



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Statistics of Extremes: Challenges and Opportunities

Citation for published version:

de Carvalho, M 2016, Statistics of Extremes: Challenges and Opportunities. in F Longin (ed.), *Extreme Events in Finance: A Handbook of Extreme Value Theory and Its Applications*. Wiley, Hoboken. <https://doi.org/10.1002/9781118650318.ch9>

Digital Object Identifier (DOI):

[10.1002/9781118650318.ch9](https://doi.org/10.1002/9781118650318.ch9)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Extreme Events in Finance: A Handbook of Extreme Value Theory and Its Applications

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



STATISTICS OF EXTREMES: CHALLENGES AND OPPORTUNITIES

M. de Carvalho^{a,*}

^a *Faculty of Mathematics, Pontificia Universidad Católica de Chile, Santiago, Chile*

e-mail: MdeCarvalho@mat.puc.cl

Abstract

In this chapter I provide a personal view on some recent concepts and methods of statistics of extremes, and I discuss challenges and opportunities which could lead to potential future developments.

Keywords: Families of spectral measures; Measure-dependent measure; Nonstationary extremal dependence structures; Proportional tails model; Predictor-dependent spectral measures; Spectral density ratio model; Statistics of extremes.

1 Introduction

My personal experience on discussing concepts of risk and statistics of extremes with practitioners started in 2009 while I was a visiting researcher at the Portuguese Central Bank (*Banco de Portugal*). At the beginning, colleague practitioners were intrigued about the methods I was applying; the questions

*This document is a copy of one of the chapters of the research monograph *Extreme Events in Finance: A Handbook of Extreme Value Theory and its Applications*, edited by François Longin, to be published by Wiley. I would like to thank, without implicating, Holger Rootzén and Ross Leadbetter for helpful comments on the penultimate version of this document, and to François Longin for encouraging discussion group participants of the ESSEC Conference on Extreme Events in Finance to write down their viewpoints. I would like to thank other conference participants including Isabel Fraga Alves, Jan Beirlant, Frederico Caeiro, Ivette Gomes, Serguei Novak, Michał Warchoł, Chen Zhou, among others, for stimulating discussions and for pointing out many fascinating directions for the future of statistics of extremes. The research was partially funded by the Chilean NSF through the Fondecyt project 11121186 "Constrained Inference Problems in Extreme Value Modeling."

were recurrent: “What is the difference between statistics of extremes and survival analysis (or duration analysis)?¹ And why don’t you apply empirical estimators?” The short answer is that when modeling rare catastrophic events, we need to extrapolate beyond observed data—into the tails of a distribution—and standard inference methods often fail to deal with this properly. To see this, suppose that we observe a random sample of losses $L_1, \dots, L_N \stackrel{\text{iid}}{\sim} S_L$, and that we estimate the survivor function $S_L(x) := P(L > x)$, using the empirical survivor function, $\widehat{S}_L(x) := N^{-1} \sum_{i=1}^N I(L_i > x)$, for $x > 0$. Now, suppose that we want to assess what is the probability of observing a loss just ε larger than the maximum observed loss, $M_N := \max\{L_1, \dots, L_N\}$. Obviously, the probability of that event turns out to be zero [$\widehat{S}_L(M_N + \varepsilon) = 0$, for all $\varepsilon > 0$], thus illustrating that the empirical survivor function fails to be able to extrapolate into the right tail of the loss distribution. As put simply by Taleb (2012, p. 46), “the fool believes that the tallest mountain in the world will be equal to the tallest one he has observed.”

In this chapter I resume some viewpoints that I shared with the Discussion Group ‘Future of Statistics of Extremes’ at the ESSEC *Conference on Extreme Events in Finance*, which took place in Royaumont Abbey, France, on December 15–17, 2014,

`extreme-events-in-finance.essec.edu`

and which originated the invitation by the Editor for writing this chapter. My goal is on providing a personal view on some recent concepts and methods of statistics of extremes, and to discuss challenges and opportunities which could lead to potential future developments. The scope is far from encyclopedic, and many other interesting perspectives are found all over this monograph.

In §2, I note that a bivariate extreme-value distribution is an example of what I call here a measure-dependent measure, I briefly review kernel density estimators for the spectral density, and I discuss families of spectral measures. In §3, I argue that the spectral density ratio model (de Carvalho and Davison, 2014), the proportional tails model (Einmahl et al., 2015), and the exponential families for heavy-tailed data (Fithian and Wager, 2015) share similar construction principles; in addition, I discuss en passant a new nonparametric estimator for the so-called scedasis function, which is one of the main estimation targets on the proportional tails model. Comments on potential future developments are

¹In econometrics, survival analysis is also known as duration analysis; see Wooldridge (2010, Chap. 22).

scattered across the chapter and a miscellanea of topics is included in §4.

Throughout this chapter I use the acronym EVD to denote Extreme Value Distribution.

2 Statistics of Bivariate Extremes

2.1 The Bivariate EVD is a Measure-Dependent Measure

Let G be a probability measure on (Ω, \mathcal{A}) , and let Θ be a parameter space. The family $\{G_\theta : \theta \in \Theta\}$ is a statistical model. Obviously not every statistical model is appropriate for modeling risk. As mentioned in §1, candidate statistical models should possess the ability to extrapolate into the tails of a distribution, beyond existing data.

Theorem 1. *If there exist sequences $\{a_n > 0\}$ and $\{b_n\}$ such that $P\{(M_n - b_n)/a_n \leq y\} \rightarrow G_\theta(y)$, as $n \rightarrow \infty$, for some non-degenerate distribution G_θ , then*

$$G_\theta(y) = \exp \left[- \left\{ 1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right\}^{-1/\xi} \right], \quad \theta = (\mu, \sigma, \xi), \quad (1)$$

defined on $\{y : 1 + \xi(y - \mu)/\sigma > 0\}$ where $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$, and $\xi \in \mathbb{R}$.²

See Coles (2001, Theorem 3.1.1). Here, μ and σ are location and scale parameters, while ξ is a shape parameter that determines the rate decay of the tail: $\xi \rightarrow 0$, light-tail (Gumbel); $\xi > 0$, heavy-tail (Fréchet); $\xi < 0$, short-tail (Weibull). The generalized EVD (G_θ in (1)) is a three parameter family which plays an important role in statistics of univariate extremes.

In some cases we want to assess the risk of observing simultaneously large values of two random variables (say, two simultaneous large losses in a portfolio), and the mathematical basis for such modeling is that of statistics of bivariate extremes. In this context, ‘extremal dependence’ is often interpreted as a synonym of risk. Moving from one dimension to two dimensions increases sharply the complexity of models for the extremes. The first challenge one faces when modeling bivariate extremes is that the estimation object of interest is infinite-dimensional, whereas in the univariate case only three parameters (μ, σ, ξ) are needed. The intuition is the following. When modeling bivariate extremes, apart from

²Following the standard convention that for $\xi = 0$, Eq. (1) is to be understood with $\xi \rightarrow 0$.

the marginal distributions we are also interested in the extremal dependence structure of the data, and—as we shall see in Theorem 2—only an infinite-dimensional object is flexible enough to capture the ‘spectrum’ of all possible types of dependence.

Let $(Y_{1,1}, Y_{1,2}), \dots, (Y_{N,1}, Y_{N,2}) \stackrel{\text{iid}}{\sim} F_{Y_1, Y_2}$, where I assume that Y_1 and Y_2 are unit Fréchet $[G_{(1,1,1)}]$ marginally distributed, i.e., $P(Y_1 \leq y) = P(Y_2 \leq y) = \exp(-1/y)$, for $y > 0$. Similarly to the univariate case the classical theory for characterizing the extremal behavior of bivariate extremes is based on block maxima, here given by the componentwise maxima $M_N = (\max\{Y_{i,1}\}_{i=1}^N, \max\{Y_{i,2}\}_{i=1}^N) = (M_{N,1}, M_{N,2})$; note that the componentwise maxima M_N needs not to be a sample point. Similarly to the univariate case, we focus on the standardized maxima, which for Fréchet marginals is given by the standardized componentwise maxima, i.e., $M_N^* = N^{-1}(\max\{Y_{i,1}\}_{i=1}^N, \max\{Y_{i,2}\}_{i=1}^N) = (M_{N,1}^*, M_{N,2}^*)$. Next, I define a special type of statistical model which plays a key role on bivariate extreme value modeling.

Definition 1. *Let \mathcal{F} be the space of all probability measures that can be defined over $(\Omega_0, \mathcal{A}_0)$. If G_H is a probability measure on $(\Omega_1, \mathcal{A}_1)$, for all $H \in \mathcal{H} \subseteq \mathcal{F}$, then we say that G_H is a measure-dependent measure. The family $\{G_H : H \in \mathcal{H}\}$ is said to be a set of measure-dependent measures, if G_H is a measure-dependent measure.*

Remark 1. *Throughout the definitions and theorems presented below, \mathcal{H} denotes the space of all probability measures H which can be defined over $([0, 1], \mathbb{B}_{[0,1]})$, where $\mathbb{B}_{[0,1]}$ is the Borel sigma-algebra on $[0, 1]$, and which obey the mean constraint*

$$\int_{[0,1]} w H(dw) = \frac{1}{2}. \quad (2)$$

What are relevant statistical models for statistics of bivariate extremes? Is there an extension of the generalized EVD for the bivariate setting? The following is a bivariate analogue to Theorem 1.

Theorem 2. *If $P(M_{N,2}^* \leq y_1, M_{N,1}^* \leq y_2) \rightarrow G_H(y_1, y_2)$, as $n \rightarrow \infty$, with G being a non-degenerate distribution function, then*

$$G_H\{(0, y_1) \times (0, y_2)\} := G_H(y_1, y_2) = \exp \left\{ -2 \int_{[0,1]} \max \left(\frac{w}{y_1}, \frac{1-w}{y_2} \right) H(dw) \right\}, \quad y_1, y_2 > 0, \quad (3)$$

for some $H \in \mathcal{H}$.

See [Coles \(2001, Theorem 8.1\)](#). Throughout I refer to G_H as a bivariate EVD. Note the similarities between (1) and (3): both start with an ‘exp,’ but for bivariate EVD $\Theta = \mathcal{H}$, whereas for univariate EVD $\Theta \subseteq \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$. To understand why H needs to be an element of \mathcal{H} , let $y_1 \rightarrow \infty$ or $y_2 \rightarrow \infty$ in (3). Some further comments are in order. First, since (2) is the only constraint on H , neither H nor G_H can have a finite parameterization. Second, a bivariate extreme value distribution G_H is an example of a *measure-dependent measure*, as introduced in [Definition 1](#).

A pseudo-polar transformation is useful for understanding the role of H , which is the so-called spectral measure. Define $(R, W) = (Y_1 + Y_2, Y_1/(Y_1 + Y_2))$, and denote R and W as the radius and pseudo-angle, respectively. If Y_1 is relatively large, then $W \approx 1$; if Y_2 is relatively large, then $W \approx 0$. [de Haan and Resnick \(1977\)](#) have shown that $P(W \in \cdot \mid R > u) \rightarrow H(\cdot)$, as $u \rightarrow \infty$. Thus, when the radius R_i is large, the pseudo-angles W_i are approximately distributed according to H . Perfect (extremal) dependence corresponds to H being degenerate at $1/2$, whereas independence corresponds to H being a binomial distribution function, with half of the mass in 0 and the other half in 1. The spectral probability measure H determines the interactions between joint extremes, and is thus an estimating target of interest; other functionals of the spectral measure are also often used, such as the spectral density $h = dH/dw$ or [Pickands \(1981\)](#) dependence function $A(w) = 1 - w + 2 \int_0^w H(v) dv$, for $w \in [0, 1]$. The cases of extremal independence and extremal dependence respectively correspond to the bivariate EVDs, $G_H(y_1, y_2) = \exp\{-1/y_1 - 1/y_2\}$ and $G_H(y_1, y_2) = \exp\{-\max(1/y_1, 1/y_2)\}$, for $y_1, y_2 > 0$.

2.2 Nonparametric Spectral Density Estimation

In practice, we have to deal with a statistical problem—lack of knowledge on H —and an inference challenge—that is, obtaining estimates which obey the marginal moment constraints, and which define a density on the unit interval. Indeed, as posed by [Coles \(2001, p. 146\)](#) “it is not straightforward to constrain nonparametric estimators to satisfy functional constraints of the type” of [Eq. \(2\)](#). Inference should be conducted by using $n = \sum_{i=1}^N I(Y_{i,1} + Y_{i,2} > u)$ pseudo-angles W_1, \dots, W_n , which are constructed from a sample of size N , thresholding the pseudo-radius at a sufficiently high threshold u . Kernel smoothing estimators for h have been recently proposed by [de Carvalho et al. \(2013\)](#) and are

based on

$$\widehat{h}(w) = \sum_{i=1}^n p_i \beta(w; W_i \nu, (1 - W_i) \nu). \quad (4)$$

Here $\beta(w; a, b)$ denotes the beta density with shape parameters $a, b > 0$, and $\nu > 0$ is a parameter responsible for the level of smoothing, and which can be obtained through cross-validation. Each beta density is centered around a pseudo-angle in the sense that $E(W_i^*) = W_i$, for $W_i^* \sim \text{Beta}(W_i \nu; (1 - W_i) \nu)$. And how can we obtain the probability masses, p_i ? There are at least two options. A simple one is to consider Euclidean likelihood methods (Owen, 2001, pp. 63–66), in which case the vector of probability masses $\mathbf{p} = (p_1, \dots, p_n)$ solves:

$$\begin{aligned} \max_{\mathbf{p} \in \mathbb{R}^n} \quad & -\frac{1}{2} \sum_{i=1}^n (np_i - 1)^2 \\ \text{s.t.} \quad & \sum_{i=1}^n p_i = 1 \\ & \sum_{i=1}^n W_i p_i = 1/2. \end{aligned} \quad (5)$$

By the method of Lagrange multipliers we obtain $p_i = n^{-1} \{1 - (\overline{W} - 1/2) S^{-2} (W_i - \overline{W})\}$, where $\overline{W} = n^{-1} \sum_{i=1}^n W_i$, and $S^2 = n^{-1} \sum_{i=1}^n (W_i - \overline{W})^2$. This yields the following estimator, known as the smooth Euclidean likelihood spectral density

$$\widehat{h}_{\text{Euc}}(w) = \frac{1}{n} \sum_{i=1}^n \{1 - (\overline{W} - 1/2) S^{-2} (W_i - \overline{W})\} \beta(w; W_i \nu, (1 - W_i) \nu). \quad (6)$$

Another option proposed by de Carvalho et al. (2013) is to consider a similar approach to that of Einmahl and Segers (2009), in which case the vector of probability masses $\mathbf{p} = (p_1, \dots, p_n)$ solves the following empirical likelihood (Owen, 2001) problem:

$$\begin{aligned} \max_{\mathbf{p} \in \mathbb{R}_+^n} \quad & \sum_{i=1}^n \log p_i \\ \text{s.t.} \quad & \sum_{i=1}^n p_i = 1 \\ & \sum_{i=1}^n W_i p_i = 1/2. \end{aligned} \quad (7)$$

Again by the method of Lagrange multipliers, the solution is $p_i = [n \{1 + \lambda (W_i - 1/2)\}]^{-1}$, for $i = 1, \dots, n$, where λ is the Lagrange multiplier associated to the second equality constraint in (7), defined implicitly as the solution to the equation

$$\frac{1}{n} \sum_{i=1}^n \frac{W_i - 1/2}{1 + \lambda (W_i - 1/2)} = 0.$$

This yields the following estimator, known as the smooth Empirical likelihood spectral density

$$\hat{h}_{\text{Emp}}(w) = \frac{1}{n} \sum_{i=1}^n \frac{\beta(w; W_i \nu, (1 - W_i) \nu)}{1 + \lambda(W_i - 1/2)}. \quad (8)$$

One can readily construct smooth estimators for the corresponding spectral measures; the smooth Euclidean spectral measure and smooth Empirical likelihood spectral measure are respectively given by

$$\begin{aligned} \hat{H}_{\text{Euc}}(w) &= \frac{1}{n} \sum_{i=1}^n \{1 - (\bar{W} - 1/2) S^{-2}(W_i - \bar{W})\} B(w; W_i \nu, (1 - W_i) \nu), \\ \hat{H}_{\text{Emp}}(w) &= \frac{1}{n} \sum_{i=1}^n \frac{B(w; W_i \nu, (1 - W_i) \nu)}{1 + \lambda(W_i - 1/2)}, \end{aligned}$$

where $B(w; a, b)$ is the regularized incomplete beta function, with $a, b > 0$. By construction both estimators, (6) and (8), obey the moment constraint, so that for example

$$\int_0^1 w \hat{h}_{\text{Euc}}(w) dw = \sum_{i=1}^n p_i \left\{ \frac{\nu W_i}{\nu W_i + \nu(1 - W_i)} \right\} = \sum_{i=1}^n p_i W_i = 1/2. \quad (9)$$

Put differently, realizations of the random probability measures \hat{H}_{Euc} and \hat{H}_{Emp} are elements of \mathcal{H} . Examples of applications of these estimators in finance can be found in Kiriliouk et al. (2015, Fig. 4). At the moment, the large sample properties of these estimators remain unknown.

Other estimators for the spectral measure (obeying (2)) can be found in Boldi and Davison (2007), Guillothe et al. (2011), and Sabourin and Naveau (2014).

2.3 Predictor-Dependent Spectral Measures

Formally, $\{F_x : x \in \mathcal{X}\}$ is a set of predictor-dependent (henceforth pd) probability measures if the F_x are probability measures on $(\Omega, \mathbb{B}_\Omega)$, indexed by a covariate $x \in \mathcal{X} \subseteq \mathbb{R}^p$; here \mathbb{B}_Ω is the Borel sigma-algebra on Ω . Analogously, I define:

Definition 2. *The family $\{H_x : x \in \mathcal{X}\}$ is a set of pd spectral measures if $H_x \in \mathcal{H}$, for all $x \in \mathcal{X}$.*

And why do we care about pd spectral measures? Pd spectral measures allow us to assess how extremal dependence evolves over a certain covariate x , i.e., they allow us to model nonstationary

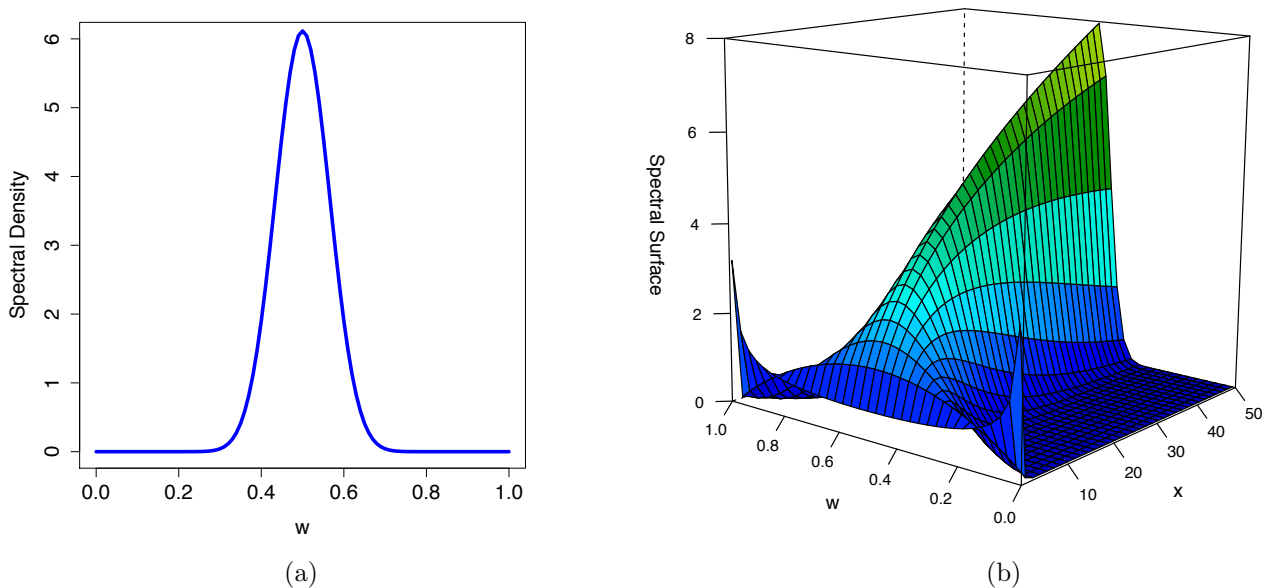


Figure 1: (a) Example of a spectral density. (b) Spectral surface from a predictor-dependent beta family, with $a_x = x$, for $x \in \mathcal{X} = [0.5, 50]$.

extremal dependence structures. Pd spectral measures are a natural probabilistic concept for modeling extremal dependence structures which may change according to a covariate. Indeed, in many settings of applied interest, it seems natural to regard risk from a covariate-adjusted viewpoint, and this leads us to ideas of ‘conditional risk.’ However, if we want to develop ideas of ‘conditional risk’ for bivariate extremes, i.e., if we want to assess systematic variation of risk according to a covariate, we need to allow for nonstationary extremal dependence structures.

To describe how extremal dependence may change over a predictor, I now introduce the concept of spectral surface.

Definition 3. Suppose $H_x \in \mathcal{H}$ is absolutely continuous for all $x \in \mathcal{X}$. The pd spectral density is defined as $h_x = dH_x/dw$, and we refer to the set $\{h_x(w) : w \in [0, 1], x \in \mathcal{X}\}$ as the spectral surface.

A simple spectral surface can be constructed with the pd spectral density $h_x(w) = \beta(w; a_x, a_x)$, where $a : \mathcal{X} \mapsto (0, \infty)$. In Figure 1, I represent a spectral surface based on this model, with $a_x = x$, for $x \in \mathcal{X} = [0.5, 50]$. (Larger values of the predictor x lead to larger levels of extremal dependence.) Other spectral surfaces can be readily constructed from parametric models for the spectral density; see, for instance, Coles (2001, §8.2.1).

Let's now regard the subject of pd bivariate extremes from another viewpoint. Modeling nonstationarity in marginal distributions has been the focus of much recent literature in applied extreme value modeling; see for instance [Coles \(2001, Ch. 6\)](#). The simplest approach in this setting was popularized long ago by [Davison and Smith \(1990\)](#), and it is based on indexing the location and scale parameters of the generalized EVD by a predictor, say by taking

$$G_{(\mu_x, \sigma_x, \xi)}(y) = \exp \left[- \left\{ 1 + \xi \left(\frac{y - \mu_x}{\sigma_x} \right) \right\}^{-1/\xi} \right], \quad x \in \mathcal{X}. \quad (10)$$

And how to model ‘nonstationary bivariate extremes’ if one must? Surprisingly, by comparison to the marginal case, approaches to modeling nonstationarity in the extremal dependence structure have received relatively little attention. These should be important to assess the dynamics governing extremal dependence of variables of interest. For example, has extremal dependence between returns of CAC 40 and DAX 30 been constant over time, or has this level been changing over the years?

By using pd spectral measures we are essentially indexing the parameter of the bivariate extreme value distribution (H) with a covariate, and thus the approach can be regarded as an analogue of the Davison–Smith paradigm in (10), but for the bivariate setting. In the same way that (10) is a covariate-adjusted version of the generalized EVD (1), the following concept can be regarded as a pd version of the bivariate EVD in (3).

Definition 4. *The family $\{G_{H_x} : H_x \in \mathcal{H}\}$ is a set of (measure-dependent) pd bivariate extreme value distributions if for $y_1, y_2 > 0$,*

$$G_{H_x}\{(0, y_1) \times (0, y_2)\} := G_{H_x}(y_1, y_2) = \exp \left\{ - 2 \int_{[0,1]} \max \left(\frac{w}{y_1}, \frac{1-w}{y_2} \right) H_x(dw) \right\}, \quad x \in \mathcal{X}.$$

Similarly to §2.2, in practice we need to obtain estimates which obey the marginal moment constraint, and which define a density on the unit interval, for all $x \in \mathcal{X}$. It is not straightforward to construct nonparametric estimators able to yield valid pd spectral measures. Indeed, any such estimator, \widehat{h}_x , needs to obey the moment constraint, i.e., $\int_0^1 w \widehat{h}_x(w) dw = 1/2$, for all $x \in \mathcal{X}$. [Castro and de Carvalho \(2015\)](#) and [Castro et al. \(2015\)](#) are currently developing models for these contexts, but there are still plenty of opportunities here.³

³A natural option could be on using dependent Bernstein polynomials ([Barrientos et al., 2012](#))—although it may be challenging to impose the moment constraint. It seems conceivable that similar ideas to those in [Guillotte et al. \(2011\)](#) could be used to construct a prior over a family $\{H_x : x \in \mathcal{X}\}$.

Needless to say that other pd objects of interest can be readily constructed. For example, a pd version of [Pickands \(1981\)](#) dependence function can be defined as $A_x(w) = 1 - w + 2 \int_0^w H_x(v) dv$, and a pd $\chi = \lim_{u \rightarrow \infty} P(Y_1 > u \mid Y_2 > u)$ can also be constructed. Using the fact that $\chi = 2 - 2A(1/2)$ ([de Carvalho and Ramos, 2012](#), p. 91) the pd χ can be defined as $\chi_x = 2 - 2A_x(1/2)$, for $x \in \mathcal{X}$.

2.4 Other Families of Spectral Measures

Beyond pd spectral measures other families of spectral measures are of interest. In a recent paper, [de Carvalho and Davison \(2014\)](#) proposed a model for a family of spectral measures $\{H_1, \dots, H_K\}$. The applied motivation for the concept was to track the effect of explanatory variables on joint extremes, i.e., the main concern was on the joint modeling of extremal events when data are gathered from several populations, to each of which corresponds a vector of covariates. Thus, conceptually, there are already in [de Carvalho and Davison \(2014\)](#) some of the ingredients of pd spectral measures and related modeling objectives. Each element in the family, should be regarded as a ‘distorted version,’ of a baseline spectral measure H_0 , in a sense that I will precise below. Formally, spectral density ratio families are defined as follows.

Definition 5. *Let $H_k \in \mathcal{H}$ be absolutely continuous, for $k = 1, \dots, K$. The family $\{H_1, \dots, H_K\}$ is a spectral density ratio family, if there exists an absolutely continuous $H_0 \in \mathcal{H}$, tilting parameters $(\alpha_k, \beta_k) \in \mathbb{R}^2$, and $c : [0, 1] \mapsto \mathbb{R}$ such that*

$$\frac{dH_k}{dH_0}(w) = \exp\{\alpha_k + \beta_k c(w)\}, \quad k = 1, \dots, K. \quad (11)$$

Example 1. Consider a family of symmetric Beta distributions, $dH_k = \beta(w; \phi_k, \phi_k) dw$, for $k = 0, \dots, K$. If $c(w) = \log\{w(1-w)\}$ we can write that $dH_k = \exp\{a_k + b_k c(w)\} dw$, where $(a_k, b_k) = (-\log B(\phi_k), \phi_k - 1)$, with $B(\phi) = \int_0^1 \{u(1-u)\}^{\phi-1} du$. Hence, $dH_k/dH_0 = \exp\{\alpha_k + \beta_k c(w)\}$, where the tilting parameters are $(\alpha_k, \beta_k) = (\log\{B(\phi_0)/B(\phi_k)\}, \phi_k - \phi_0)$. Note that $(\alpha_0, \beta_0) = (0, 0)$, and thus this parametrization is identifiable. This version of the model is closed, since tilting always produces a symmetric beta distribution.

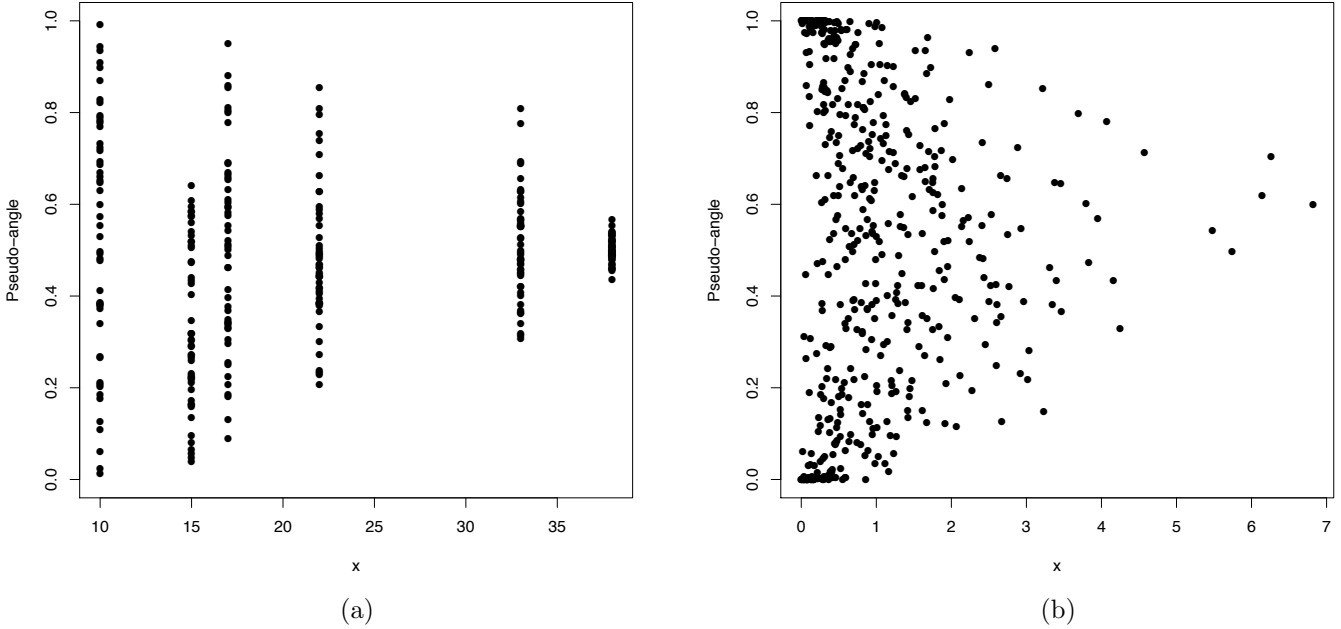


Figure 2: Scatterplots presenting two configurations of data (predictor, pseudo-angles): one (a) where there is sample pseudo-angles per each observed covariate, and another (b) where to each observed covariate may correspond a single pseudo-angle.

From (11), we can write all the normalization and moment constraints for this family as a function of the baseline spectral measure and the tilting parameters, i.e.,

$$\left\{ \begin{array}{ll} \int_0^1 dH_0(w) = 1, & \int_0^1 w dH_0(w) = 1/2, \\ \int_0^1 \exp\{\alpha_1 + \beta_1 c(w)\} dH_0(w) = 1, & \int_0^1 w \exp\{\alpha_1 + \beta_1 c(w)\} dH_0(w) = 1/2, \\ \quad \quad \quad \vdots & \quad \quad \quad \vdots \\ \int_0^1 \exp\{\alpha_K + \beta_K c(w)\} dH_0(w) = 1, & \int_0^1 w \exp\{\alpha_K + \beta_K c(w)\} dH_0(w) = 1/2. \end{array} \right. \quad (12)$$

Inference is based on the combined sample $\{W_{1,0}, \dots, W_{n_0,0}, \dots, W_{1,K}, \dots, W_{n_K,K}\}$ from the spectral distributions H_0, \dots, H_k . Details on estimation and inference through empirical likelihood methods can be found in [de Carvalho and Davison \(2011, 2014\)](#). An extremely appealing feature of their model is that it allows for borrowing strength across samples, in the sense that the estimate of H_k is based on $n = n_0 + \dots + n_K$ pseudo-angles, instead of simply n_k . Although flexible, their approach requires however a substantial computational investment; in particular, inference entails intensive constrained

optimization problems—even for a moderate K —so that estimates of H_k obey empirical versions of the normalization and moment constraints in (12). Their approach allows for modeling extremal dependence in settings such as Fig. 2 (a) but it excludes data configurations such as Fig. 2 (b). The pd-based approach of Castro et al. (2015) allows for inference to be conducted in both settings in Fig. 2.

3 Models Based on Families of Tilted Measures

The main goal of this section is on describing the link between the specifications underlying the spectral density ratio model, discussed in §2.4, the proportional tails model (Einmahl et al., 2015), and the exponential families for heavy-tailed data (Fithian and Wager, 2015).

3.1 Proportional Tails Model

The proportional tails model is essentially an approach for modeling nonstationary extremes. Suppose that at time points $t = 1, \dots, N$ we gather independent observations $Y_1^{(N)}, \dots, Y_N^{(N)}$ respectively sampled from the continuous distribution functions $F_{N,1}, \dots, F_{N,N}$, all with a common right end point $y^* = \sup\{y : F_{N,t}(y) < 1\}$. Suppose further that there exists a (time-invariant) baseline distribution function F_0 , also with right end point y^* , and a continuous function $s : [0, 1] \mapsto [0, \infty)$, such that

$$s\left(\frac{t}{N}\right) := \lim_{y \rightarrow y^*} \frac{1 - F_{N,t}(y)}{1 - F_0(y)}, \quad t = 1, \dots, N. \quad (13)$$

Here s is the so-called scedasis density, and following Einmahl et al. (2015) I assume the following normalization constraint $\int_0^1 s(u) du = 1$. Equation (13) is the key specification of the proportional tails model. Roughly speaking, the scedasis density tells us how much more/less mass there is on the tail $1 - F_{N,t}$, relatively to the baseline tail, $1 - F_0$, for a large y ; uniform scedasis corresponds to a constant frequency of extremes over time.

The question arises naturally: “If the scedasis density provides an indication of the ‘relative frequency’ of extremes over time, it would seem natural that such function could be somehow connected to the intensity measure of the point process characterization of univariate extremes (Coles, 2001, §7.3)?” To have an idea on how the concepts relate I sketch here an heuristic argument. I insist, the argument is heuristic, and my aim here does not go beyond shedding some light on how these ideas

connect. Consider the following artificial setting. Suppose that we could gather a large sample from F_0 , say $\{Y_{1,0}, \dots, Y_{m,0}\}$, and that at each time point we could also collect a large sample from $F_{N,t}$, say $\{Y_{1,t}, \dots, Y_{m,t}\}$, for $t = 1, \dots, N$. For concreteness let's focus on $t = 1$. Then, the definition of scedasis in (13), and similar arguments as in Coles (2001, §4.2.2) suggest that for a sufficiently large y ,

$$s\left(\frac{1}{N}\right) \approx \frac{1 - F_{N,1}(y)}{1 - F_0(y)} \approx \frac{\{1 + \xi(y - \mu_1)/\sigma_1\}^{-1/\xi}}{\{1 + \xi(y - \mu_0)/\sigma_0\}^{-1/\xi}} = \frac{\Lambda_1\{(0, 1) \times (y, \infty)\}}{\Lambda_0\{(0, 1) \times (y, \infty)\}}, \quad (14)$$

where $\Lambda_i\{[t_1, t_2] \times (z, \infty)\} := (t_2 - t_1)\{1 + \xi(z - \mu_i)/\sigma_i\}^{-1/\xi}$, for $i = 0, 1$, is the intensity measure of the limiting Poisson process for univariate extremes (cf Coles, 2001, Theorem 7.1.1). Thus, it can be seen from (14) that in this artificial setting the scedasis density can be literally interpreted as a measure of the relative intensity of the extremes at period $t = 1$, with respect to a (time-invariant) baseline.

Another important question is: ‘‘How can we estimate the scedasis density?’’ Einmahl et al. (2015) propose a kernel-based estimator

$$\hat{s}(w) = \frac{1}{n} \sum_{t=1}^N I(Y_t^{(N)} > Y_{N, N-n}) K_b(w - t/N), \quad w \in (0, 1), \quad (15)$$

where $K_b(\cdot) = (1/b)K(\cdot/b)$, with $b > 0$ being a bandwidth and K being a kernel; in addition, $Y_{N,1} \leq \dots \leq Y_{N,N}$ are the order statistics of $Y_1^{(N)}, \dots, Y_N^{(N)}$. Specifically, Einmahl et al. (2015) recommend K to be a symmetric kernel on $[-1, 1]$. A conceptual problem with using a kernel on $[-1, 1]$ is that it allows for the scedasis density to put mass outside $[0, 1]$.⁴ Using similar ideas to the ones involved in the construction of the smooth spectral density estimators in §2.2, I propose here the following estimator

$$\tilde{s}(w) = \frac{1}{n} \sum_{t=1}^N I(Y_t^{(N)} > Y_{N, N-n}) \beta(w; \nu t/(N+1), \nu\{1 - t/(N+1)\}), \quad w \in (0, 1). \quad (16)$$

Indeed, each beta density is centered close to t/N in the sense that $E(Z_t) = t/(N+1)$, for $Z_t \sim \text{Beta}(\nu t/(N+1); \nu\{1 - t/(N+1)\})$, where ν is the parameter controlling the level of smoothing. My goal here will not be on trying to recommend an estimator over the other, but rather on providing a brief description of strengths and limitations with both approaches. In Fig. 3, I illustrate how the two estimators, (15) and (16), perform on the same data used by Einmahl et al. (2015) and on simulated

⁴For a discussion on the challenges surrounding kernel density estimation on the unit interval see, for instance, Chen (1999), Jones and Henderson (2007), de Carvalho et al. (2013), Geenens (2014), and references therein.

data (single run experiment).⁵ The data consist of daily negative returns of the Standard and Poor’s index from 1988 till 2007 ($N = 5043$), and I use the same value for n (130), the same bandwidth ($h = 0.1$), and (biweight) kernel [$K(y) = 15(1 - y^2)^2/16$, for $y \in [-1, 1]$] as authors; I also follow the authors’s settings for the simulated data. Finally, I consider $\nu = 100$ for illustration.

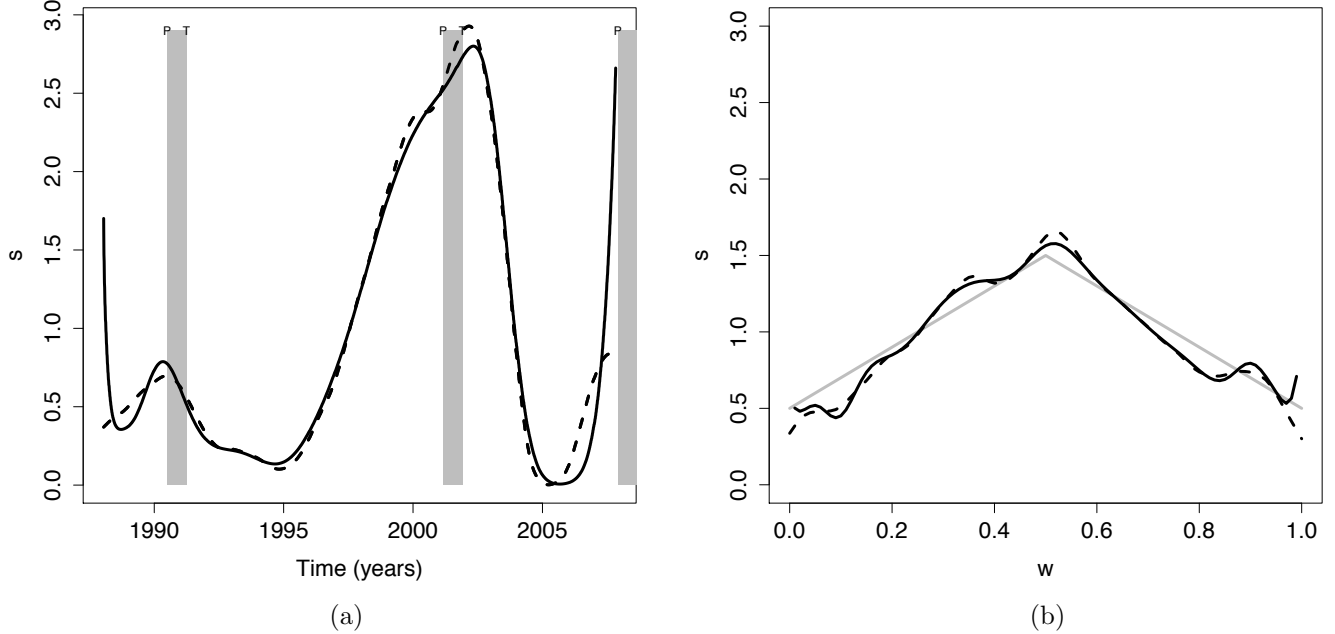


Figure 3: Scedasis density estimates. The solid line represents the beta-kernel estimate from (16), whereas the dashed line represents the estimate from (15). (a) Daily Standard and Poor’s index from 1988 till 2007; the gray rectangles correspond to contraction periods in the US economy. (b) Simulated data illustration from $F_{N,t}(y) = \exp\{-s(t/N)/y\}$, for $y > 0$, with $N = 5000$ and $n = 400$; the grey line represents the true scedasis $s(w) = 2w + 0.5$, for $w \in [0, 0.5)$, and $s(w) = -2w + 2.5$, for $w \in [0.5, 1]$.

In the Standard and Poor’s example in Fig. 3 (a) it can be seen that both estimators capture similar dynamics; the gray rectangles represent contraction periods of the US economy as dated by the National Bureau of Economic Research (NBER). It is interesting to observe that the local maxima of the scedasis density are relatively close to economic contraction periods. Indeed, ‘turning points’ (local maxima and minima) of the scedasis density seem like an interesting estimation target for many

⁵The dashed line in Fig. 3 (a) differs slightly (close to 0 and 1) from Einmahl et al. (2015, Fig. 1), because here I do not use boundary correction.

settings of applied interest.

The estimator in (16) has the appealing feature of putting all mass of the scedasis density inside the $(0, 1)$ interval, and some further numerical experiments suggest that it tends to have a similar behavior to that in (15) except at the boundary. However, a shortcoming with the method in (16) is that it may not be defined at the vertices 0 or 1, and hence it could be inappropriate for forecasting purposes.

The proportional tails model is extremely appealing, and simple to fit. A possible shortcoming is that it does not allow for $\xi > 0$ to change over time. For applications in which we suspect that ξ may change over time, the generalized additive approach by [Chavez-Demoulin and Davison \(2005\)](#) is a sensible alternative and although the model is more challenging to implement, it can be readily fitted with the R package `QRM` by typing in the command `game`.

A problem which seems relevant for practice is that of cluster analysis for the proportional tails model. To see this, suppose that one estimates the scedasis density and tail index for several stocks. It seems natural to wonder: “How can we cluster stocks whose scedasis looks more alike, or—perhaps more interestingly—how can we cluster stocks with a similar scedasis and tail index?”

Lastly, I would like to comment that it seems conceivable that Bernstein polynomials could be used for scedasis density estimation. In particular, a natural question is “Would it be possible to construct a prior over the space of all integrated scedasis functions?” Random Bernstein polynomials could seem like the way to go; see [Petrone \(1999\)](#) and references therein.

3.2 Exponential Families for Heavy-Tailed Data

In this section I sketch some basic ideas on exponential families for heavy-tailed data; I will be more brief here than in §3.1. My goal is mainly on introducing the model specification and to move on; further details can be found in [Fithian and Wager \(2015\)](#).

The starting point for the Fithian–Wager approach is on modeling the conditional right tail law from a population, $G_1^*(y) = P(Y_1 - u \leq y \mid Y_1 > u)$, as an exponential family with carrier measure $G_0^* = P(Y_0 - u \leq y \mid Y_0 > u)$, for a sufficiently large threshold u . Two random samples are assumed to be available, $Y_{1,0}, \dots, Y_{N_0,0} \stackrel{\text{iid}}{\sim} F_0$ and $Y_{1,1}, \dots, Y_{N_1,1} \stackrel{\text{iid}}{\sim} F_1$, with $N_0 \gg N_1$; hence, the applied setting of interest is one where the size of the sample from F_0 is much larger than the one from F_1 . The model

specification is

$$\frac{dG_1^*}{dG_0^*}(y) = \exp\{\eta T(y) - \psi(\eta)\}, \quad y \in [0, \infty), \quad (17)$$

where the sufficient statistic $T(y)$ is of the form $y/(y + \kappa)$ for a certain κ ; the functional form of $T(y)$ is motivated from the case where G_0^* and G_1^* are generalized Pareto distributions (cf [Fithian and Wager, 2015](#), p. 487).

In common with the spectral density ratio model the Fithian–Wager model is motivated by the gains from borrowing strength across samples. Fithian and Wager are not however concerned about spectral measures but rather on estimating a (small-sample) mean of a heavy-tailed distribution, by borrowing information from a much larger sample from a related population with the same $0 < \xi < 1$. More concretely, the authors propose a semiparametric method for estimating the mean of Y_1 , by using the decomposition $\mu = p\mu_L + (1 - p)\mu_R$, where $\mu = E(Y_1)$, $p = P(Y_1 \leq u)$, $\mu_L = E(Y_1 \mid Y_1 \leq u)$, and $\mu_R = E(Y_1 \mid Y_1 > u)$. The Fithian–Wager estimator for the (small-sample) mean can be written as

$$\begin{aligned} \hat{\mu} &= \hat{p} \hat{\mu}_L + (1 - \hat{p}) \hat{\mu}_R \\ &= \frac{n_L}{N_1} \frac{1}{n_L} \sum_{\{i: Y_{i,1} \leq u\}} Y_{i,1} + \frac{n_R}{N_1} \frac{1}{\sum_{\{i: Y_{i,0} > u\}} \exp\{\hat{\eta} T(Y_{i,0} - u)\}} \sum_{\{i: Y_{i,0} > u\}} Y_{i,0} \exp\{\hat{\eta} T(Y_{i,0} - u)\}, \end{aligned} \quad (18)$$

where $n_L = |\{i : Y_{i,1} \leq u\}|$ and $n_R = |\{i : Y_{i,1} > u\}|$, for a large threshold u . Here $\hat{\eta}$ can be computed through a logistic regression with an intercept and predictor $T(y - u)$, as a consequence of results on imbalanced logistic regression ([Owen, 2007](#)). As it can be observed from (18) the main trick on the estimation of μ is on the exponential tilt-based estimator for the mean residual lifetime μ_R .

3.3 Families of Tilted Measures

From previous sections it may have become obvious that the common link underlying the specification of the spectral density ratio model, the proportional tails model, and the exponential families for heavy-tailed data, was the assumption that all members in a family of interest were obtained through a suitable ‘distortion’ of a certain baseline measure. In this section I make this link more precise.

Definition 6. *Let \mathcal{F} be the space of all probability measures that can be defined on (Ω, \mathcal{A}) . Let $g_{i,I} : \Omega \mapsto \mathbb{R}$, for $i = 1, \dots, I$. A family of probability measures in \mathcal{F} , $\{F_1, \dots, F_I\}$, is a g -tilted family if*

there exists $F_0 \in \mathcal{F}$ and a functional θ such that

$$\Phi_\theta(y) := \left(\frac{\theta(F_i)}{\theta(F_0)} \right)(y) = g_{i,I}(y), \quad y \in \Omega.$$

Some examples are presented below.

Example 2 (Spectral density ratio model). For the spectral density ratio model the family of interest is $\{H_1, \dots, H_K\}$ and thus $I = K$. Tilting is conducted through $g_k(w) = \exp\{\alpha_k + \beta_k c(w)\}$, for $w \in (0, 1)$; let $\theta(H) = dH$, for absolutely continuous $H \in \mathcal{H}$. Thus, Eq. (11) can be written as

$$\Phi_\theta(w) := \frac{dH_k}{dH_0}(w) = \exp\{\alpha_k + \beta_k c(w)\} =: g_k(w), \quad w \in (0, 1).$$

Example 3 (Proportional tails model). For the proportional tails model, for a fixed N , the family of interest is $\{F_{N,1}, \dots, F_{N,N}\}$ and hence $I = N$. Tilting is conducted through $g_{t,N}(y) = s(t/N)$, for $y \in \mathbb{R}$; let $\theta(F) = \lim_{y \rightarrow y^*} 1 - F(y)$, with F denoting a continuous distribution function. Thus, Eq. (13) can be rewritten as

$$\Phi_{\theta_{\text{pt}}}(y) := \lim_{y \rightarrow y^*} \frac{1 - F_{N,t}(y)}{1 - F_0(y)} = s\left(\frac{t}{N}\right) =: g_{t,N}(y), \quad y \in \mathbb{R}.$$

Example 4 (Exponential families for heavy-tailed data). In §3.2 the ‘family’ of interest is $\{G_1^*\}$ and thus $I = 1$. Tilting is conducted through $g(y) = \exp\{\eta T(y) + \psi(\eta)\}$, for $y \in [0, \infty)$; let θ to be defined as in Example 2. Thus, Eq. (17) can be rewritten as

$$\Phi_\theta^*(y) := \frac{dG_1^*}{dG_0^*}(y) = \exp\{\eta T(y) - \psi(\eta)\} =: g(y), \quad y \in [0, \infty).$$

4 Miscellanea

Asymptotic (In)Dependence

Here, I comment on the need for further developing models compatible with both asymptotic dependence and asymptotic independence. In two influential papers, [Poon et al. \(2003, 2004\)](#) put forward that asymptotic independence was observed on many pairs of stock market returns. This had important consequences in finance, mostly because inferences in a seminal paper ([Longin and Solnik, 2001](#)) had been based on the assumption of asymptotic dependence, and hence perhaps risk had been overestimated earlier. However, an important question is: “What if pairs of financial losses can move over time

from asymptotic independence to asymptotic dependence, and the other way around?” Some markets are believed to be more integrated these days than in the past, so for such markets it is relevant to ask whether they could have entered an ‘asymptotic dependence regime’? An accurate answer to this question, would require however models able to allow for smooth transitions from asymptotic independence to asymptotic dependence, and vice versa, but as already mentioned in §2.3 at the moment there is a shortage of models for nonstationary extremal dependence structures. [Wadsworth et al. \(2015\)](#) presents an interesting approach for modeling asymptotic (in)dependence.

Spatial Multivariate Extremes

An important reference here is [Genton et al. \(2015\)](#), but there is a wealth of problems to work in this direction, so I stop my comment here.

Dimension Reduction for Multivariate Extremes

Is there a way to reduce dimension in such a way that the interesting features of the data—in terms of tails of multivariate distributions—are preserved?⁶ I think it is fair to say that, apart from some remarkable exceptions, most models for multivariate extremes have been applied only to low-dimensional settings. I remember that at a seminal workshop on high-dimensional extremes, organized by Anthony Davison, at the Ecole Polytechnique Fédérale da Lausanne (14–18 September, 2009), for most talks high-dimensional actually meant ‘two-dimensional,’ and all speakers were top scientists in the field.

Principal Component Analysis (PCA) itself would seem inappropriate, since principal axes are constructed in a way to find the directions that account for most variation, and for our axes of interest (whatever they are...) variation does not seem to be the most reasonable objective? A naive approach could be to use PCA for compositional data ([Jolliffe, 2002](#), §13.3) and apply it to the pseudo-angles themselves. Such approach could perhaps provide a simple way to disentangle dependence into components that could be of practical interest?

⁶An interesting paper on dimension reduction for multivariate extremes appeared in the meantime at the *Electronic Journal of Statistics* ([Chautru, 2015](#)), after the discussion took place. Anne Sabourin and colleagues are also currently working on the topic.

Should the Journal *Extremes* include an *Applications and Case Studies* section?

Theory and methods are the backbone of our field, without regular variation we wouldn't have gone far anyway. But, beyond theory, should our community be investing even more than it already is, in modeling and applications? As put simply by [Box \(1979\)](#), “all models are wrong, but some are useful.” However, while most of us agree that models only provide an approximation to reality, we seem to be very demanding about the way that we develop theory about such—wrong yet useful—models. Some models entail ingenious approximations to reality, and yet are very successful in practice. Should we venture more on this direction in the future? Applied work can also motivate new, and useful, theory. Should we venture more on collaborating with researchers from other fields, or on creating more conferences such as the ESSEC Conference on Extreme Events in Finance, where one has the opportunity to regard risk and extremes from a broader perspective, so to think out of the box? Should the Journal *Extremes* include an *Applications and Case Studies* section?

Communicating Risk and Extremes

What has our community been supplying in terms of communication of risk and extremes? Silence, for the most part. Definitely there have been some noteworthy initiatives, but perhaps mostly from people outside of our field such as those of David Spiegelhalter and David Hand? My own view is that it would be excellent if in a recent future, leading scientists in our field could be more involved in communicating risk and extremes to the general public, either by writing newspaper and magazine articles, or by promoting science vulgarization. Our community is becoming more and more aware of this need, I think. I was happy to see Paul Embrechts showing recently his concern about this matter at EVA 2015 in Ann Arbor.

Prior Elicitation in Contexts where a Conflict of Interest Exists

How can we accurately elicit prior information when modeling extreme events in finance, in cases where a conflict of interest may exist? Suppose that a regulator requires a Bank to report an estimate. If prior information is gathered from a Bank expert—and if the Bank is better off by misreporting—then how can we trust in the accuracy of the inferences? In such cases, I think the only Bayesian analysis a regulator should be willing to accept would be an objective Bayes-based analysis; see [Berger \(2006\)](#)

for a review on objective Bayes.

References

- Barrientos, A. F., Jara, A., and Quintana, F. A. (2012), “Fully Nonparametric Regression for Bounded Data using Dependent Bernstein Polynomials,” Technical report.
- Berger, J. (2006), “The Case For Objective Bayesian Analysis,” *Bayesian Analysis*, 1, 385–402.
- 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.
- Box, G. E. P. (1979), “Some Problems of Statistics and Everyday Life,” *Journal of the American Statistical Association*, 74, 1–4.
- Castro, D., de Carvalho, M., and Wadsworth, J. (2015), “Time-Varying Extremal Dependence with Application to Leading European Stock Markets,” Submitted.
- Castro, D., and de Carvalho, M. (2015), “Spectral Density Regression for Bivariate Extremes,” Preprint.
- Chautru, E. (2015), “Dimension Reduction in Multivariate Extreme Value Analysis,” *Electronic Journal of Statistics*, 9, 383–418.
- Chavez-Demoulin, V., and Davison, A. C. (2005), “Generalized Additive Modelling of Sample Extremes,” *Journal of the Royal Statistical Society, Ser. C*, 54, 207–222.
- Chen, S. X. (1999), “Beta Kernel Estimators for Density Functions,” *Computational Statistics and Data Analysis*, 31, 131–145.
- Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, New York: Springer.
- Davison, A. C., and Smith, R. L. (1990), “Models for Exceedances over High Thresholds (with Discussion),” *Journal of the Royal Statistical Society, Ser. B*, 52, 393–442.
- de Carvalho, M., and Davison, A. C. (2011), “Semiparametric Estimation for K -Sample Multivariate Extremes,” In: *Proceedings 58th World Statistical Congress*, pp. 2961–2969.
- de Carvalho, M., and Davison, A. C. (2014), “Spectral Density Ratio Models for Multivariate Extremes,” *Journal of the American Statistical Association*, 109, 764–776.
- 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.
- de Carvalho, M., and Ramos, A. (2012), “Bivariate Extreme Statistics, II,” *RevStat—Statistical Journal*, 10, 81–104.
- de Haan, L. and Resnick, S. I. (1977), “Limit Theory for Multivariate Sample Extremes,” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40, 317–377.
- 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.

- Einmahl, J. H. J., de Haan, L., and Zhou, C. (2015), “Statistics of Heteroscedastic Extremes,” *Journal of the Royal Statistical Society, Ser. B*, in press (DOI: 10.1111/rssb.12099).
- Fithian, W., and Wager, S. (2015), “Semiparametric Exponential Families for Heavy-Tailed Data,” *Biometrika*, 102, 486–493.
- Geenens, G. (2014), “Probit Transformation for Kernel Density Estimation on the Unit Interval,” *Journal of the American Statistical Association*, 109, 346–359.
- Genton, M. G., Padoan, S. A., and Sang, H. (2015), “Multivariate Max-Stable Spatial Processes,” *Biometrika*, 102, 215–230.
- Guillotte, S., Perron, F., and Segers, J. (2011), “Non-Parametric Bayesian Inference on Bivariate Extremes,” *Journal of the Royal Statistical Society, Ser. B*, 73, 377–406.
- Jolliffe, I. T. (2002), *Principal Component Analysis*, New York: Springer.
- Jones, M. C., and Henderson, D. A. (2007), “Kernel-Type Density Estimation on the Unit Interval,” *Biometrika*, 94, 977–984.
- Kiriliouk, A. Segers, J., and Warchol, M. (2015), “Nonparametric Estimation of Extremal Dependence,” In: *Extreme Value Modelling and Risk Analysis: Methods and Applications*, Eds. Dipak Dey, Jun Yan. Boca Raton, FL: Chapman and Hall/CRC.
- Longin, F., and Solnik, B. (2001), “Extreme Correlation of International Equity Markets,” *The Journal of Finance*, 56, 649–676
- Owen, A. B. (2001), *Empirical Likelihood*, Boca Raton, FL: Chapman and Hall/CRC.
- Owen, A. B. (2007), “Infinitely Imbalanced Logistic Regression,” *Journal of Machine Learning Research*, 8, 761–773.
- Petrone, S. (1999), “Random Bernstein Polynomials,” *Scandinavian Journal of Statistics*, 26, 373–393.
- Pickands, J. (1981), “Multivariate Extreme Value Distributions,” In: *Proceedings 43rd World Statistical Congress*, pp. 859–878.
- Poon, S-H., Rockinger, M., and Tawn, J. (2003), “Modelling Extreme-Value Dependence in International Stock Markets,” *Statistica Sinica*, 13, 929–953.
- Poon, S-H., Rockinger, M., and Tawn, J. (2004), “Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications,” *Review of Financial Studies*, 17, 581–610.
- Sabourin, A., and Naveau, P. (2014), “Bayesian Dirichlet Mixture Model for Multivariate Extremes: A Re-Parametrization,” *Computational Statistics and Data Analysis*, 71, 542–567.
- Taleb, N. (2012), *Antifragile: Things that Gain from Disorder*, New York: Random House.
- Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. M. (2015), “Modelling Across Extremal Dependence Classes,” Under revision.
- Wooldridge, J. M. (2015), *Econometric Analysis of Cross Section and Panel Data*, 2nd Ed., Cambridge, MA: MIT Press.