Edinburgh Research Explorer

Spectral Density Ratio Models for Multivariate Extremes

Spectral Density Ratio Models for Multivariate Extremes

Miguel DE CARVALHO and Anthony C. DAVISON

Abstract

The modeling of multivariate extremes has received increasing recent attention because of its importance in risk assessment. In classical statistics of extremes, the joint distribution of two or more extremes has a nonparametric form, subject to moment constraints. This paper develops a semiparametric model for the situation where several multivariate extremal distributions are linked through the action of a covariate on an unspecified baseline distribution, through a so-called density ratio model. Theoretical and numerical aspects of empirical likelihood inference for this model are discussed, and an application is given to pairs of extreme forest temperatures.

KEY WORDS: Air temperature; Empirical likelihood; Exponential tilting; Forest microclimate; Multivariate extreme values; Semiparametric modeling; Spectral distribution.

1   INTRODUCTION

As human society becomes more complex, it becomes more vulnerable to rare but catastrophic events, such as the Fukushima nuclear accident, widespread flooding, or turbulence in financial markets. Assessment of the risks of such events involves the estimation of small probabilities and hence entails

Miguel de Carvalho is Assistant Professor, Department of Statistics, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Santiago, Chile (*MdeCarvalho@mat.puc.cl*). Anthony Davison is Professor of Statistics, EPFL-FSB-MATHAA-STAT, Station 8, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (*Anthony.Davison@epfl.ch*).

extrapolation into the tails of multivariate distributions, often beyond any existing data. The mathematical basis for such extrapolation is the statistics of extremes, which is an active domain of current research. Published applications include air quality monitoring (Heffernan and Tawn 2004), wave surge analysis (Ramos and Ledford 2009), aviation safety (Einmahl et al. 2009), precipitation studies (Jarušková 2009) and finance (Embrechts et al. 2009). A central concept in classical multivariate extreme value theory is the so-called spectral distribution, which determines the degree of dependence among extremes of different variables. Such a distribution must satisfy certain moment constraints, and 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; Ballani and Schlather 2011; Cooley et al. 2010; Kotz and Nadarajah 2000, chap. 3). Boldi and Davison (2007) introduced a constrained mixture of Dirichlet distributions which is weakly dense in the class of all possible spectral measures (see also Sabourin and Naveau, 2013), and Einmahl and Segers (2009) 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, whereas it is sometimes necessary to consider a family of spectral measures, $H_0, \ldots, H_K$, in order to capture the effect of explanatory variables on joint extremes. In this paper, we discuss the joint modeling of extremal events when data are gathered from several populations, to each of which corresponds a vector of covariates. The work is motivated by bivariate data on air temperatures at several sites around Switzerland, collected in order to understand how extremely high temperatures under the forest canopy relate to those in a nearby the open field (Renaud and Rebetez, 2009; Ferrez et al., 2011; Renaud et al., 2011). Each site is characterized by 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 quantification of how their joint distribution depends on the features. This is of interest for reasons such as maximizing the shelter offered by different types of forest during heat waves or the design of wooded parks in urban areas.

A simple approach to the modeling of such data would be to fit parametric distributions to the individual subpopulations, with the parameters depending on the covariates through 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 indirect. Second, the focus of interest in our application is on the effects of the covariates, so we prefer to avoid choosing a particular parametric model. Third, just as there are efficiency gains from using threshold exceedances rather than block maxima when modeling extrema (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_k using all the 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 innovation here is a semiparametric spectral density ratio model, wherein the spectral distribution functions are unspecified but related through a known weight function modulated by so-called tilting parameters. In a second step a regression is conducted, where the tilting parameters themselves are dependent variables. Related models have been applied in 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 modeling. One issue is that in the extremal context the multivariate model applies only as a limit, which is not attained in finite samples, so it cannot be true that our model, or indeed any other simple formulation, applies exactly. In common with much of the extremal literature we ignore this issue in the main development, but discuss it in Section 5.

The structure of our paper is as follows. In the next section we introduce our model and corresponding inference methods. In Section 3 we assess the performance of our methods by simulation, and in Section 4 we provide a real data application on extreme forest temperatures. The appendix includes proofs of the main results; further details can be found in the online Supplementary Material.

2 SPECTRAL DENSITY RATIO MODEL

2.1 Background on Statistics of Extremes

Before introducing models for multivariate extremes, we start with basic notions on univariate extremes. Let $Y_1, \ldots, Y_N$ be independent and identically distributed random variables with distribution function F. A central result is the extremal types theorem, which states that if there exist sequences of constants $\{a_N > 0\}$ and $\{b_N\}$, such that as $N \to \infty$ the distribution of the rescaled sample maximum $M_N = a_N^{-1}\{\max(Y_1, \ldots, Y_N) - b_N\}$ converges to a non-degenerate limit, then this limit must be the generalized extreme value distribution (Coles, 2001, Theorem 3.1)

$$G(y) = \exp\left[-\left\{ 1 + \xi\left(\frac{y - \mu}{\sigma}\right) \right\}_+^{-1/\xi} \right]. \tag{1}$$

Here we write $x_+ = \max(x, 0)$, and $\mu \in \mathbb{R}$, $\sigma > 0$, and $\xi \in \mathbb{R}$ are respectively location, scale, and shape parameters. If $\xi > 0$, $\xi = 0$, or $\xi < 0$ we say that F is in the domain of attraction of a Fréchet, Gumbel, or reverse Weibull distribution; the unit Fréchet distribution has $\xi = \sigma = \mu = 1$.

The analogue of the extremal types theorem in the D-dimensional setting is Pickands' (1981) representation theorem. Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N$ be independent and identically distributed vectors of continuous random variables on $\mathbb{R}^D$. Pickands' result states that if there exist sequences of vector constants $\{\boldsymbol{a}_N > 0\}$ and $\{\boldsymbol{b}_N\}$ such that the rescaled component-wise maximum $\boldsymbol{M}_N = \boldsymbol{a}_N^{-1}\{\max(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N) - \boldsymbol{b}_N\}$ converges in distribution as $N \to \infty$, then its limiting distribution must be a multivariate extreme value distribution (Beirlant et al., 2004, p. 254), i.e.,

$$P(\boldsymbol{M}_N \leq \boldsymbol{y}) \to G(\boldsymbol{y}), \quad \boldsymbol{y} \in [-\infty, \infty]^D.$$

It is convenient to transform this limiting distribution to have unit Fréchet marginal distributions, $\exp(-1/y_d)$, for $y_d > 0$ and $d = 1, \ldots, D$; since the univariate margins must be of form (1), this is always possible using the inverse of the transformation $y \mapsto \{1 + \xi(y - \mu)/\sigma\}_+^{1/\xi}$. It then turns out that G has the form

$$G_H(\boldsymbol{y}) = \exp\left\{ -D \int_{S_D} \max\left(\frac{w_1}{y_1}, \ldots, \frac{w_D}{y_D}\right) \mathrm{d}H(\boldsymbol{w}) \right\}, \quad \boldsymbol{y} = (y_1, \ldots, y_D) \in [0, \infty]^D, \tag{2}$$

where H is the so-called spectral distribution function, or spectral measure, which has support in the unit simplex $S_D = \left\{ \boldsymbol{w} \in \mathbb{R}^D : \sum_{d=1}^D w_d = 1, w_d \geq 0, d = 1, \ldots, D \right\}$. Comparison of (1) and (2) reveals

4

that a major difference between the limiting distributions of univariate and multivariate maxima is that (1) is parametric, whereas (2) depends upon a probability measure H. The measure H determines the degree of association between the extremes of the D random variables, and must only satisfy the normalization and moment conditions

$$\int_{S_D} \mathrm{d}H(\boldsymbol{w}) = 1, \quad \int_{S_D} \boldsymbol{w}\, \mathrm{d}H(\boldsymbol{w}) = D^{-1}\boldsymbol{1}_D, \tag{3}$$

where $\boldsymbol{1}_D$ is the D-vector of ones. The moment constraint in (3) can be obtained by letting $y_d \to \infty$, for $d \neq d'$, in equation (2), and noting that G_H has unit Fréchet margins; analogous constraints would be produced by transformation to any other set of known marginal distributions.

A pseudo-polar transformation is helpful for describing the level of dependence between the extremes of two random variables, and hence for grasping the role played by H: if we map $\boldsymbol{Y}_i$ into the pseudo-polar coordinates $(\boldsymbol{W}_i = R_i^{-1}\boldsymbol{Y}_i, R_i = \sum_{d=1}^D Y_{i,d})$, where R_i is the pseudo-radius, then the pseudo-angle $\boldsymbol{W}_i$ has measure H on S_D in the limit as $R_i \to \infty$. When $D = 2$, the unit simplex S_D is simply the unit interval, so we may write $w_1 = w$, $w_2 = 1 - w$, where $w \in [0, 1]$. Then if extreme values of the two variables tend to occur individually, H places most of its mass close to the limits $w = 0$ and $w = 1$, whereas if the extremes tend to occur together then H places mass mainly near $w = 1/2$; $w = 0$ and $w = 1$ correspond to the axes $y_1 = 0$ and $y_2 = 0$, whereas $w = 1/2$ corresponds to the bisector of the first quadrant. In the multivariate setting the interpretation is analogous: extremal independence corresponds to a spectral distribution placing equal masses D^{-1} at the vertices of S_D, giving $G_H(\boldsymbol{y}) = \exp(-\sum_{d=1}^D y_d^{-1})$, whereas perfect extremal dependence corresponds to a spectral distribution having unit mass at the barycenter $D^{-1}\boldsymbol{1}_D$ of S_D, giving $G_H(\boldsymbol{y}) = \exp\{-1/\min(y_1, \ldots, y_D)\}$.

This pseudo-polar representation also arises in a point process characterization (Coles, 2001, sec. 8.3), under which the rescaled observations converge in distribution to a Poisson process whose intensity measure on $\mathbb{R}_+ \times S_D$ factorizes as

$$\Lambda(\mathrm{d}r \times \mathrm{d}\boldsymbol{w}) = D\frac{\mathrm{d}r}{r^2} \times \mathrm{d}H(\boldsymbol{w}); \tag{4}$$

the angular dispersion of its points is controlled by H, independent of the pseudo-radial component r.

In our setting we assume that we have independent sets of observations $\{\boldsymbol{w}_{k,1}, \ldots, \boldsymbol{w}_{k,n_k}\}$ ($k = 0, \ldots, K$) from $K + 1$ unknown spectral distributions $H_0, \ldots, H_K$. The measures must satisfy the

normalization and moment constraints

$$\int_{S_D} \mathrm{d}H_k(\boldsymbol{w}) = 1, \quad \int_{S_D} \boldsymbol{w} \, \mathrm{d}H_k(\boldsymbol{w}) = D^{-1}\mathbf{1}_D, \quad k = 0, \ldots, K. \tag{5}$$

The next section introduces the spectral density ratio model for modeling multi-sample multivariate extremes.

2.2 Spectral Density Model

Later we focus our attention on the simplex S_D on which spectral measures are defined, but for now we let H_k denote any distribution function. Our interest lies in the measures $H_0, \ldots, H_K$, which are assumed to be linked through a positive function g of known form,

$$\left\{ (H_0, \ldots, H_K) : \frac{\mathrm{d}H_k(\boldsymbol{w})}{\mathrm{d}H_0(\boldsymbol{w})} = g(\boldsymbol{w}, \boldsymbol{\psi}_k), \text{ for some } g(\boldsymbol{w}, \boldsymbol{\psi}_k) > 0; \; k = 0, \ldots, K \right\}. \tag{6}$$

Here $\boldsymbol{\psi} = (\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_K)^{\mathrm{T}}$ is a finite-dimensional parameter vector, and we set $g(\boldsymbol{w}, \boldsymbol{\psi}_0) \equiv 1$ for identifiability. The specification (6) is common to many models (Efron and Tibshirani, 1996; Qin and Zhang, 1997; Fokianos et al., 2001; Fokianos, 2004; Cheng et al., 2009). Under (6) the distributions H_k are left unspecified but are related through a known weight function, with the measure H_0 acting as a reference from which the other K measures are obtained through a distortion controlled by g and $\boldsymbol{\psi}_k$. This type of semiparametric approach provides gains in efficiency by estimating each of the H_k on the basis of the entire dataset (Gilbert et al., 1999; Fokianos, 2004; Kedem et al., 2009).

To motivate this approach, consider the two-sample univariate case in which $g(w; \alpha, \beta) = \exp\{\alpha + \beta c(w)\}$, where α and β are called tilting parameters and $c(w)$ is a known distortion function; for concreteness, let $c(w) = w$. Then the log-ratio of the densities $\log\{\mathrm{d}H_1(w)/\mathrm{d}H_0(w)\} = \alpha + \beta w$ is linear in the parameters, and thus $H_1(w)$ is a linear exponential family with baseline distribution $H_0(w)$ and canonical parameter and statistic β and w respectively; $-\alpha \equiv -\alpha(\beta)$ is the cumulant generator. This model has links to the logistic regression model, and an important property, the independence of irrelevant alternatives, implies that inferences are independent of the baseline (Fokianos, 2004).

Specification (6) turns out to be natural for modeling the multi-sample multivariate extreme value framework discussed in Section 2.1, with the constraints (5) restated as

$$\int_{S_D} g(\boldsymbol{w}, \boldsymbol{\psi}_k) \, \mathrm{d}H_k(\boldsymbol{w}) = 1, \quad \int_{S_D} \boldsymbol{w} \, g(\boldsymbol{w}, \boldsymbol{\psi}_k) \, \mathrm{d}H_k(\boldsymbol{w}) = D^{-1}\mathbf{1}_D, \quad k = 0, \ldots, K. \tag{7}$$

Below we refer to the general semiparametric setting (6), subject to the normalization and marginal moment constraints (7), as the spectral density ratio model. We propose to fit it through empirical likelihood methods (Owen, 1988, 2001).

2.3 Estimation

For simplicity below we sometimes denote the combined sample $\{\boldsymbol{w}_{0,1}, \ldots, \boldsymbol{w}_{0,n_0}, \ldots, \boldsymbol{w}_{K,1}, \ldots, \boldsymbol{w}_{K,n_K}\}$ by $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$. The likelihood of the $(K+1)$-sample multivariate extreme value problem under (6) is

$$\mathscr{L}(\boldsymbol{\psi}, H_0) = \prod_{k=0}^{K} \prod_{j=1}^{n_k} \mathrm{d}H_k(\boldsymbol{w}_{k,j}) = \prod_{i=1}^{n} p_i \prod_{k=0}^{K} \prod_{j=1}^{n_k} g(\boldsymbol{w}_{k,j}, \boldsymbol{\psi}_k), \tag{8}$$

where $p_i = \mathrm{d}H_0(\boldsymbol{v}_i)$ denotes the size in the jump of the baseline spectral distribution function H_0 at the observed $\boldsymbol{v}_i$.

We restrict our attention to the tilting function $g(\boldsymbol{w}, \boldsymbol{\psi}_k) = \exp\{\alpha_k + \beta_k c(\boldsymbol{w})\}$, where $\boldsymbol{\psi}_k = (\alpha_k, \beta_k)^{\mathrm{T}}$, and for identifiability set $\alpha_0 = \beta_0 = 0$. We suppress the dependence of $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^{\mathrm{T}}$ on $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^{\mathrm{T}}$, and vice versa, though the normalization constraints in (7) link these parameters. Thus the loglikelihood obtained from (8) is

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, H_0) = \sum_{i=1}^{n} \log p_i + \sum_{k=0}^{K} \sum_{j=1}^{n_k} \{\alpha_k + \beta_k c(\boldsymbol{w}_{k,j})\}. \tag{9}$$

Empirical likelihood estimation of the spectral density ratio model involves maximizing (9), subject to the empirical versions of constraints (7), viz,

$$
\begin{aligned}
p_i \geq 0, \qquad & \boldsymbol{v}_i \in S_D, \\
\sum_{i=1}^{n} p_i = 1, \qquad & \sum_{i=1}^{n} p_i \boldsymbol{v}_i = D^{-1} \mathbf{1}_D, \\
\sum_{i=1}^{n} p_i \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\} = 1, \qquad & \sum_{i=1}^{n} p_i \boldsymbol{v}_i \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\} = D^{-1} \mathbf{1}_D, \quad k = 1, \ldots, K.
\end{aligned}
\tag{10}
$$

Using an approach similar to that of Qin and Lawless (1994), it is shown in Appendix A.1 that if we use Lagrange multiplier procedures to profile p_i subject to the normalization and marginal moment

constraints (10), then the jump sizes for the baseline spectral distribution function can be written as

$$p_i \equiv p_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{1}{n} \frac{1}{\sum_{k=0}^{K} \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\}\{\rho_k + \boldsymbol{\lambda}_k^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D)\}}$$

$$= \frac{1}{n} \frac{1}{\rho_0 + \boldsymbol{\lambda}_0^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D) + \sum_{k=1}^{K} \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\}\{\rho_k + \boldsymbol{\lambda}_k^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D)\}},$$

(11)

where $\rho_k = n_k/n$, and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_0, \ldots, \boldsymbol{\lambda}_K)^{\mathrm{T}}$ denotes the Lagrange multipliers corresponding to the constraints in (10). Thus apart from a constant the profile empirical loglikelihood for $\boldsymbol{\beta}$ can be written as

$$\ell_{\mathrm{p}}(\boldsymbol{\beta}) = \max_{\boldsymbol{\alpha}, \boldsymbol{\lambda}, H_0} \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, H_0)$$

$$= -\sum_{i=1}^{n} \log\left[\rho_0 + \widehat{\boldsymbol{\lambda}}_0^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D) + \sum_{k=1}^{K} \exp\{\widehat{\alpha}_k + \beta_k \boldsymbol{v}_i\}\{\rho_k + \widehat{\boldsymbol{\lambda}}_k^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D)\}\right]$$

$$+ \sum_{k=1}^{K} \sum_{j=0}^{n_k} \{\widehat{\alpha}_k + \beta_k c(\boldsymbol{w}_{k,j})\},$$

(12)

where $\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta})$ and $\widehat{\boldsymbol{\lambda}} = \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta})$ are the joint solutions to the equations

$$\begin{cases} \dfrac{1}{n} \sum_{i=1}^{n} \dfrac{\exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\}}{\sum_{k=0}^{K} \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\}\{\rho_k + \boldsymbol{\lambda}_k^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D)\}} = 1, \\ \dfrac{1}{n} \sum_{i=1}^{n} \dfrac{\boldsymbol{v}_i \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\}}{\sum_{k=0}^{K} \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\}\{\rho_k + \boldsymbol{\lambda}_k^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D)\}} = D^{-1}\mathbf{1}_D, \end{cases} \quad k = 0, \ldots, K. \quad (13)$$

Setting $\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \ell_{\mathrm{p}}(\boldsymbol{\beta})$, the maximum likelihood estimator of the size of the jump of the baseline spectral density function turns out to be

$$\widehat{p}_i = \frac{1}{n} \frac{1}{\rho_0 + \widehat{\boldsymbol{\lambda}}_0^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D) + \sum_{k=1}^{K} \exp\{\widehat{\alpha}_k + \widehat{\beta}_k c(\boldsymbol{v}_i)\}\{\rho_k + \widehat{\boldsymbol{\lambda}}_k^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D)\}},$$

so the maximum likelihood estimator of the baseline spectral measure is

$$\widehat{H}_0(\cdot) = \sum_{i=1}^{n} \widehat{p}_i \delta_{\boldsymbol{v}_i}(\cdot)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_{\boldsymbol{v}_i}(\cdot)}{\rho_0 + \widehat{\boldsymbol{\lambda}}_0^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D) + \sum_{k=1}^{K} \exp\{\widehat{\alpha}_k + \widehat{\beta}_k c(\boldsymbol{v}_i)\}\{\rho_k + \widehat{\boldsymbol{\lambda}}_k^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D)\}},$$

(14)

where $\delta_\cdot(\cdot)$ denotes the Dirac measure. The other spectral distributions are estimated as

$$\widehat{H}_k(\cdot) = \sum_{i=1}^{n} \widehat{p}_i \exp\{\widehat{\alpha}_k + \widehat{\beta}_k c(\boldsymbol{v}_i)\} \delta_{\boldsymbol{v}_i}(\cdot)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\exp\{\widehat{\alpha}_k + \widehat{\beta}_k c(\boldsymbol{v}_i)\} \delta_{\boldsymbol{v}_i}(\cdot)}{\rho_0 + \widehat{\boldsymbol{\lambda}}_0^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D) + \sum_{k=1}^{K} \exp\{\widehat{\alpha}_k + \widehat{\beta}_k c(\boldsymbol{v}_i)\}\{\rho_k + \widehat{\boldsymbol{\lambda}}_k^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\mathbf{1}_D)\}},$$

(15)

8

for $k = 1, \ldots, K$.

2.4 Inference

The following result establishes the asymptotic normality of the semiparametric empirical likelihood estimator and a version of Wilks' theorem for the spectral density ratio model; throughout all asymptotic considerations, we assume that $\rho_k = n_k/n \to \rho_k^* \in (0, 1)$, so that every population is non-trivially represented as the total sample size $n = n_0 + \cdots + n_K \to \infty$. Proofs may be found in the Appendix and online Supplementary Material.

Theorem 1. Suppose that the density ratio model applies with data generated using 'true' parameter $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, and that $\int_{S_D} \boldsymbol{v}^{\mathsf{T}} \boldsymbol{v} \, \mathrm{d}H_k(\boldsymbol{v}), \int_{S_D} c(\boldsymbol{v})^2 \, \mathrm{d}H_k(\boldsymbol{v}) < \infty$, for $k = 0, \ldots, K$. Then:

(a) the equation $\partial \ell_\mathrm{p}/\partial \boldsymbol{\beta} = \mathbf{0}$ admits a solution $\widehat{\boldsymbol{\beta}}$ that is weakly consistent for $\boldsymbol{\beta}^*$, and

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \xrightarrow{\mathrm{d}} N(\mathbf{0}, \boldsymbol{\Sigma}^{-1}), \quad n \to \infty,$$

where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution, $\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^K} \ell_\mathrm{p}(\boldsymbol{\beta})$ is the maximum empirical likelihood estimator of $\boldsymbol{\beta}$, and the symmetric positive definite matrix $\boldsymbol{\Sigma}$ is defined in (35); and

(b) the empirical likelihood ratio statistic satisfies $2\{\ell_\mathrm{p}(\widehat{\boldsymbol{\beta}}) - \ell_\mathrm{p}(\boldsymbol{\beta}^*)\} \xrightarrow{\mathrm{d}} \chi_K^2$ as $n \to \infty$.

A corollary gives the limiting distribution of the empirical likelihood ratio statistic for $\boldsymbol{\beta} \in \mathbb{B}$, where $\mathbb{B} = \{\boldsymbol{\beta} \in \mathbb{R}^K : \beta_1 = \cdots = \beta_m = 0\}$ and $m < K$. Although this $\mathbb{B}$ is likely to be the most used in practice, it is easy to check that the result also holds for null hypotheses of the form $\beta_k = \beta_k^0$, for $k = 1, \ldots, m$.

Corollary 2. Under the conditions of Theorem 1 and if $\boldsymbol{\beta}^* \in \mathbb{B}$, then $2\{\ell_\mathrm{p}(\widehat{\boldsymbol{\beta}}) - \ell_\mathrm{p}(\widehat{\boldsymbol{\beta}}_{\mathbb{B}})\} \xrightarrow{\mathrm{d}} \chi_m^2$ as $n \to \infty$, where $\widehat{\boldsymbol{\beta}}_{\mathbb{B}} = \arg\max_{\boldsymbol{\beta} \in \mathbb{B}} \ell_p(\boldsymbol{\beta})$ is the maximum empirical likelihood estimator of $\boldsymbol{\beta}$ restricted to $\mathbb{B}$.

Although Theorem 1 and Corollary 2 assure us that the usual properties of maximum likelihood estimators and likelihood ratio tests apply in large samples, it will often be wiser to base inference on bootstrap resamples. Empirical likelihood estimators and test statistics based on finite samples are

known to deviate further from their asymptotic distributions than do their parametric counterparts, and use of the bootstrap provides some insurance against this.

The one-sample empirical likelihood estimator $\ddot{H}_k$ (Einmahl and Segers, 2009) provides a benchmark to verify whether the specification of the weight function $g(w, \boldsymbol{\psi}_k)$ is appropriate. When modeling bivariate extremes, we suggest following Qin and Zhang (1997) and using the statistics

$$\Delta_k = n^{1/2} \sup_{w \in [0,1]} |\widehat{H}_k(w) - \ddot{H}_k(w)|, \quad k = 0, \dots, K, \tag{16}$$

to assess whether $\widehat{H}_k$ and $\ddot{H}_k$ differ significantly: if so, then this suggests that the weight function is badly chosen. A p-value can be obtained by generating bootstrap samples $\{w_{k,1}^{\star(b)}, \dots, w_{k,n_k}^{\star(b)}\}$ from $g(w, \widehat{\boldsymbol{\psi}}_k) \mathrm{d}\widehat{H}_0$, and using the combined bootstrap sample $V^{\star(b)} = \{v_1^{\star(b)}, \dots, v_n^{\star(b)}\}$ to compute

$$\Delta_k^{\star(b)} = n^{1/2} \sup\{|\widehat{H}_k(w) - \ddot{H}_k(w)| : w \in V^{\star(b)}\}, \quad k = 0, \dots, K, \quad b = 1, \dots, B. \tag{17}$$

We then compare the B bootstrapped statistics $\Delta^{\star(b)}$ with the values of (16) obtained with the combined sample $V = \{v_1, \dots, v_n\}$, i.e., $\Delta_k^{\mathrm{obs}} = \sup\{|\widehat{H}_k(w) - \ddot{H}_k(w)| : w \in V\}$. The p-value for the goodness of fit test is then given, using the indicator function I, as

$$\widehat{P}_k = \frac{1}{B} \sum_{b=1}^{B} I(\Delta_k^{\star(b)} \geq \Delta_k^{\mathrm{obs}}), \quad k = 0, \dots, K.$$

Related approaches are believed to be among the best techniques for testing the tilt model (Cheng et al., 2009, sec. 4), so their performance has been widely studied.

In principle it would be natural to extend (16) to the D-dimensional setting by replacing the sup-norm with the multivariate Kolmogorov–Smirnov statistic of Justel et al. (1997), though its computation seems difficult for $D > 2$.

2.5 Covariates

To avoid complex model selection problems, we introduce covariate modeling 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 tilting parameters has been conducted, we study their association with p covariates of interest

stored in a $K \times p$ design matrix $\boldsymbol{X} = (\boldsymbol{x}_1 - \boldsymbol{x}_0, \ldots, \boldsymbol{x}_K - \boldsymbol{x}_0)^{\mathrm{T}}$; here $\boldsymbol{x}_0$ is the covariate associated to the baseline population, for which $\beta_0 = 0$. Motivated by the asymptotic normality of the semiparametric empirical likelihood estimator $\widehat{\boldsymbol{\beta}}$, we use a simple regression model wherein only $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_K)^{\mathrm{T}} \sim Q$ is considered, so that

$$\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = \boldsymbol{X}\boldsymbol{b}, \quad \boldsymbol{b} = (b_1, \ldots, b_p)^{\mathrm{T}}, \quad p = 1, \ldots, K. \tag{18}$$

Using the data $\{(\boldsymbol{x}_k, \widehat{\boldsymbol{\beta}}_k)\}_{k=0}^K$, we propose to assess the effect of each of the p covariates on the extremal dependence by conducting inference using $\boldsymbol{b}$. Consider the auxiliary variable $Z_k(\boldsymbol{b}) = (\boldsymbol{x}_k - \boldsymbol{x}_0)^{\mathrm{T}}\{\widehat{\boldsymbol{\beta}}_k - (\boldsymbol{x}_k - \boldsymbol{x}_0)^{\mathrm{T}}\boldsymbol{b}\}$ and let $p_k = \mathrm{d}Q(\widehat{\beta}_i)$. The empirical likelihood of $\boldsymbol{b}$ is

$$\ell(\boldsymbol{b}) = \sup\left\{ \sum_{k=0}^K \log p_k : p_k \geq 0, \ \sum_{k=0}^K p_k = 1, \ \sum_{k=0}^K p_k Z_k(\boldsymbol{b}) = 0 \right\}.$$

We choose model (18) for its simplicity but more general models could be used to describe the degree of association between extremal dependence, as measured by the tilting parameters, and the covariates of interest. Inference for the maximum empirical likelihood estimator $\widetilde{\boldsymbol{b}} = \arg\max_{\boldsymbol{b} \in \mathbb{R}^p} \ell(\boldsymbol{b})$ can then be performed using a Wilks' theorem, which suggests a χ^2 calibration with coverage error of order $O(K^{-1})$ (Owen, 1991), though bootstrap calibration is preferable when K is small.

3   SIMULATION STUDY

3.1   Bivariate Spectral Density Ratio Model

We now set up a small simulation study on the properties of our procedures. To model the spectral density corresponding to each subpopulation, we consider a family of symmetric Beta distributions,

$$\mathrm{d}H_k(w) = \frac{1}{B(\phi_k)} w^{\phi_k - 1}(1 - w)^{\phi_k - 1}\mathrm{d}w, \quad \phi_k > 0, \quad k = 0, \ldots, K, \tag{19}$$

where $B(\phi) = \int_0^1 \{u(1 - u)\}^{\phi - 1}\mathrm{d}u$. The mean for each spectral density equals $1/2$, so the constraints (7) are satisfied. Using the distortion function $c(w) = \log\{w(1 - w)\}$ we can rewrite (19) as

$$\begin{cases} \mathrm{d}H_k(w) &= \exp\{a_k + b_k c(w)\}\mathrm{d}w, \\ (a_k, b_k) &= (-\log B(\phi_k), \phi_k - 1), \quad k = 0, \ldots, K. \end{cases}$$

11

Hence, making use of (19), we obtain the following spectral density representation of the $(K+1)$-sample bivariate extreme value beta model,

$$\frac{\mathrm{d}H_k(w)}{\mathrm{d}H_0(w)} = \exp\{\alpha_k + \beta_k c(w)\}, \tag{20}$$

where the tilting parameters are

$$(\alpha_k, \beta_k) = (\log\{B(\phi_0)/B(\phi_k)\}, \phi_k - \phi_0), \quad k = 0, \ldots, K. \tag{21}$$

A consequence of (21) is that $(\alpha_0, \beta_0) = (0,0)$, so this parametrization of the spectral density ratio model (20) is identifiable. This model is closed, since tilting always produces a symmetric beta distribution.

3.2 Numerical Exercise

We now report computational experience with the model described in Section 3.1. We use 1000 simulated data sets with $K = 2$ and $n_0 = n_1 = n_2 = 60$, chosen for comparability with the data in Section 3.3. 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 optimization estimates, we have found it best to use a two-step strategy, wherein one maximizes the unconstrained outer objective function (9) with respect to ψ and then minimizes the inner dual problem with respect to the nuisance parameter λ. For large n the inner dual problem not only has much lower dimensionality than the corresponding primal problem, but also has the advantage of being subject to a set of linear constraints that can be removed using a pseudo-logarithmic function (Owen, 2001, p. 62). A similar two-step estimation strategy is used by Chaudhuri et al. (2007), and a more general discussion on two-step nearly optimal estimators that behave asymptotically 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 (21). The estimates of the exponential tilting 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 optimization estimates. QQ-plots of the estimates, not shown here, confirm the approximate normality of the tilting parameter estimators claimed in Theorem 1. To illustrate the asymptotic results obtained in the previous section

Table 1: Average estimates of tilting parameters and Lagrange multipliers, and their root mean square errors (RMSE), based on 1000 simulations

Tilting parameters	k		
	0	1	2
True values	$(0,0)$	$(1.14, 0.50)$	$(2.94, 1.50)$
Average full optimization estimates	—	$(1.19, 0.53)$	$(2.96, 1.52)$
RMSE	—	$(0.52, 0.24)$	$(0.81, 0.45)$
Average two-step estimates	—	$(1.21, 0.53)$	$(3.08, 1.59)$
RMSE	—	$(0.46, 0.21)$	$(0.76, 0.43)$

Lagrange multipliers	k		
	0	1	2
True values	0	0	0
Average full optimization estimate	0.04	-0.13	0.08
RMSE	2.03	3.03	2.26
Average two-step estimates	0.03	-0.09	0.04
RMSE	1.97	3.01	2.23

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., we tested $\alpha_1 = \beta_1 = 0$. The rejection rate for comparing the empirical likelihood ratio statistic with the 95% quantile of a χ_1^2 distribution was 78.7%, so a bootstrap approach is preferable; to illustrate the performance of the bootstrap-based goodness of fit test in Section 2.4, the Supplementary Material reports the p-values obtained for each of the 1000 simulated data sets.

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 represents a situation where one of the samples has more extremes, from which the others can borrow strength.
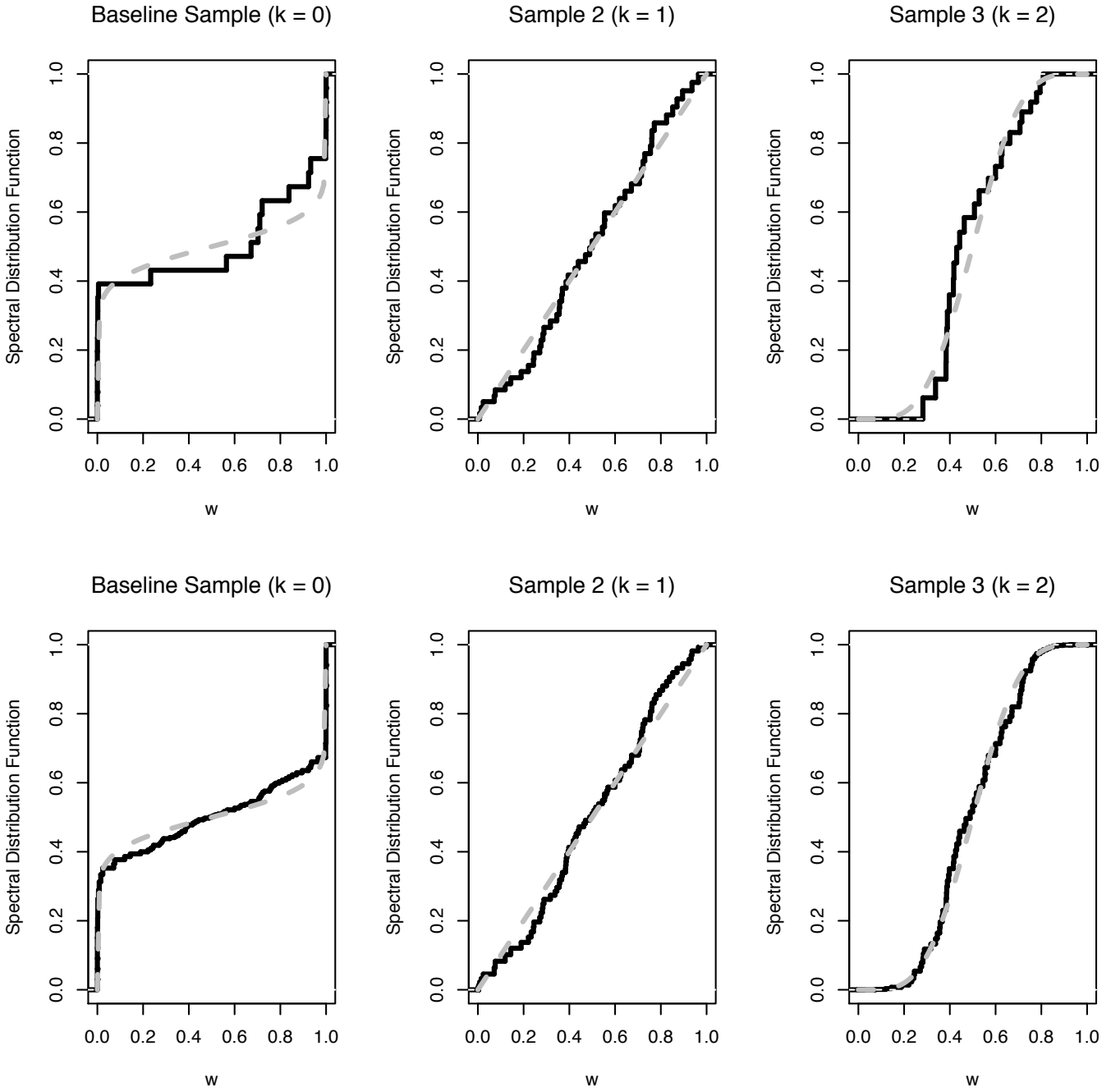
Figure 1: Spectral distribution functions estimated by empirical likelihood (upper panels, solid lines) and using the spectral density ratio model (lower panels, solid line); the dashed lines represent the true spectral distribution functions.
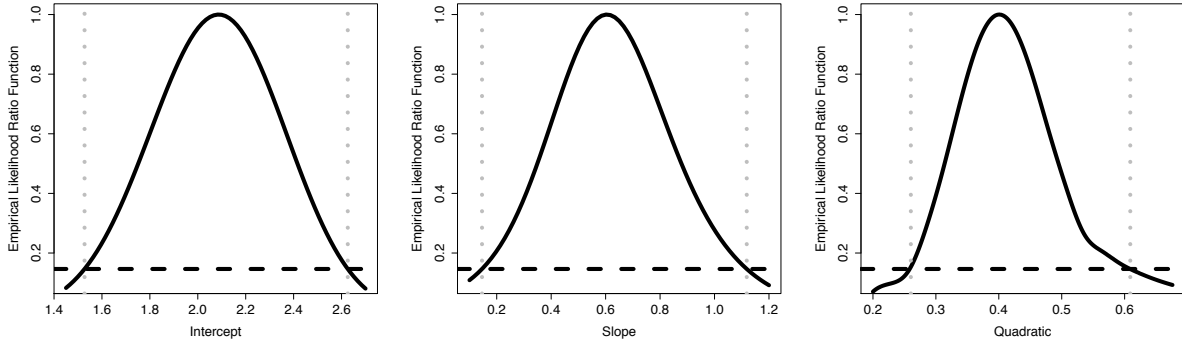
Figure 2: Empirical likelihood ratio functions for the intercept, slope and quadratic effects, for a run of the simulation model considered in Section 3.3. The grey vertical reference lines delimit the 95% empirical likelihood confidence intervals, given by the parameter values for which the empirical likelihood ratio exceeds the horizontal dashed line.

3.3 Regression

We now concentrate on the regression step, considering a simulation model based on the specification

$$\widehat{\beta}_k = \beta_k + \varepsilon_k, \quad \varepsilon_k \overset{\text{iid}}{\sim} N(0,1), \quad k = 0, \ldots, K, \tag{22}$$

with $\beta_k = 2 + 0.5x_k + 0.3x_k^2$; hence, conditional on obtaining $\widehat{\beta}_k$, the estimation target becomes $(b_1, b_2, b_3) = (2, 0.5, 0.3)$. This simple model has connections with that to be applied in Section 3.3, and, motivated by our real data application, we focus on simulating and assessing performance on a setting where the number of samples is small to moderate, with $K+1 = 15, 20, 30$. Figure 2 shows empirical likelihood ratio functions for the intercept, slope and quadratic effects, obtained by simulating 20-tuples of tilting parameters according to (22). The corresponding 95% confidence intervals for b_1 (`intercept`), b_2 (`slope`), and b_3 (`quadratic`) are respectively $[1.53, 2.63]$, $[0.15, 1.12]$, and $[0.26, 0.61]$.

The average interval lengths and coverage probabilities were estimated from 250 independent replicates of $(K+1)$-tuples of tilting parameters simulated according to (22); we focus on the 95% nominal confidence level, but similar conclusions apply for the 90% and 99% nominal confidence levels. Table 2 shows that the average interval length tends to reduce with K, with this reduction being relatively more marked for slope and the quadratic effect. In all cases the coverage probability is lower than the nominal level, but its accuracy tends to increase with K.

Table 2: Average lengths and empirical coverage probabilities (%) for 95% nominal confidence intervals based on 250 random $(K+1)$-tuples of tilting parameters, simulated according to the model considered in Section 3.3.

| Average interval length | Total number of samples $(K+1)$ | | |
	15	20	30
Intercept	1.15	1.05	0.89
Slope	1.04	0.87	0.74
Quadratic	0.83	0.69	0.57
Coverage probabilities	Total number of samples $(K+1)$		
	15	20	30
Intercept	85	92	95
Slope	80	87	88
Quadratic	85	86	89

4 EXTREME TEMPERATURE ANALYSIS CASE STUDY

4.1 Description of the Analysis, Data, and Preprocessing

In this section we describe an application to modeling the dependence between extreme air temperatures under the forest canopy and in a nearby open field at 14 sites in Switzerland (Fig. 3). Our aim is to assess how extremely high temperatures in the open are related to those under the canopy, a topic of interest in forestry and meteorology (Renaud and Rebetez, 2009; Ferrez et al., 2011; Renaud et al., 2011). The raw data consist of two series of air temperatures (in °C) per site, measured in circular metal shelters two meters above ground every 10 minutes since 1997. The slope of the forest ranges from 3% to 80%, the smallest corresponding to Jussy and the largest to Visp, and will be the regressor of interest in Section 4.3. The data are from the Long-term Forest Ecosystem Research database maintained by LWF (Langfristige Waldökosystem-Forschung); see `http://www.wsl.ch`.

We use the same preprocessing steps as in Ferrez et al. (2011) and de Carvalho et al. (2013): we take daily maxima of the residual series resulting from removal of the annual cycle in both location and scale,
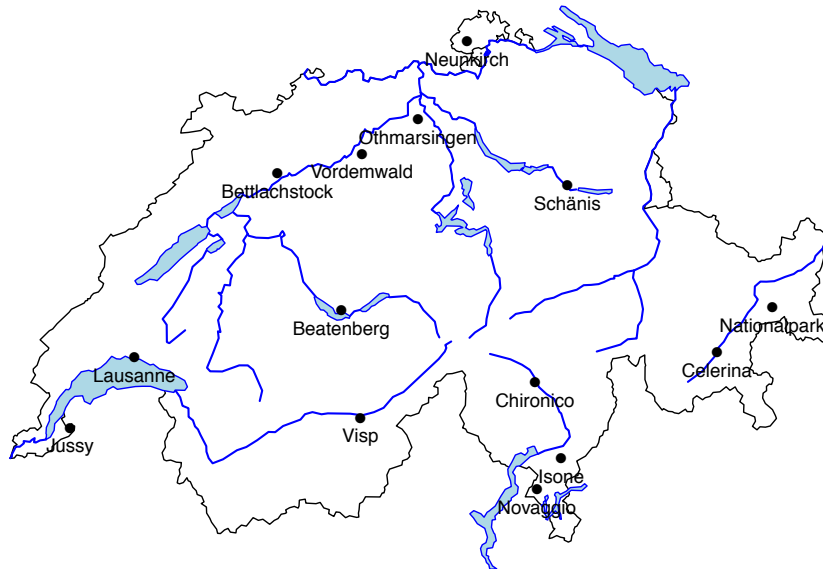
Figure 3: Location of the temperature monitoring stations; for each site, air temperature data under the forest canopy and in a nearby open field are available, and we aim modeling extremal dependence between both temperatures.

i.e., we subtract a periodic mean and divide by a periodic standard deviation. Related preprocessing approaches for seasonal data are discussed in Gong et al. (2011), along with their consequences for modeling extremes. 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. A sensitivity analysis is conducted in Section 4.2 to assess the influence of the threshold on the analysis. We then use the pairs of residuals to construct the samples of pseudo-angles, with the two largest samples corresponding to Novaggio and Visp, of sizes 65 and 64, and with the smallest ones, both of size 45, corresponding to Chironico and Isone. Our model allows us to fit each of the 14 spectral distribution functions using all 785 pseudo-angles from the combined sample; thus for Chironico and Isone we use roughly 17 times more data than with the individual samples.

Data after transformation to unit Fréchet scale, and corresponding threshold boundaries $\{\boldsymbol{y} \in [0, \infty]^2 : y_1 + y_2 = \widehat{F}_{R^k}^{-1}(0.98)\}$, with R^k denoting the pseudo-radius of the kth sample, are reported in the online Supplementary Material.

17

Table 3: Characteristics of sites, estimated tilting parameters and standard errors. Beatenberg is the baseline station

Location	No.	Slope	Estimates		Asymptotic S.E.	
k	n_k	(%)	$\widehat{\alpha}_k$	$\widehat{\beta}_k$	$\widehat{\Sigma}_{\widehat{\alpha}_k}$	$\widehat{\Sigma}_{\widehat{\beta}_k}$
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

4.2 Dependence of Extreme Temperatures

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 3; the baseline sample is Beatenberg. To evaluate the fit of the spectral density ratio model, 1000 bootstrap samples were generated and used to compute (17). All p-values suggest that the spectral density ratio model is appropriate for the data; the smallest p-value is 0.59 for Schänis and the largest is 0.72 for National Park.

The interpretation of the tilting parameters in Table 3 is as follows. 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 $\widehat{\alpha}_k$ and $\widehat{\beta}_k$ display weaker dependence of extremes than the baseline. The estimated
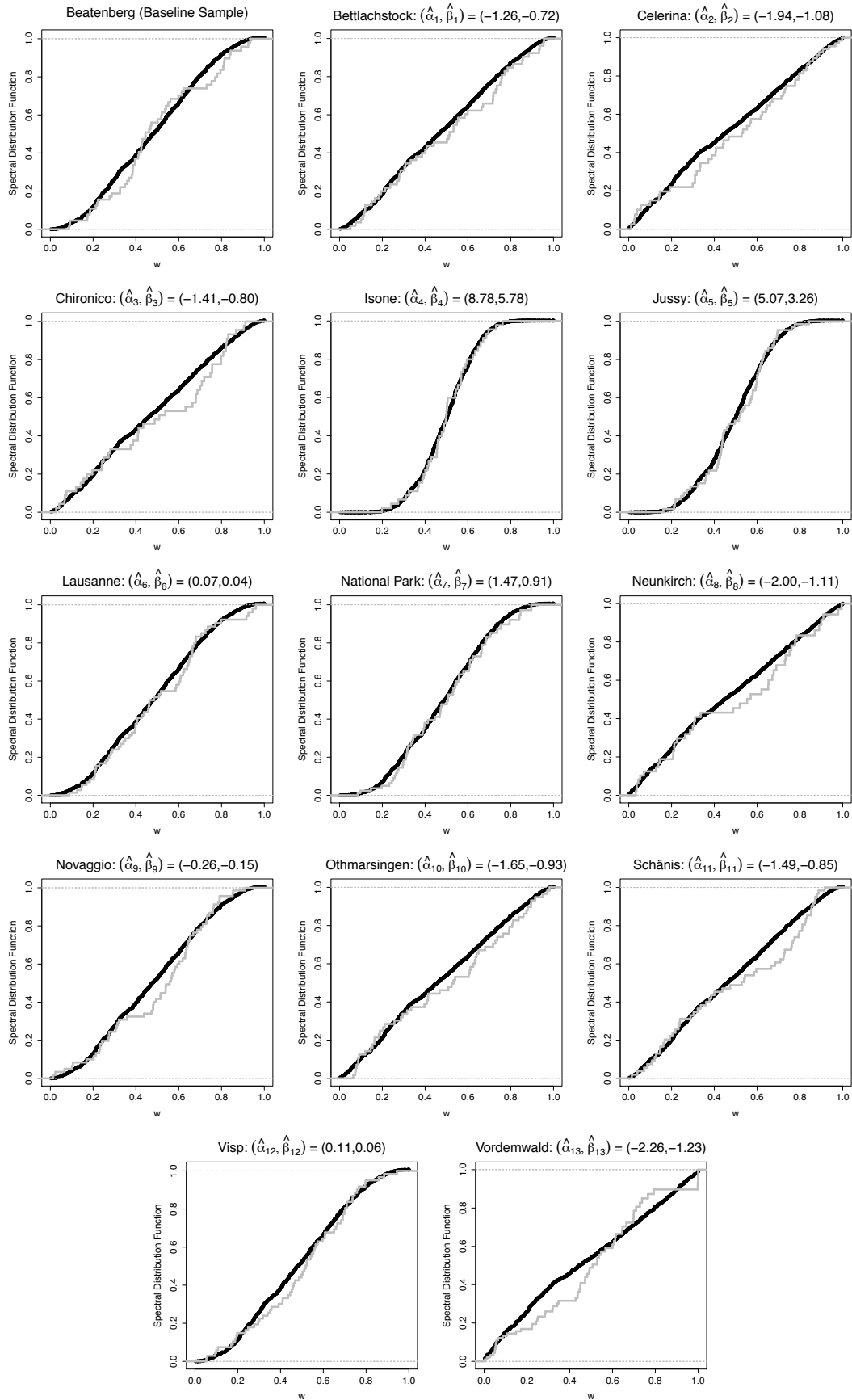
18

Figure 4: Estimated spectral distribution functions obtained using the spectral density ratio model (solid) and empirical likelihood for individual samples (grey).
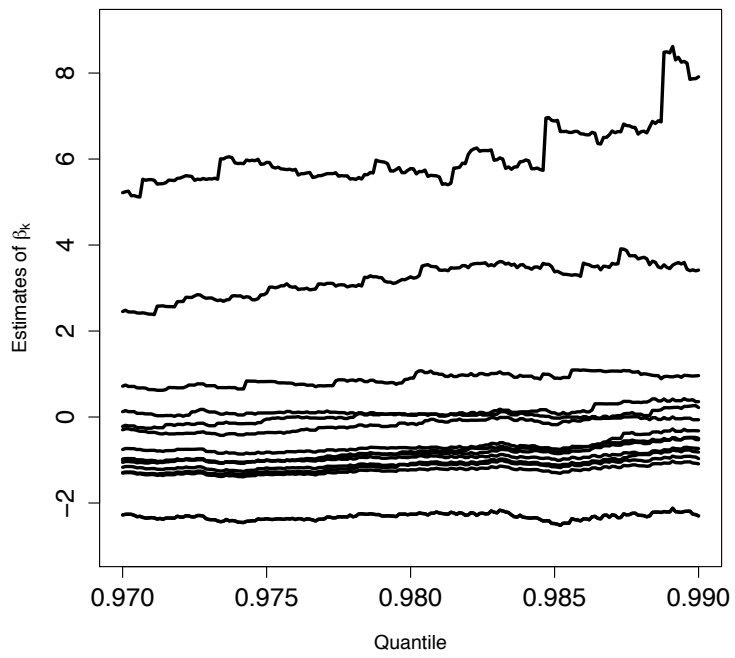
Figure 5: Trajectories of estimated tilting parameters for the 14 temperature monitoring stations obtained by thresholding the data over different quantiles.

spectral distributions are plotted in Figure 4. The more positive the estimated tilting parameters, the more the distribution concentrates around 1/2. Two forests with much larger estimates, Isone and Jussy, correspond to more 'S-shaped' spectral distributions.

To assess the sensitivity of the estimated tilting parameters to the threshold selected in Section 4.1, Figure 5 shows trajectories of the $\widehat{\beta}_k$ computed by thresholding the data at a grid around the 98% quantile. The intersection of a vertical line at the 98% quantile with the trajectories in Figure 5 yields the estimates in Table 3. The 99% quantile is the largest considered, and by thresholding the data at this quantile the number of observations of the combined sample drops from 785 to 397. Due to a bias–variance tradeoff, some trajectories in Figure 5 wiggle more for higher thresholds; this is clearly the case for Isone and Jussy. The number of crossings in Figure 5 is moderate, and thus, apart from minor variation, the ranks of extremal dependence with respect to the baseline remain fairly constant for different thresholds.

4.3   Effect of Slope on Extremal Dependence of Air Temperatures

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. Ferrez et al. (2011) found that slope has a complex impact on the dependence of open-site and below-canopy extreme temperatures, and here we assess this further. A plot of tilting parameters against slope in Figure 6 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 by dropping Isone and Jussy and conducting nonparametric inference by empirical likelihood over the fitted linear model: $\mathrm{E}(\widehat{\beta} \mid \text{slope}) = 0.62 - 7 \times \text{slope}/100 + 7 \times (\text{slope}/100)^2$. A bootstrap calibration yields 95% confidence intervals approximately $(-12, 1)$ and $(-1, 2)$ for the coefficients of the linear and quadratic effects of slope/100, thus casting doubt on the significance of a quadratic association between the sheltering capacity of the forest and its slope. This doubt is strengthened by the undercoverage displayed in the simulation results of Section 3.3.
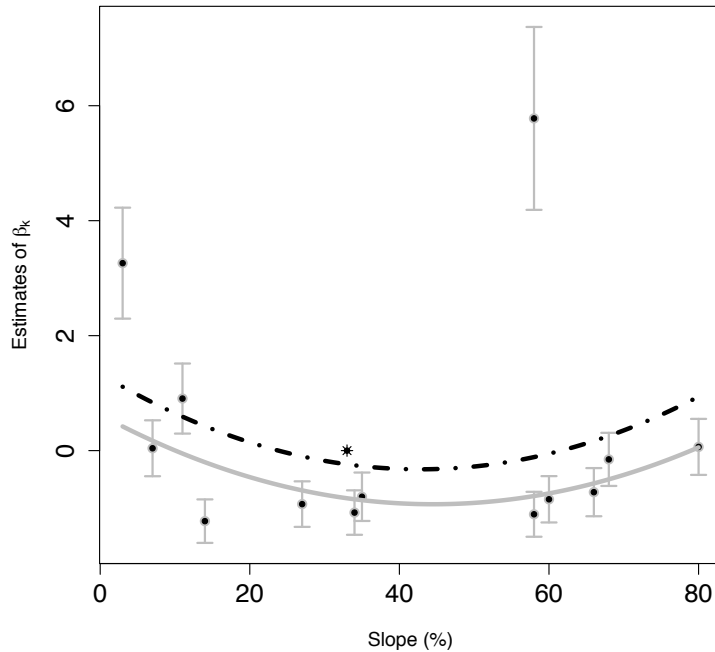
Figure 6: Slope of the forest plotted against estimated tilting parameters. The solid line corresponds to the model without Isone and Jussy; the dashed line corresponds to the model with all forests. The asterisk corresponds to the baseline sample.

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, subject to moment constraints, using a tilting function and regression specification that can 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. Although in theory one should prefer having K as large as possible, so to be able to more accurately describe how extremal dependence evolves over the predictor of interest and to avoid possible numerical convergence issues in the second stage of the approach, a larger K also involves more normalization and moment constraints, and hence it makes

the first-stage optimization harder.

Our theoretical discussion treats the case with a single tilting function $c(\boldsymbol{w})$, but clearly extends to several tilting functions, $g(\boldsymbol{w}; \boldsymbol{\psi}_k) = \exp\{\alpha_k + \boldsymbol{\beta}_k^{\mathrm{T}} \boldsymbol{c}(\boldsymbol{w})\}$. Examination of the argument suggests that the theoretical results will be essentially unchanged with a general function $g(\boldsymbol{w}; \boldsymbol{\psi}_k)$, though the argument will be much messier. Computational aspects however may be appreciably more challenging.

Our development assumes that the underlying data are known to have unit Fréchet marginal distributions, but in most cases a transformation to such margins must be estimated. Such estimation induces dependence along the variables $\boldsymbol{w}_{k,1}, \ldots, \boldsymbol{w}_{k,n_k}$, for each $k \in \{0, \ldots, K\}$, so the asymptotic development in Section 2.4 does not apply directly. In extremal analysis of dependence it is common to ignore the effect of marginal transformation, which is often found to introduce limited additional variation relative to that in the extremal model itself. From a formal mathematical viewpoint, however, this is incorrect: to accommodate this extra uncertainty it would be necessary to modify the loglikelihood (9) to allow for the unknown marginal transformation. While this would complicate the derivations of the theoretical results, some sketch calculations suggest that limiting normality and chi-squared results would continue to apply, though their proofs would be appreciably more complicated. One solution to this is use of the bootstrap, as suggested above: variation in the margins would be included by bootstrapping the original data and re-estimating the marginal transformations and dependence structure. A second solution, used in the analysis in Section 4, is to apply different thresholds when estimating the margins and dependence structure. To be concrete, suppose that the original sample is of size N, and that marginal and dependence fitting for the extremes are based on m_N and d_N observations, respectively. Then consistent estimation requires that $m_N, d_N \to \infty$, and successful application of the extremal models entails $m_N/N, d_N/N \to 0$. If, in addition, $d_N/m_N \to 0$, then in large samples the uncertainty in estimating the marginal transformation will be negligible relative to that due to the dependence estimation, and the theory described in Section 2.4 will apply. Such an approach matches applications, where it is commonly found that the fit of marginal distributions is better than that of the joint part of the model, and so a lower threshold can be used for the margins. In our application, m_N and d_N were taken to correspond to 10% and 2% respectively, so it seems reasonable to apply our theory.

23

APPENDIX

A.1 Profiling the Baseline Spectral Distribution Function

The Lagrangian corresponding to the empirical likelihood optimization problem of interest here is

$$L = \sum_{i=1}^{n} \log p_i + \sum_{k=0}^{K} \sum_{j=1}^{n_k} \{\alpha_k + \beta_k c(\boldsymbol{w}_{k,j})\}$$
$$- n \sum_{k=0}^{K} \eta_k \left[\sum_{i=1}^{n} p_i \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\} - 1 \right] - n \sum_{k=0}^{K} \boldsymbol{\lambda}_k^{\mathrm{T}} \left[\sum_{i=1}^{n} p_i \boldsymbol{v}_i \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\} - D^{-1}\boldsymbol{1}_D \right],$$

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

$$\frac{1}{p_i} - n \sum_{k=0}^{K} \eta_k \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\} - n \sum_{k=0}^{K} \boldsymbol{\lambda}_k^{\mathrm{T}} \boldsymbol{v}_i \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\} = 0, \tag{23}$$

and hence

$$p_i = \frac{1}{n} \frac{1}{\sum_{k=0}^{K} \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\}(\eta_k + \boldsymbol{\lambda}_k^{\mathrm{T}} \boldsymbol{v}_i)}. \tag{24}$$

We now profile with respect to $\boldsymbol{\alpha}$, and thus we set to zero the derivative of L with respect to α_k, and use (10) to obtain $\eta_k = \rho_k - D^{-1}\boldsymbol{\lambda}_k^{\mathrm{T}}\boldsymbol{1}_D$. Thus (24) can be written as

$$p_i = \frac{1}{n} \frac{1}{\sum_{k=0}^{K} \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\}\{\rho_k + \boldsymbol{\lambda}_k^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\boldsymbol{1}_D)\}}.$$

A.2 Auxiliary Function and Lemmas

Following Tan (2009) and Huang and Rathouz (2012), and with $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\lambda})^{\mathrm{T}}$, we define the auxiliary function

$$\kappa(\boldsymbol{\beta}, \boldsymbol{\gamma}) = -\sum_{i=1}^{n} \log \left\{ \sum_{k=0}^{K} \exp\{\alpha_k + \beta_k c(\boldsymbol{v}_i)\}\{\rho_k + \boldsymbol{\lambda}_k^{\mathrm{T}}(\boldsymbol{v}_i - D^{-1}\boldsymbol{1}_D)\} \right\} + \sum_{k=0}^{K} \sum_{j=0}^{n_k} \{\alpha_k + \beta_k c(\boldsymbol{w}_{k,j})\}. \tag{25}$$

Equation (12) yields $\kappa\{\boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta})\} = \ell_{\mathrm{p}}(\boldsymbol{\beta})$, where $\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}) = (\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}), \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}))$ are the joint solutions to (13). Implicit differentiation then gives

$$\frac{\mathrm{d}\ell_{\mathrm{p}}(\boldsymbol{\beta})}{\mathrm{d}\boldsymbol{\beta}} = \left.\frac{\partial\kappa}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta})}, \quad \frac{\mathrm{d}^2\ell_{\mathrm{p}}(\boldsymbol{\beta})}{\mathrm{d}\boldsymbol{\beta}\mathrm{d}\boldsymbol{\beta}^{\mathrm{T}}} = \left.\left\{ \frac{\partial^2\kappa}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}} - \frac{\partial^2\kappa}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}^{\mathrm{T}}} \left(\frac{\partial^2\kappa}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^{\mathrm{T}}}\right)^{-1} \frac{\partial^2\kappa}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\beta}^{\mathrm{T}}} \right\}\right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta})}. \tag{26}$$

The following lemmas are used in the proof of Theorem 1. Lemmas 1 and 2 are essentially versions of Lemmas 2 and 3 in Huang and Rathouz (2012, Supplementary Material, p. 2); details may be found in the online Supplementary Material to the present paper.

Lemma 1. *Let* $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda})^{\mathrm{T}}$ *and* $\boldsymbol{\theta}^* = (\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \mathbf{0})^{\mathrm{T}}$, *and suppose that the conditions of Theorem 1 hold. Then*

$$-\frac{1}{n}\frac{\partial^2 \kappa}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \xrightarrow{p} \boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{\beta\beta} & \boldsymbol{S}_{\beta\alpha} & \boldsymbol{S}_{\beta\lambda} \\ & \boldsymbol{S}_{\alpha\alpha} & \boldsymbol{S}_{\alpha\lambda} \\ & & \boldsymbol{S}_{\lambda\lambda} \end{pmatrix}, \quad n \to \infty,$$

where the matrix $\boldsymbol{S}$ *is symmetric, and, with* $I(\cdot)$ *the indicator function and* $\rho_0 = n_0/n, \ldots, \rho_K = n_K/n$, *we have*

$$
\begin{aligned}
(\boldsymbol{S}_{\beta\beta})_{k,m} &= I(k=m)\rho_k J_k^{cc} - \rho_k\rho_m J_{k,m}^{cc}, \quad (\boldsymbol{S}_{\beta\alpha})_{k,m} = I(k=m)\rho_k J_k^c - \rho_k\rho_m J_{k,m}^c, \\
(\boldsymbol{S}_{\beta\lambda})_{k,m} &= I(k=m)\rho_k J_k^{cv} - \rho_k J_{k,m}^{cv}, \quad (\boldsymbol{S}_{\alpha\alpha})_{k,m} = I(k=m)\rho_k - \rho_k\rho_m J_{k,m}, \\
(\boldsymbol{S}_{\alpha\lambda})_{k,m} &= -\rho_k J_{k,m}^v, \quad (\boldsymbol{S}_{\lambda\lambda})_{k,m} = -J_{k,m}^{vv},
\end{aligned}
$$

for $k, m = 1, \ldots, K$ *when considering* $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ *and* $k, m = 0, \ldots, K$ *when considering* $\boldsymbol{\lambda}$, *where*

$$
\begin{aligned}
J_k^{cc} &= \int_{S_D} c^2(\boldsymbol{v})\exp\{\alpha_k + \beta_k c(\boldsymbol{v})\}\mathrm{d}H_0(\boldsymbol{v}), \\
J_{k,m}^{cc} &= \int_{S_D} c^2(\boldsymbol{v})\frac{\exp\{\alpha_k + \beta_k c(\boldsymbol{v})\}\exp\{\alpha_m + \beta_m c(\boldsymbol{v})\}}{\sum_{l=0}^K \rho_l \exp\{\alpha_l + \beta_l c(\boldsymbol{v})\}}\mathrm{d}H_0(\boldsymbol{v}), \\
J_k^c &= \int_{S_D} c(\boldsymbol{v})\exp\{\alpha_k + \beta_k c(\boldsymbol{v})\}\mathrm{d}H_0(\boldsymbol{v}), \\
J_{k,m}^c &= \int_{S_D} c(\boldsymbol{v})\frac{\exp\{\alpha_k + \beta_k c(\boldsymbol{v})\}\exp\{\alpha_m + \beta_m c(\boldsymbol{v})\}}{\sum_{l=0}^K \rho_l \exp\{\alpha_l + \beta_l c(\boldsymbol{v})\}}\mathrm{d}H_0(\boldsymbol{v}), \\
J_k^{cv} &= \int_{S_D} c(\boldsymbol{v})(\boldsymbol{v} - D^{-1}\mathbf{1}_D)\exp\{\alpha_k + \beta_k c(\boldsymbol{v})\}\mathrm{d}H_0(\boldsymbol{v}), \\
J_{k,m}^{cv} &= \int_{S_D} c(\boldsymbol{v})(\boldsymbol{v} - D^{-1}\mathbf{1}_D)\frac{\exp\{\alpha_k + \beta_k c(\boldsymbol{v})\}\exp\{\alpha_m + \beta_m c(\boldsymbol{v})\}}{\sum_{l=0}^K \rho_l \exp\{\alpha_l + \beta_l c(\boldsymbol{v})\}}\mathrm{d}H_0(\boldsymbol{v}), \\
J_{k,m} &= \int_{S_D} \frac{\exp\{\alpha_k + \beta_k c(\boldsymbol{v})\}\exp\{\alpha_m + \beta_m c(\boldsymbol{v})\}}{\sum_{l=0}^K \rho_l \exp\{\alpha_l + \beta_l c(\boldsymbol{v})\}}\mathrm{d}H_0(\boldsymbol{v}), \\
J_{k,m}^v &= \int_{S_D} (\boldsymbol{v} - D^{-1}\mathbf{1}_D)\frac{\exp\{\alpha_k + \beta_k c(\boldsymbol{v})\}\exp\{\alpha_m + \beta_m c(\boldsymbol{v})\}}{\sum_{l=0}^K \rho_l \exp\{\alpha_l + \beta_l c(\boldsymbol{v})\}}\mathrm{d}H_0(\boldsymbol{v}), \\
J_{k,m}^{vv} &= \int_{S_D} (\boldsymbol{v} - D^{-1}\mathbf{1}_D)(\boldsymbol{v} - D^{-1}\mathbf{1}_D)^{\mathrm{T}}\frac{\exp\{\alpha_k + \beta_k c(\boldsymbol{v})\}\exp\{\alpha_m + \beta_m c(\boldsymbol{v})\}}{\sum_{l=0}^K \rho_l \exp\{\alpha_l + \beta_l c(\boldsymbol{v})\}}\mathrm{d}H_0(\boldsymbol{v}).
\end{aligned}
$$

These expressions are understood to be evaluated at $\alpha_k = \alpha_k^*$, $\beta_k = \beta_k^*$, for $k, m = 1, \ldots, K$.

The conditions of Theorem 1, the Cauchy–Schwarz inequality, and the fact that

$$\frac{\exp\{\alpha_k + \beta_k c(\boldsymbol{v})\} \exp\{\alpha_m + \beta_m c(\boldsymbol{v})\}}{\sum_{l=0}^{K} \rho_l \exp\{\alpha_l + \beta_l c(\boldsymbol{v})\}} \le \rho_m^{-1} \exp\{\alpha_k + \beta_k c(\boldsymbol{v})\}$$

imply that quantities such as $J_{k,m}^v, J_{k,m}^{cc}, J_{k,m}^{cv}$ appearing in Lemma 1 are finite, and continuity implies that they are also finite in a neighbourhood of $\boldsymbol{\theta}^*$.

Lemma 2. *Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda})^{\mathrm{T}}$ and $\boldsymbol{\theta}^* = (\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*, \boldsymbol{0})^{\mathrm{T}}$, and suppose that the conditions of Theorem 1 hold. Then*

$$\frac{1}{\sqrt{n}} \frac{\partial \kappa}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}^*} \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V}), \quad \boldsymbol{V} = \boldsymbol{S} - \boldsymbol{T}_0 - \boldsymbol{T}_1 - \boldsymbol{T}_2, \quad n \to \infty, \tag{27}$$

where $\boldsymbol{S}$ is defined in Lemma 1, and, with $\boldsymbol{R} = diag(\rho_1, \ldots, \rho_K)$,

$$\boldsymbol{T}_0 = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_{\beta\lambda} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_{\alpha\lambda} \\ \boldsymbol{S}_{\lambda\beta} & \boldsymbol{S}_{\lambda\alpha} & 2\boldsymbol{S}_{\lambda\lambda} \end{pmatrix}, \quad \boldsymbol{T}_1 = \boldsymbol{S} \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{S}, \quad \boldsymbol{T}_2 = \rho_0^{-1} \boldsymbol{S} \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1}\boldsymbol{1}^{\mathrm{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{S}.$$

Owing to their unfortunate omission of $\boldsymbol{T}_0$ and some incorrect signs, the decomposition of $\boldsymbol{V}$ in (27) differs from that of Huang and Rathouz (2012, Lemma 3), but (27) also holds in their setting.

Lemma 3. *Let $\boldsymbol{S}$ be as in Lemma 1, but partitioned according to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\lambda})^{\mathrm{T}}$ as*

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{\beta\beta} & \boldsymbol{S}_{\beta\gamma} \\ \boldsymbol{S}_{\gamma\beta} & \boldsymbol{S}_{\gamma\gamma} \end{pmatrix}.$$

Then, provided the necessary inverses exist,

$$(\boldsymbol{I}, -\boldsymbol{S}_{\beta\gamma}\boldsymbol{S}_{\gamma\gamma}^{-1})\boldsymbol{S}(\boldsymbol{I}, -\boldsymbol{S}_{\gamma\gamma}^{-1}\boldsymbol{S}_{\gamma\beta})^{\mathrm{T}} = \boldsymbol{S}_{\beta\beta} - \boldsymbol{S}_{\beta\gamma}\boldsymbol{S}_{\gamma\gamma}^{-1}\boldsymbol{S}_{\gamma\beta}, \tag{28}$$

$$(\boldsymbol{I}, -\boldsymbol{S}_{\beta\gamma}\boldsymbol{S}_{\gamma\gamma}^{-1})\boldsymbol{T}_0(\boldsymbol{I}, -\boldsymbol{S}_{\gamma\gamma}^{-1}\boldsymbol{S}_{\gamma\beta})^{\mathrm{T}} = \boldsymbol{0}, \tag{29}$$

$$(\boldsymbol{I}, -\boldsymbol{S}_{\beta\gamma}\boldsymbol{S}_{\gamma\gamma}^{-1})\boldsymbol{T}_1(\boldsymbol{I}, -\boldsymbol{S}_{\gamma\gamma}^{-1}\boldsymbol{S}_{\gamma\beta})^{\mathrm{T}} = \boldsymbol{0}, \tag{30}$$

$$(\boldsymbol{I}, -\boldsymbol{S}_{\beta\gamma}\boldsymbol{S}_{\gamma\gamma}^{-1})\boldsymbol{T}_2(\boldsymbol{I}, -\boldsymbol{S}_{\gamma\gamma}^{-1}\boldsymbol{S}_{\gamma\beta})^{\mathrm{T}} = \boldsymbol{0}. \tag{31}$$

A.3 Proof of Theorem 1

Proof. Our argument is similar to that of Huang and Rathouz (2012).

(a) We first establish that, as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \frac{\mathrm{d}\ell_{\mathrm{p}}}{\mathrm{d}\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \xrightarrow{\mathrm{d}} N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{32}$$

$$-\frac{1}{n} \frac{\mathrm{d}^2\ell_{\mathrm{p}}}{\mathrm{d}\boldsymbol{\beta}\mathrm{d}\boldsymbol{\beta}^{\mathrm{T}}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \xrightarrow{\mathrm{P}} \boldsymbol{\Sigma}, \tag{33}$$

where $\boldsymbol{\Sigma} = \boldsymbol{S}_{\beta\beta} - \boldsymbol{S}_{\beta\gamma}\boldsymbol{S}_{\gamma\gamma}^{-1}\boldsymbol{S}_{\gamma\beta}$, with $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\lambda})^{\mathrm{T}}$. By the asymptotic theory of M-estimators (van der Vaart, 1998, ch. 5), the equation $\mathbf{0} = \partial\kappa/\partial\boldsymbol{\gamma}|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}$ admits a solution $\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}^*) = \boldsymbol{\gamma}^* + o_{\mathrm{p}}(n^{-1/2})$, and in particular we have the expansion

$$\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}^*) - \boldsymbol{\gamma}^* = -\left\{ \left(\frac{\partial^2\kappa}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^{\mathrm{T}}}\right)^{-1} \frac{\partial\kappa}{\partial\boldsymbol{\gamma}} \right\}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*, \boldsymbol{\gamma}=\boldsymbol{\gamma}^*} + o_{\mathrm{p}}(n^{-1/2}).$$

Taylor expansion of $(\mathrm{d}\ell_{\mathrm{p}}/\mathrm{d}\boldsymbol{\beta})(\boldsymbol{\beta}^*)$ in terms of $\boldsymbol{\gamma}(\boldsymbol{\beta}^*)$ around $\boldsymbol{\beta}^*$ yields

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \frac{\mathrm{d}\ell_{\mathrm{p}}}{\mathrm{d}\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} &= \frac{1}{\sqrt{n}} \left\{ \frac{\partial\kappa}{\partial\boldsymbol{\beta}} + \frac{\partial^2\kappa}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}^{\mathrm{T}}}(\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}^*) - \boldsymbol{\gamma}^*) \right\}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*, \boldsymbol{\gamma}=\boldsymbol{\gamma}^*} + o_{\mathrm{p}}(n^{-1/2}) \\
&= \frac{1}{\sqrt{n}} \left\{ \frac{\partial\kappa}{\partial\boldsymbol{\beta}} - \frac{\partial^2\kappa}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}^{\mathrm{T}}}\left(\frac{\partial^2\kappa}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^{\mathrm{T}}}\right)^{-1} \frac{\partial\kappa}{\partial\boldsymbol{\gamma}} \right\}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*, \boldsymbol{\gamma}=\boldsymbol{\gamma}^*} + o_{\mathrm{p}}(n^{-1/2}) \\
&= \left(\boldsymbol{I}, -\frac{1}{n}\frac{\partial^2\kappa}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}^{\mathrm{T}}}\left(\frac{1}{n}\frac{\partial^2\kappa}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^{\mathrm{T}}}\right)^{-1} \right) \frac{1}{\sqrt{n}}\left(\frac{\partial\kappa}{\partial\boldsymbol{\beta}}, \frac{\partial\kappa}{\partial\boldsymbol{\gamma}}\right)^{\mathrm{T}} + o_{\mathrm{p}}(n^{-1/2}), \\
&\xrightarrow{\mathrm{d}} \left(\boldsymbol{I}, -\boldsymbol{S}_{\beta\gamma}\boldsymbol{S}_{\gamma\gamma}^{-1} \right)\boldsymbol{Z}, \quad n \to \infty, \tag{34}
\end{aligned}
$$

where $\boldsymbol{Z}$ has a $N(\mathbf{0}, \boldsymbol{V})$ distribution, the last expression is evaluated at $\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*$, and we have used Lemmas 1 and 2 and Slutsky's lemma. Expression (32) follows from (34), since Lemma 3 yields

$$
\begin{aligned}
\boldsymbol{\Sigma} &= (\boldsymbol{I}, -\boldsymbol{S}_{\beta\gamma}\boldsymbol{S}_{\gamma\gamma}^{-1})\boldsymbol{V}(\boldsymbol{I}, -\boldsymbol{S}_{\gamma\gamma}^{-1}\boldsymbol{S}_{\gamma\beta})^{\mathrm{T}} \\
&= (\boldsymbol{I}, -\boldsymbol{S}_{\beta\gamma}\boldsymbol{S}_{\gamma\gamma}^{-1})(\boldsymbol{S} - \boldsymbol{T}_0 - \boldsymbol{T}_1 - \boldsymbol{T}_2)(\boldsymbol{I}, -\boldsymbol{S}_{\beta\gamma}\boldsymbol{S}_{\gamma\gamma}^{-1})^{\mathrm{T}} \\
&= \boldsymbol{S}_{\beta\beta} - \boldsymbol{S}_{\beta\gamma}\boldsymbol{S}_{\gamma\gamma}^{-1}\boldsymbol{S}_{\gamma\beta}. \tag{35}
\end{aligned}
$$

In order to establish (33), we note that (26), Lemma 1 and the consistency of $\widehat{\boldsymbol{\gamma}}$ as $n \to \infty$ yield

$$-\frac{1}{n} \frac{\mathrm{d}^2\ell_{\mathrm{p}}}{\mathrm{d}\boldsymbol{\beta}\mathrm{d}\boldsymbol{\beta}^{\mathrm{T}}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} = -\frac{1}{n}\left\{ \frac{\partial^2\kappa}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}} - \frac{\partial^2\kappa}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}^{\mathrm{T}}}\left(\frac{\partial^2\kappa}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^{\mathrm{T}}}\right)^{-1}\frac{\partial^2\kappa}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\beta}^{\mathrm{T}}} \right\}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*, \boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}^*)} \xrightarrow{\mathrm{P}} \boldsymbol{S}_{\beta\beta} - \boldsymbol{S}_{\beta\gamma}\boldsymbol{S}_{\gamma\gamma}^{-1}\boldsymbol{S}_{\gamma\beta}.$$

Now $\widehat{\boldsymbol{\beta}}$ satisfies $\mathbf{0} = \mathrm{d}\ell_\mathrm{p}/\mathrm{d}\boldsymbol{\beta}$ if and only if $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}(\widehat{\boldsymbol{\beta}}))$ satisfies $(\partial\kappa/\partial\boldsymbol{\beta}, \partial\kappa/\partial\boldsymbol{\gamma}) = \mathbf{0}$, and by the asymptotic theory of M-estimators such an equation admits a solution $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}(\widehat{\boldsymbol{\beta}})) = (\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*) + o_\mathrm{p}(n^{-1/2})$. Taylor expansion yields

$$\mathbf{0} = \frac{\mathrm{d}\ell_\mathrm{p}}{\mathrm{d}\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = \frac{\mathrm{d}\ell_\mathrm{p}}{\mathrm{d}\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} + \frac{\mathrm{d}^2\ell_\mathrm{p}}{\mathrm{d}\boldsymbol{\beta}\mathrm{d}\boldsymbol{\beta}^\mathrm{T}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) + o_\mathrm{p}(n^{-1/2}),$$

so that, again by Lemmas 1 and 2 and Slutsky's lemma,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = \left\{ \left(-\frac{1}{n}\frac{\mathrm{d}^2\ell_\mathrm{p}}{\mathrm{d}\boldsymbol{\beta}\mathrm{d}\boldsymbol{\beta}^\mathrm{T}} \right)^{-1} \frac{1}{\sqrt{n}}\frac{\mathrm{d}\ell_\mathrm{p}}{\mathrm{d}\boldsymbol{\beta}} \right\}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} + o_\mathrm{p}(1) \xrightarrow{\mathrm{d}} N(\mathbf{0}, \boldsymbol{\Sigma}^{-1}), \quad n \to \infty. \tag{36}$$

(b) By Taylor expansion, it holds that

$$\ell_\mathrm{p}(\boldsymbol{\beta}^*) = \ell_\mathrm{p}(\widehat{\boldsymbol{\beta}}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\mathrm{T} \frac{\mathrm{d}\ell_\mathrm{p}}{\mathrm{d}\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} + \frac{1}{2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\mathrm{T} \frac{\mathrm{d}^2\ell_\mathrm{p}}{\mathrm{d}\boldsymbol{\beta}\mathrm{d}\boldsymbol{\beta}^\mathrm{T}}\bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) + o_\mathrm{p}(1),$$

and hence by (33), (36) and Slutsky's lemma,

$$2\{\ell_\mathrm{p}(\widehat{\boldsymbol{\beta}}) - \ell_\mathrm{p}(\boldsymbol{\beta}^*)\} = \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\mathrm{T} \left(-\frac{1}{n}\frac{\mathrm{d}^2\ell_\mathrm{p}}{\mathrm{d}\boldsymbol{\beta}\mathrm{d}\boldsymbol{\beta}^\mathrm{T}} \right)\bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) + o_\mathrm{p}(1) \xrightarrow{\mathrm{d}} \chi_K^2, \quad n \to \infty.$$

$\square$

SUPPLEMENTARY MATERIAL

The online supplement to this article contains details on existing methods, further empirical results, and the proofs of Lemmas 1–3 and Corollary 2.

REFERENCES

Ballani, F. and Schlather, M. (2011), "A Construction Principle for Multivariate Extreme Value Distributions," *Biometrika*, 98, 633–645.

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," *Journal of the Royal Statistical Society,* Series B, 69, 217–229.

Chaudhuri, 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," *Journal of the Royal Statistical Society,* Series B, 71, 881–904.

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

Coles, S. G. and Tawn, J. A. (1991), "Modelling Extreme Multivariate Events," *Journal of the Royal Statistical Society,* Series B, 53, 377–392.

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

de Carvalho, M., Oumow, B., Segers, J., and Warchoł, M. (2013), "A Euclidean Likelihood Estimator for Bivariate Tail Dependence," *Communications in Statistics: Theory and Methods*, 42, 1176–1192.

Efron, B. and Tibshirani, R. J. (1996), "Using Specially Designed Exponential Families for Density Estimation," *The Annals of Statistics*, 24, 2431–2461.

Einmahl, J. H., Li, J., and Liu, R. Y. (2009), "Thresholding Events of Extreme in Simultaneous Monitoring of Multiple Risks," *Journal of the American Statistical Association*, 104, 982–992.

Einmahl, J. H. J. and Segers, J. (2009), "Maximum Empirical Likelihood Estimation of the Spectral Measure of an Extreme-value Distribution," *The Annals of Statistics*, 37, 2953–2989.

Embrechts, P., Lambrigger, D. D., and Wüthrich, M. V. (2009), "Multivariate Extremes and the Aggregation of Dependent Risks: Examples and Counter-examples," *Extremes*, 12, 107–127.

Ferrez, J., Davison, A., and Rebetez, M. (2011), "Extreme Temperature Analysis Under Forest Cover Compared to an Open Field," *Agricultural and Forest Meteorology*, 151, 992–1001.

Fokianos, K. (2004), "Merging Information for Semiparametric Density Estimation," *Journal of the Royal Statistical Society,* Series B, 66, 941–958.

Fokianos, K., Kedem, B., Qin, J., and Short, D. A. (2001), "A Semiparametric Approach to the One-way Layout," *Technometrics*, 43, 56–65.

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*, 86, 27–43.

Gong, Z., Kiessler, P., and Lund, R. (2011), "A Prediction-residual Approach for Identifying Rare Events in Periodic Time Series," *Journal of Time Series Analysis*, 32, 407–419.

Heffernan, J. E. and Tawn, J. A. (2004), "A Conditional Approach for Multivariate Extreme Values (with Discussion)," *Journal of the Royal Statistical Society,* Series B, 66, 497–530.

Huang, A. and Rathouz, P. J. (2012), "Proportional Likelihood Ratio Models for Mean Regression," *Biometrika*, 99, 223–229.

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

Justel, A., Peña, D., and Zamar, R. (1997), "A Multivariate Kolmogorov–Smirnov Test of Goodness of Fit," *Statistics & Probability Letters*, 35, 251–259.

Kedem, B., Kim, E.-Y., Voulgaraki, A., and Graubard, B. I. (2009), "Two-dimensional Semiparametric Density Ratio Modeling of Testicular Germ Cell Data," *Statistics in Medicine*, 28, 2147–2159.

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

Owen, A. B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249.

— (1991), "Empirical Likelihood for Linear Models," *The Annals of Statistics*, 19, 1725–1747.

— (2001), *Empirical Likelihood*, Boca Raton: Chapman & Hall.

Pickands, J. (1981), "Multivariate Extreme Value Distributions," *Proceedings of the 43rd Session of the International Statistical Institute*, 859–878.

Qin, J. and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 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. W. (2009), "A New Class of Models for Bivariate Joint Tails," *Journal of the Royal Statistical Society,* Series B, 71, 219–241.

Renaud, V., Innes, J., Dobbertin, M., and Rebetez, M. (2011), "Comparison between open-site and below-canopy climatic conditions in Switzerland for different types of forests over 10 years (1998-2007)," *Theoretical and Applied Climatology*, 105, 119–127.

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

Sabourin, A. and Naveau, P. (2013), "Bayesian Dirichlet Mixture Model for Multivariate Extremes: A Reparametrization," *Computational Statistics and Data Analysis (in press)*.

Tan, Z. (2009), "A Note on Profile Likelihood for Exponential Tilt Mixture Models," *Biometrika*, 96, 229–236.

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.