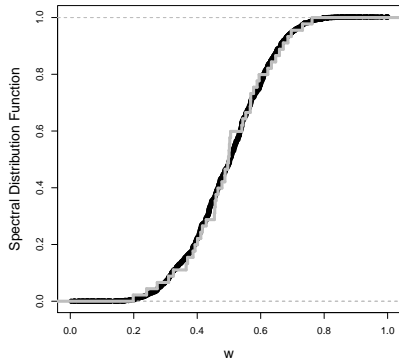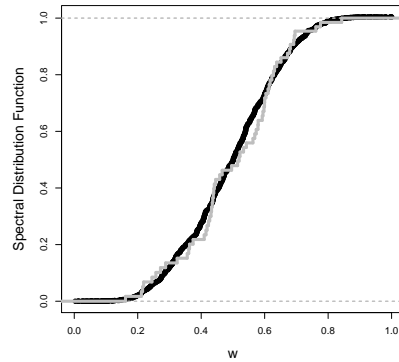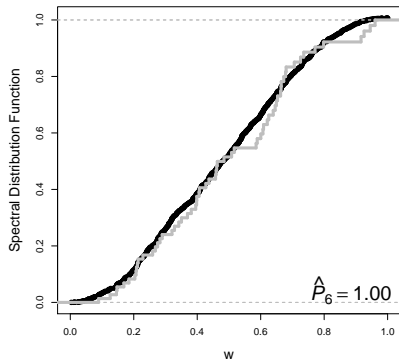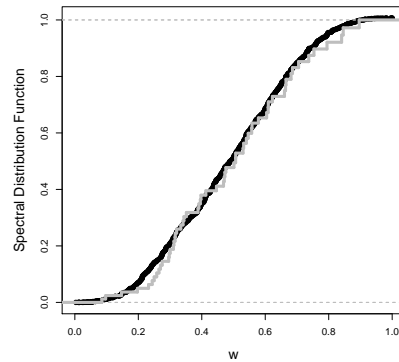
Figure 2: Estimates of the spectral distribution functions obtained by empirical likelihood estimation of the spectral density ratio model; the grey line represents the individual sample empirical likelihood estimates.
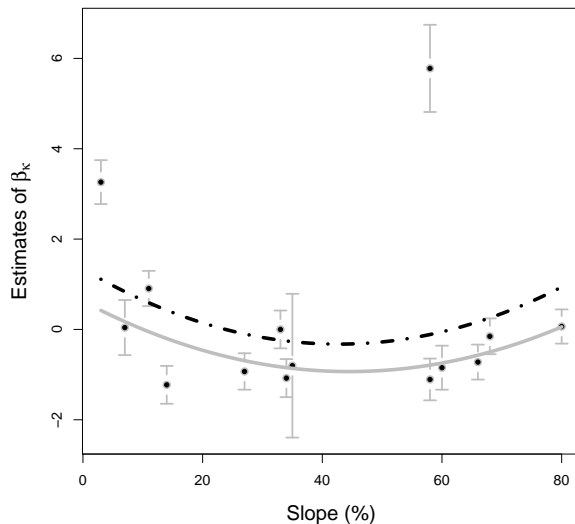
18

Figure 3: Slope of the forest plotted against estimated tilting parameters. The solid line corresponds to the model without Isone and Jussy; the dashed line corresponds to the model with all forests.

by dropping Isone and Jussy and conducting nonparametric inference by empirical likelihood over the fitted linear model: $E(\widehat{\beta} \mid X) = 0.62 - 0.07 \times \text{slope} + 7 \times 10^{-4} \times \text{slope}^2$. A bootstrap calibration yields the confidence intervals $(-0.12; 0.01)$ and $(-10^{-4}; 2 \times 10^{-4})$, for the coefficients of the linear and quadratic effects of slope, respectively, thus casting doubt on the significance of a quadratic association between the sheltering capacity of the forest and its slope.

5 Discussion

This paper introduces the spectral density ratio model for multivariate extremes, designed for contexts where K samples are available and there is the need to understand how extremal dependence is related to covariates. The rich semiparametric formulation allows us to link a family of K unknown spectral densities, constrained by marginal moment constraints, using a tilting function and regression specification that may be tailored to the problem at hand. Although flexible, the approach requires a substantial computational investment, and it would valuable to find faster algorithms, particularly as bootstrap resampling may be needed for reliable inference in applications.

Acknowledgments

Appendix A: Profiling the baseline spectral distribution function with normalization and marginal moment constraints

Let $v_i = (v_i^{(1)}, \dots, v_i^{(d)})^{\mathrm{T}}$. The Lagrangian corresponding to the empirical likelihood optimisation problem of interest is

$$L = \sum_{i=1}^{n} \log p_i + \sum_{k=1}^{K-1} \sum_{j=1}^{n_k} \{\alpha_k + \beta_k c(w_{kj})\}$$
$$- \eta_0 \Big(\sum_{i=1}^{n} p_i - 1 \Big) - n \sum_{k=1}^{K-1} \eta_k \Big\{ \sum_{i=1}^{n} p_i(g(v_i, \gamma_k) - 1) \Big\} - n_0 \sum_{d=1}^{D} \sum_{k=0}^{K-1} \delta_{dk} \Big\{ \sum_{i=1}^{n} p_i(v_i^{(d)} g(v_i, \gamma_k) - D^{-1}) \Big\}.$$

Setting the derivative of L with respect to p_i to 0, we obtain

$$\frac{1}{p_i} - \eta_0 - n \sum_{k=1}^{K-1} (g(v_i, \gamma_k) - 1) - n_0 \sum_{d=1}^{D} \sum_{k=0}^{K-1} \delta_{dk}(v_i^{(d)} g(v_i, \gamma_k) - D^{-1}) = 0, \quad i = 1, \dots, n.$$

Thus $\sum_{i=1}^{n} p_i \partial L / \partial p_i = n - \eta_0 = 0$, whence $n = \eta_0$, and hence

$$p_i = \frac{1}{n} \frac{1}{1 + \sum_{k=1}^{K-1} \eta_k(g(v_i, \gamma_k) - 1) + n_0/n \sum_{d=1}^{D} \sum_{k=0}^{K-1} \delta_{dk}(v_i^{(d)} g(v_i, \gamma_k) - D^{-1})}.$$

In order to obtain expressions for the η_κ, set $\partial L / \partial \alpha_\kappa = 0$, for $\kappa = 1, \dots, K-1$. It follows that

$$\eta_\kappa = \frac{n_\kappa}{n} - \frac{n_0}{n} \sum_{d=1}^{D} \delta_{d\kappa} D^{-1},$$

and

$$p_i = \frac{1}{n_0} \frac{1}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \sum_{d=1}^{D} \sum_{k=0}^{K-1} \delta_{dk} g(v_i, \gamma_k)(v_i^{(d)} - D^{-1})}.$$

If we define $\delta = (\delta_0, \dots, \delta_{K-1})^{\mathrm{T}}$, then

$$p_i = \frac{1}{n_0} \frac{1}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}} \mathcal{M}(v_i, \gamma)},$$

where

$$\mathcal{M}(v, \gamma) = \mathcal{G}(v, \gamma) \otimes (v - D^{-1} \mathbf{1}_D), \quad \mathcal{G}(v, \gamma) = (1, g(v, \gamma_1), \dots, g(v, \gamma_{K-1}))^{\mathrm{T}}.$$

Appendix B: Convex dual representation and the inner optimisation problem

The convex dual representation of the empirical likelihood problem under analysis is helpful for computational purposes. Here the necessary dual involves minimising

$$\mathcal{L}(\delta) = -\sum_{i=1}^{n} \log\left\{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}} \mathcal{M}(v_i, \gamma)\right\},$$

subject to the linear constraints $1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}} \mathcal{M}(v_i, \gamma) > 0$, for $i = 1, \ldots, n$. These constraints can be removed by using the pseudo-logarithmic function introduced by Owen (2001, p. 235), i.e.,

$$\log_\#(s) = \begin{cases} \log(s), & s > \epsilon, \\ \log(\epsilon) - 1.5 + 2s/\epsilon - s^2/(2\epsilon^2), & s \le \epsilon, \end{cases}$$

for some small $\epsilon > 0$. Then the initial problem of interest simplifies into one of minimising

$$\mathcal{L}_\#(\delta) = -\sum_{i=1}^{n} \log_\#\left\{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}} \mathcal{M}(v_i, \gamma)\right\}.$$

over $\delta \in \mathbb{R}^{KD}$, for which a Newton algorithm can be implemented by recursive least squares. We write the gradient and Hessian of $\mathcal{L}_\#$ as

$$\frac{\partial \mathcal{L}_\#}{\partial \delta} = -U^{\mathrm{T}} y, \qquad \frac{\partial \mathcal{L}_\#^2}{\partial \delta \partial \delta^{\mathrm{T}}} = U^{\mathrm{T}} U,$$

where $U = (u_1, \ldots, u_n)^{\mathrm{T}}$ and $y = (y_1, \ldots, y_n)$, are defined as

$$u_i = \left[-\log_\#^{(2)}\left\{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}} \mathcal{M}(v_i, \gamma)\right\}\right]^{1/2} \times [\mathcal{M}(v_i, \gamma)]^{\mathrm{T}},$$

$$y_i = \frac{\log_\#^{(1)}\left\{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}} \mathcal{M}(v_i, \gamma)\right\}}{\left[-\log_\#^{(2)}\left\{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}} \mathcal{M}(v_i, \gamma)\right\}\right]^{1/2}},$$

and $\log_\#^{(i)}$ denotes the i-th derivative of the pseudo-logarithmic function.

Numerical optimisation can then be performed by updating δ according to the rule $\delta \to \delta + (U^{\mathrm{T}} U)^{-1} U^{\mathrm{T}} y$, which uses the preceding values of the Lagrange multipliers δ corresponding to the marginal moment constraints, and an increment $(U^{\mathrm{T}} U)^{-1} U^{\mathrm{T}} y$; the latter is readily obtained by least squares regression of y on U.

Appendix C: Proof of Theorem 1

a) By the asymptotic theory of M-estimators (van der Vaart, 1998, §5) and Taylor expansion, we obtain

$$\frac{\partial \ell_{\mathrm{p}}}{\partial \gamma}\bigg|_{\gamma=\widehat{\gamma}} = \frac{\partial \ell_{\mathrm{p}}}{\partial \gamma}\bigg|_{\gamma=\gamma_o} + \frac{\partial \ell_{\mathrm{p}}}{\partial \gamma \partial \gamma^{\mathrm{T}}}\bigg|_{\gamma=\gamma_o} (\widehat{\gamma} - \widehat{\gamma}_o) + o_p(n^{-1/2}).$$

Since by definition the estimate $\widehat{\gamma}$ solves the equation $0 = \partial \ell_{\mathrm{p}}/\partial \gamma$, the latter expansion can be rewritten as

$$n^{1/2}(\widehat{\gamma} - \gamma_0) = \left\{-\frac{1}{n}\left(\frac{\partial^2 \ell_{\mathrm{p}}}{\partial \gamma \partial \gamma^{\mathrm{T}}}\right)^{-1} \frac{1}{n^{1/2}} \frac{\partial \ell_{\mathrm{p}}}{\partial \gamma}\right\}\bigg|_{\gamma=\gamma_o} + o_p(1). \tag{20}$$

By the weak law of large numbers and the central limit theorem, with γ evaluated at the true value γ_o, a second Bartlett identity and the asymptotic normality of $\partial \ell_{\mathrm{p}}/\partial \gamma$ follow, i.e.,

$$-\frac{1}{n}\frac{\partial^2 \ell_{\mathrm{p}}}{\partial \gamma \partial \gamma^{\mathrm{T}}} - \mho = o_p(1), \qquad \frac{1}{n^{1/2}}\frac{\partial \ell_{\mathrm{p}}}{\partial \gamma} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Omega). \tag{21}$$

Let $\{\mathrm{E}_k\}_{k=1}^{K-1}$ denote the canonical basis of $\mathbb{R}^{K-1}$, and use the auxiliary notation

$$M_{\alpha_\kappa}(v,\gamma) = (\mathrm{E}_\kappa \odot \mathcal{G}(v,\gamma)) \otimes (v - D^{-1}\mathbf{1}_D), \quad M_{\beta_\kappa}(v,\gamma) = c(v)M_{\alpha_\kappa}(v,\gamma), \tag{22}$$

with $\odot$ and $\otimes$ respectively denoting the Hadamard and Kronecker products. To avoid unnecessary complications below we omit the arguments of most functions and use the shorthand notation g_κ to denote $g(v,\gamma_k)$. The matrix $\mho$ has dimensions $(K-1)(1+p) \times (K-1)(1+p)$ and for $\kappa, \kappa' = 1, \ldots, K-1$ its entries are

$$-\frac{1}{n}\frac{\partial^2 \ell_{\mathrm{p}}}{\partial \alpha_\kappa^2} = \int \left\{ \frac{(\rho_\kappa g_\kappa + \delta^{\mathrm{T}} M_{\alpha_\kappa})(1 + \sum_{k \neq \kappa} \rho_k g_k + \delta^{\mathrm{T}}(\mathcal{M} - M_{\alpha_\kappa}))\mathrm{d}H_0}{(1 + \sum_{k=1}^{K-1} \rho_k)(1 + \sum_{k=1}^{K-1} \rho_k g_k + \delta^{\mathrm{T}}\mathcal{M})} \right\} + o_p(1),$$

$$-\frac{1}{n}\frac{\partial \ell_{\mathrm{p}}}{\partial \alpha_\kappa \alpha_{\kappa'}} = \int \left\{ \frac{-(\rho_\kappa g_\kappa + \delta^{\mathrm{T}} M_{\alpha_\kappa})(\rho_{\kappa'} g_{\kappa'} + \delta^{\mathrm{T}} M_{\alpha_{\kappa'}})\mathrm{d}H_0}{(1 + \sum_{k=1}^{K-1} \rho_k)(1 + \sum_{k=1}^{K-1} \rho_k g_k + \delta^{\mathrm{T}}\mathcal{M})} \right\} + o_p(1),$$

$$-\frac{1}{n}\frac{\partial \ell_{\mathrm{p}}}{\partial \alpha_\kappa \partial \beta_\kappa} = \int \left\{ \frac{(\rho_\kappa c g_\kappa + \delta^{\mathrm{T}} M_{\beta_\kappa})(1 + \sum_{k \neq \kappa} \rho_k g_k + \delta^{\mathrm{T}}(\mathcal{M} - M_{\alpha_\kappa}))\mathrm{d}H_0}{(1 + \sum_{k=1}^{K-1} \rho_k)(1 + \sum_{k=1}^{K-1} \rho_k g_k + \delta^{\mathrm{T}}\mathcal{M})} \right\} + o_p(1),$$

$$-\frac{1}{n}\frac{\partial \ell_{\mathrm{p}}}{\partial \alpha_\kappa \partial \beta_{\kappa'}} = \int \left\{ \frac{-(\rho_\kappa g_\kappa + \delta^{\mathrm{T}} M_{\alpha_\kappa})(\rho_{\kappa'} c g_{\kappa'} + \delta^{\mathrm{T}} M_{\beta_{\kappa'}})\mathrm{d}H_0}{(1 + \sum_{k=1}^{K-1} \rho_k)(1 + \sum_{k=1}^{K-1} \rho_k g_k + \delta^{\mathrm{T}}\mathcal{M})} \right\} + o_p(1),$$

$$-\frac{1}{n}\frac{\partial \ell_{\mathrm{p}}}{\partial \beta_\kappa^2} = \int \left\{ \frac{\rho_\kappa(c)^2 g_\kappa + x_\kappa c\delta^{\mathrm{T}} M_{\beta_\kappa} \mathrm{d}H_0}{1 + \sum_{k=1}^{K-1} \rho_k} \right\} + \int \left\{ \frac{(\rho_\kappa c g_\kappa + \delta^{\mathrm{T}} M_{\beta_\kappa})^2 \mathrm{d}H_0}{(1 + \sum_{k=1}^{K-1} \rho_k)(1 + \sum_{k=1}^{K-1} \rho_k g_k + \delta^{\mathrm{T}}\mathcal{M})} \right\} + o_p(1),$$

$$-\frac{1}{n}\frac{\partial \ell_{\mathrm{p}}}{\partial \beta_\kappa \partial \beta_{\kappa'}} = \int \left\{ \frac{-(\rho_\kappa c g_\kappa + \delta^{\mathrm{T}} M_{\beta_\kappa})(\rho_{\kappa'} c g_{\kappa'} + \delta^{\mathrm{T}} M_{\beta_{\kappa'}})\mathrm{d}H_0}{(1 + \sum_{k=1}^{K-1} \rho_k)(1 + \sum_{k=1}^{K-1} \rho_k g_k + \delta^{\mathrm{T}}\mathcal{M})} \right\} + o_p(1).$$

The entries in the matrix $\Omega = \mathrm{var}\left(n^{-1/2}\partial \ell_{\mathrm{p}}/\partial \gamma \right)$ are

$$\frac{1}{n}\mathrm{var}\left(\frac{\partial \ell_{\mathrm{p}}}{\partial \alpha_\kappa} \right) = \sum_{l=0}^{K-1} \frac{\rho_l\{\mathsf{A}(\kappa,l) - \{\mathsf{a}(\kappa,l)\}^2\}}{1 + \sum_{k=1}^{K-1} \rho_k}, \quad \frac{1}{n}\mathrm{cov}\left(\frac{\partial \ell_{\mathrm{p}}}{\partial \alpha_\kappa}, \frac{\partial \ell_{\mathrm{p}}}{\partial \alpha_{\kappa'}} \right) = \sum_{l=0}^{K-1} \frac{\rho_l\{\mathsf{ab}(\kappa,\kappa',l) - \mathsf{a}(\kappa,l)\mathsf{a}(\kappa',l)\}}{1 + \sum_{k=1}^{K-1} \rho_k},$$

$$\frac{1}{n}\mathrm{cov}\left(\frac{\partial \ell_{\mathrm{p}}}{\partial \alpha_\kappa}, \frac{\partial \ell_{\mathrm{p}}}{\partial \beta_\kappa} \right) = \frac{\rho_\kappa \Gamma + \sum_{l=0}^{K-1} \rho_l\{\mathsf{ab}(\kappa,\kappa,l) - \mathsf{a}(\kappa,l)\mathsf{b}(\kappa,l)\}}{1 + \sum_{k=1}^{K-1} \rho_k},$$

$$\frac{1}{n}\mathrm{cov}\left(\frac{\partial \ell_{\mathrm{p}}}{\partial \alpha_\kappa}, \frac{\partial \ell_{\mathrm{p}}}{\partial \beta_{\kappa'}} \right) = \frac{\rho_{\kappa'} \Delta + \sum_{l=0}^{K-1} \rho_l\{\mathsf{ab}(\kappa,\kappa',l) - \mathsf{a}(\kappa,l)\mathsf{a}(\kappa',l)\}}{1 + \sum_{k=1}^{K-1} \rho_k},$$

$$\frac{1}{n}\mathrm{var}\left(\frac{\partial \ell_{\mathrm{p}}}{\partial \beta_\kappa} \right) = \frac{\rho_\kappa \Pi + \sum_{l=0}^{K-1} \rho_l\{\mathsf{B}(\kappa,l) - \mathsf{b}(\kappa,l)\}}{1 + \sum_{k=1}^{K-1} \rho_k},$$

$$\frac{1}{n}\mathrm{cov}\left(\frac{\partial \ell_{\mathrm{p}}}{\partial \beta_\kappa}, \frac{\partial \ell_{\mathrm{p}}}{\partial \beta_{\kappa'}} \right) = \frac{\rho_{\kappa'} \Phi + \sum_{l=0}^{K-1} \rho_l\{\mathsf{ab}(\kappa,\kappa',l) - \mathsf{a}(\kappa,l)\mathsf{a}(\kappa',l)\}}{1 + \sum_{k=1}^{K-1} \rho_k},$$

where

$$\mathsf{A}(i,j) = \int \frac{(\rho_i g_i + \delta^{\mathrm{T}} M_{\alpha_i})^2 g_j \mathrm{d}H_0}{(1 + \sum_{k=1}^{K-1} \rho_k g_k + \delta^{\mathrm{T}}\mathcal{M})^2}, \quad \mathsf{a}(i,j) = \int \frac{(\rho_i g_i + \delta^{\mathrm{T}} M_{\alpha_i}) g_j \mathrm{d}H_0}{1 + \sum_{k=1}^{K-1} \rho_k g_k + \delta^{\mathrm{T}}\mathcal{M}},$$

$$\mathsf{ab}(i,j,k) = \int \frac{(\rho_i g_i + \delta^{\mathrm{T}} M_{\alpha_i})(\rho_j g_j + \delta^{\mathrm{T}} M_{\alpha_j}) g_k \mathrm{d}H_0}{(1 + \sum_{k=1}^{K-1} \rho_k g_k + \delta^{\mathrm{T}}\mathcal{M})^2}, \quad \mathsf{ac}(i,j) = \int \frac{(\rho_i g_i + \delta_i M_{\alpha_i}) c g_j \mathrm{d}H_0}{1 + \sum_{k=1}^{K-1} \rho_k g_k + \delta^{\mathrm{T}}\mathcal{M}},$$

$$\mathsf{B}(i,j) = \int \frac{(\rho_i c g_i + \delta^{\mathrm{T}} M_{\beta_i})^2 g_j \mathrm{d}H_0}{(1 + \sum_{k=1}^{K-1} \rho_k g_k + \delta^{\mathrm{T}}\mathcal{M})^2}, \quad \mathsf{b}(i,j) = \int \frac{(\rho_i c g_i + \delta^{\mathrm{T}} M_{\beta_i})^2 g_j \mathrm{d}H_0}{1 + \sum_{k=1}^{K-1} \rho_k g_k + \delta^{\mathrm{T}}\mathcal{M}},$$

$$\mathsf{C}(i) = \int (c)^2 g_i \mathrm{d}H_0, \quad \mathsf{c}(i) = \int c g_i \mathrm{d}H_0,$$

and $\Gamma = -\mathsf{ac}(\kappa,\kappa) + \mathsf{a}(\kappa,\kappa)\mathsf{c}(\kappa)$, $\Delta = -\mathsf{ac}(\kappa,\kappa') + \mathsf{a}(\kappa,\kappa')\mathsf{c}(\kappa')$, $\Phi = -\mathsf{ac}(\kappa,\kappa') + \mathsf{a}(\kappa,\kappa')\mathsf{c}(\kappa')$, and $\Pi = \mathsf{C}(\kappa) - \{\mathsf{c}(\kappa)\}^2 - 2\mathsf{ac}(\kappa,\kappa) + 2\mathsf{a}(\kappa,\kappa)\mathsf{c}(\kappa)$.

Combining the large sample results in (21) with (20), and making use of Slutsky's theorem, we get

$$n^{1/2}(\widehat{\gamma} - \gamma_o) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma),$$

where

$$\Sigma = \mho^{-1}\Omega\mho^{-1}. \tag{23}$$

b) A second-order Taylor expansion yields

$$\ell_{\mathrm{p}}(\widehat{\gamma}) = \ell_{\mathrm{p}}(\gamma^0) + \frac{1}{2}(\widehat{\gamma} - \gamma^0)\frac{\partial^2 \ell_{\mathrm{p}}}{\partial \gamma \partial \gamma^{\mathrm{T}}}_{|\gamma = \gamma^0}(\widehat{\gamma} - \gamma^0) + o_p(1),$$

thus implying that

$$2\{\ell_{\mathrm{p}}(\widehat{\gamma}) - \ell_{\mathrm{p}}(\gamma^0)\} = n^{1/2}(\widehat{\gamma} - \gamma^0)\left\{\frac{1}{n}\frac{\partial^2 \ell_{\mathrm{p}}}{\partial \gamma \partial \gamma^{\mathrm{T}}}\right\}_{|\gamma = \gamma^0} n^{1/2}(\widehat{\gamma} - \gamma^0)^{\mathrm{T}} + o_p(1). \tag{24}$$

The final result now follows by combining (24) with the large sample results stated in (21), and making use of Slutsky's theorem.

References

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

Boldi, M.-O. and Davison, A. C. (2007) A mixture model for multivariate extremes. *J. R. Statist. Soc. B*, **69**, 217–229.

Chauduri, S., Drton, M. and Richardson, T. S. (2007) Estimation of a covariance matrix with zeros. *Biometrika*, **94**, 199–216.

Cheng, J., Qin, J. and Zhang, B. (2009) Semiparametric estimation and inference for distributional and general treatment effects. *J. R. Statist. Soc. B*, **71**, 881–904.

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

Coles, S. G. (2001) *An Introduction to Statistical Modelling of Extreme Values*. London: Springer.

Coles, S. G. and Tawn, J. A. (1991) Modelling extreme multivariate events. *J. R. Statist. Soc. B*, **53**, 377–392.

Efron, B. and Tibshirani, R. (1996) Using specially designed exponential families for density estimation. *Ann. Statist.*, **24**, 2431–2461.

Einmahl, J. H. J, Li, J. and Liu, R. Y (2009) Thresholding events of extremes in simultaneous monitoring of multiple risks. *J. Am. Statist. Assoc.*, **104**, 982–992.

Einmahl, J. H. J. and Segers, J. (2009) Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist.*, **37**, 2953–2989.

Embrechts, P., Lambrigger, D. D. and Wütrich, M. V. (2009) Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes*, **12**, 107–127.

Ferrez, J., Davison, A. C. and Rebetez (2011) Extreme temperature analysis under forest cover compared to an open field. *Agricultural and Forest Meteorology*, to appear.

Fokianos, K., Kedem, B., Qin, J. and Short, D. A. (2001) A semiparametric approach to the one-way layout. *Technometrics*, **43**, 56–65.

Fokianos, K. (2004) Merging information for semiparametric density estimation. *J. R. Statist. Soc. B*, **66**, 941–958.

Gilbert, P. B., Lele, S. R. and Vardi, Y. (1999) Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika*, **10**, 839–866.

Heffernan, J. E. and Tawn, J. A. (2004) A conditional approach for multivariate extreme values (with discussion). *J. R. Statist. Soc. B*, **66**, 497–546.

Jarušková, D. (2009) Modelling Multivariate Extremes of Precipitation Series in Northern Moravia. *Environmetrics*, **20**, 751–775.

Kedem, B., Kim, E.-Y., Voulgaraki, A., Graubard, B. I. (2009) Two-dimensional semiparametric density ratio modeling of testicular germ cell data. *Stat. Med.*, **28**, 2147–2159.

Kotz, S. and Nadarajah, S. (2000) *Extreme Value Distributions: Theory and Applications*. London: Imperial College Press.

Owen, A. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A. (1991) Empirical likelihood for linear models. *Ann. Statist.*, **19**, 1725–1747.

Owen, A. (2001) *Empirical Likelihood*. Boca Raton: Chapman and Hall.

Pickands, J. (1981) Multivariate extreme value distributions. *Proc. 43rd Sess. Int. Statist. Inst.*, 859–878.

Qin, J. and Lawless, J. F. (1994) Empirical likelihood and general estimating equations. *Ann. Statist.*, **22**, 300–325.

Qin, J. and Zhang, B. (1997) A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, **84**, 609–618.

Ramos, A. and Ledford, A. (2009) A new class of models for bivariate joint tails. *J. R. Statist.*

Soc. B, **71**, 219–241.

Renaud, V. and Rebetez, M. (2009) Comparison between open-site and below-canopy climatic conditions in Switzerland during the exceptionally hot summer of 2003. *Agr. Forest. Meteorol.*, **149**, 873–880.

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

van der Vaart, A. W. (1998) *Asymptotic Statistics*. Cambridge: Cambridge University Press.