



PONTIFICIA  
UNIVERSIDAD  
CATÓLICA  
DE CHILE

FACULTAD DE MATEMÁTICAS  
DEPARTAMENTO DE ESTADÍSTICA

# Bivariate Extremes: Modeling, Smoothing, and Regression

BY DANIELA ANDREA CASTRO CAMILO

A dissertation submitted to the Department of Statistics, Faculty of Mathematics, in partial fulfillment of the requirements for the degree of Doctor in Statistics

Advisor: MIGUEL DE CARVALHO

Members of the Examination Committee:

PROF. WILFREDO PALMA - PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

PROF. MANUEL GALEA - PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

PROF. RODRIGO HERRERA - UNIVERSIDAD DE TALCA

PROF. RAPHAËL HUSER - KING ABDULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

August, 2015.

Santiago, Chile.

Copyright ©2015 by Daniela Andrea Castro Camilo.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the author.

# Acknowledgements

---

First and foremost, I would like to express my deepest gratitude to my advisor Miguel de Carvalho for his fundamental role in my doctoral work. Miguel provided the guidance and expertise that I needed to get introduced to the theory of extreme values. I am grateful for his patience, motivation, enthusiasm, and friendship. I am also thankful for all the time we spent talking not only about statistics, but life itself. His advice on both research and life's experiences has been priceless.

I give my utmost thanks to Jennifer Wadsworth, for her guidance, enthusiasm, and for helping me broaden my academic horizons. Additionally, I wish to thank the University of Cambridge for the hospitality shown during my research stay with Jennifer.

I would like to gratefully acknowledge the financial support from Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), through the programs “*Becas para Estudios de Doctorado en Chile*” and the Fondecyt project 11121186 “*Constrained Inference Problems in Extreme Value Modeling*”.

My sincere thanks to the members of the Department of Statistics, for all the years in which they welcomed me. To all the professors who in some way or another we crossed paths or who helped me grow. To Guido del Pino, for all the guidance, patience and help provided. To my fellow students and now colleagues, for their support and friendship during the past years. Specially to Claudia, for sharing my hours of happiness and frustration. I will miss our talks and our very special way of not showing affection. To Maria Soledad and Fabiola, for being an active part of the process, and for the many things they do during the day that are too often unappreciated. (With wisdom and temperance you both rule the roost around here!)

I would also like to express my eternal gratitude to my friends, for their loving care, understanding and support. To Carolina, for reminding me that we can all aspire to be the best version of ourselves.

---

To my wonderful family and relatives, for keeping my feet on the ground and showing me what really matters. To my parents Nelson and Mónica for their unconditional support and for giving me the freedom to pursue my dreams. To my sister Paulina, because being her little sister is one of the best things that ever happened to me.

To my life partner Eduardo: Thank you for your *kindness, patience, encouragement, help, and faith in me. For being my fellow adventurer, my personal angel, and for choosing me, day after day.*

# Contents

---

<b>Acknowledgements</b>	<b>i</b>
<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>ix</b>
<b>Abstract</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background and literature review . . . . .	1
1.1.1 Univariate extreme value theory . . . . .	2
1.1.2 Extremes for nonstationary sequences . . . . .	8
1.1.3 Bivariate extreme value theory . . . . .	10
1.1.4 Regression analysis in bivariate extremes . . . . .	28
1.1.5 Multivariate extreme value theory . . . . .	32
1.2 Motivation . . . . .	33
1.2.1 Nonstationarity in bivariate extreme value distributions . . . . .	33
1.2.2 Nonparametric estimation of the spectral density . . . . .	35
1.3 Thesis outline . . . . .	35
<b>2 Spectral density regression for bivariate extremes</b>	<b>37</b>
2.1 Introduction . . . . .	37
2.2 Spectral density regression model . . . . .	39
2.2.1 Bivariate statistics of extremes and $K$ -sample setting . . . . .	39

---

2.2.2	Predictor-dependent spectral measures . . . . .	41
2.2.3	Double kernel estimator . . . . .	43
2.2.4	Details on implementation . . . . .	45
2.3	Simulation study . . . . .	46
2.3.1	Models, configurations, and preliminary experiments . . . . .	46
2.3.2	Simulation results . . . . .	50
2.4	Extreme forest temperature illustration . . . . .	52
2.4.1	Data description and preprocessing . . . . .	52
2.4.2	Altitude-adjusted extremal dependence . . . . .	54
2.5	Final remarks . . . . .	55
2.6	Appendix . . . . .	56
2.6.1	Appendix A: Monte Carlo mean spectral surfaces . . . . .	56
2.6.2	Appendix A: Additional data application report . . . . .	57
<b>3</b>	<b>Time-Varying Extreme Value Dependence with Application to Leading European Stock Markets</b> . . . . .	<b>59</b>
3.1	Introduction . . . . .	59
3.2	Predictor-dependent modeling for bivariate extremes . . . . .	61
3.2.1	Bivariate statistics of extremes . . . . .	61
3.2.2	Predictor-dependent modeling framework . . . . .	62
3.2.3	Related predictor-dependent objects of interest . . . . .	64
3.3	Estimation and inference . . . . .	65
3.3.1	Derivation of pseudo-angles . . . . .	65
3.3.2	Predictor-dependent spectral density estimation . . . . .	66
3.3.3	Connections to smoothing on the unit interval . . . . .	67
3.3.4	Tuning parameter selection and bootstrap . . . . .	68
3.4	Simulation study . . . . .	70
3.4.1	Data-generating configurations and preliminary experiments . . . . .	70
3.4.2	Simulation results . . . . .	71

---

3.5	Dynamics of joint extremal losses in leading European stock markets . . . . .	75
3.5.1	Background and motivation for empirical analysis . . . . .	75
3.5.2	Data description and exploratory considerations . . . . .	76
3.5.3	Modeling time-varying extremal dependence . . . . .	78
3.6	Final remarks . . . . .	82
3.7	Appendix . . . . .	83
3.7.1	Appendix A: tuning parameter selection . . . . .	83
3.7.2	Appendix B: Monte Carlo mean spectral surfaces . . . . .	84
3.7.3	Appendix D: supplementary data analysis reports . . . . .	85
<b>4</b>	<b>Conclusions and Further Modeling</b>	<b>87</b>
4.1	Conclusions . . . . .	87
4.2	Further Modeling . . . . .	89



## List of Figures

---

1-1	Pickands dependence functions for logistic and Dirichlet models . . . . .	23
1-2	Trajectories of nonparametric estimates of the spectral measure . . . . .	27
2-1	Predictor-dependent beta spectral densities . . . . .	42
2-2	True and estimated cross sections of the spectral surface from the symmetric and asymmetric Dirichlet predictor-dependent models . . . . .	47
2-3	True and estimated symmetric Dirichlet spectral surfaces . . . . .	48
2-4	True and estimated asymmetric Dirichlet spectral surfaces . . . . .	49
2-5	Cross sections of the symmetric and asymmetric Dirichlet spectral surfaces with their corresponding true values and Monte Carlo means . . . . .	51
2-6	Location of air temperature monitoring stations and scatterplots of the pairs of extreme temperatures using an altitude-varying color palette . . . . .	52
2-7	Spectral density estimates for 14 sites in Switzerland using the smoothed Euclidean likelihood estimator . . . . .	53
2-8	Spectral surface estimate using the double kernel estimator . . . . .	55
2-9	True and Monte Carlo mean spectral surfaces for the symmetric Dirichlet predictor-dependent model . . . . .	56
2-10	True and Monte Carlo mean spectral surfaces for the asymmetric Dirichlet predictor-dependent model . . . . .	57
3-1	Spectral surfaces from predictor-dependent beta and logistic families . . . . .	63

---

3-2	True and estimated spectral surfaces for logistic, symmetric and asymmetric Dirichlet predictor-dependent models . . . . .	72
3-3	Cross sections of logistic, symmetric and asymmetric Dirichlet spectral surfaces with their corresponding true values and Monte Carlo means . . . . .	74
3-4	Scatterplots using a time-varying color palette for daily returns for CAC 40, DAX 30, and FTSE 100 . . . . .	76
3-5	Rolling window estimates for $\chi$ and $\bar{\chi}$ for CAC 40, DAX 30, and FTSE 100 . . . . .	78
3-6	Cross sections of spectral surface estimates for CAC–DAX, FTSE–CAC, and FTSE–DAX	80
3-7	Spectral surfaces estimates for CAC–DAX, FTSE–CAC, and FTSE–DAX . . . . .	81
3-8	Time-dependent extremal coefficients . . . . .	82
3-9	True and Monte Carlo mean spectral surfaces for logistic, symmetric and asymmetric Dirichlet predictor-dependent model . . . . .	84
3-10	Daily returns for CAC 40, DAX 30 and FTSE 100 . . . . .	85

## List of Tables

---

2.1	Estimated MIAE for all twelve data-generating configurations discussed in Sections 2.3.1 and 2.3.2 . . . . .	50
2.2	Locations of monitoring stations in 14 sites in Switzerland, along with the number of pseudo-angles ( $n_k$ ) in each site . . . . .	58
3.1	Estimated MIAE for all six data-generating configurations discussed in Section 3.4.1 . .	73
3.2	Summary statistics of the return series for CAC 40, DAX 30, and FTSE 100 . . . . .	85



# Abstract

---

The statistical modeling of extreme events provides a framework to develop techniques and models for describing the unusual rather than the usual. Many problems involving extreme values are inherently bivariate, which mean that we are concerned with the problem of modeling the joint tail of a bivariate distribution. Extreme Value Theory allow us to address this problem by modeling the marginal distributions and the dependence structure separately. In applied extreme value modeling, nonstationarity in marginal distributions has been the focus of much recent literature, but approaches to modeling nonstationarity in the extremal dependence structure have received relatively little attention. Working within a framework of asymptotic dependence, we propose two regression models for the spectral density of a bivariate extreme value distribution that allows us to assess how extremal dependence evolves over a predictor. Considering that the spectral density is a moment-constrained density, we propose methodologies to estimate the predictor-dependent spectral density that produce moment-constrained estimators.

Nonstationarity is analyzed from two extremal dependence prospects. First, in a setting where different dependence structures are related with different values of the predictor, and second, in a setting where the dependence is changing over the predictor. Numerical experiments show that our methods are computationally compelling. Two data applications are given to assess extremal dependence of air temperatures at different altitudes, and to assess the dynamics governing extremal dependence of some leading European stock markets over the last decades. Our empirical analyses allow us to uncover interesting dynamics governing extremal dependence in both cases.



# 1

### 1.1 Background and literature review

The statistical modeling of extreme values has received increasing attention in the statistical literature since it provides a framework to develop techniques and models for describing events that are far from being branded as *usual*. The foundations of statistics of extremes can be traced back to 1928, when Fisher and Tippett settled the possible limit laws of the sample maximum, a theory that was later unified and extended by Gnedenko (1948). Nevertheless, it was not until 1958, when E. J. Gumbel (Gumbel, 1958) published what it has been considered the main referential work for applications of extremes. Later, de Haan (1975) established the probabilistic and stochastic properties of sample extremes, giving rise to important theoretical developments.

Early applications of extreme value models were developed primarily in the field of engineering, but now they have been extended to a wide range of areas, such as environmental impact assessment (Coles and Tawn, 1994; Joe, 1994; de Haan and Ronde, 1998; Schlather and Tawn, 2003; Gardes and Girard, 2010; Dobler *et al.*, 2012; Condon *et al.*, 2015), financial risk (Embrechts *et al.*, 1997; Poon *et al.*, 2004; Liu, 2013), aviation safety (Einhmahl *et al.*, 2009), and internet traffic modeling (Maulik *et al.*, 2002; Resnick and Rootzén, 2000), among others. This variety of applications illustrates how statistics of extremes has made its way into several branches of applied statistics, causing a dramatic acceleration in the developments of suitable methodologies over the past fifty years.

By definition, extreme values are unusual, meaning that estimates are often required for levels of a process that are much higher (or smaller) than what has already been observed. This makes the

extrapolation a necessary tool to provide answers in terms of modeling and prediction. The statistics of extremes supplies a class of models to enable such extrapolation, whether we are interested in representations and modeling techniques for extremes of a single or multiple processes.

In the context of marginal extreme value distributions, nonstationary model structures have been widely studied (Coles, 2001, Chapter 6). The seminal paper of Davison and Smith (1990), popularized the general approach of indexing the parameters of the extreme value distribution by a predictor. This modeling setting can then be used in conjunction with regression techniques, and the results can be extended to deal with serially dependent and seasonal data, which cover most of the practical situations met with univariate series. This progress in terms of nonstationary modeling for marginal distributions has not had a similar impetus in the bivariate setting. Some related results have been provided by Eastoe (2009), Jonathan *et al.* (2014), de Carvalho and Davison (2014) and Huser and Genton (2014), but none of them consider the evolution of the dependence structure as a modeling framework. The gap between developments in nonstationary marginal and bivariate distributions represent the main motivation for the developments made in this thesis, with emphasis on the modeling, smoothing and regression of nonstationary bivariate extremes. In the remainder of this chapter we review some key results in stationary and nonstationary univariate extreme value theory, as well as main concepts and modeling strategies for bivariate extremes. We also outline some key ideas in regression analysis in bivariate extremes, and give a small overview of multivariate extreme value theory.

### 1.1.1 Univariate extreme value theory

#### Asymptotic characterization of block maxima

Let  $\{Y_1, \dots, Y_N\}$  be a random sample from a distribution function  $F$  and assume that we are interested in the distribution of  $M_N = \max_{1 \leq i \leq N} \{Y_i\}$ . Theoretically, the distribution of  $M_N$  can be derived exactly for all values of  $N$ :

$$\begin{aligned} \mathbb{P}\{M_N \leq z\} &= \mathbb{P}\{Y_1 \leq z, \dots, Y_N \leq z\} \\ &= \mathbb{P}\{Y_1 \leq z\} \times \mathbb{P}\{Y_2 \leq z\} \times \dots \times \mathbb{P}\{Y_N \leq z\} \\ &= \{F(z)\}^N. \end{aligned} \tag{1.1}$$

However, this result is not very useful, since the distribution function  $F$  is typically unknown in practice. One way to proceed, is to use standard statistical techniques to estimate  $F$  from observed data, and then substitute this estimate in (1.1). Unfortunately, as  $N$  grows, small discrepancies in the estimate of  $F$  can lead to huge discrepancies for  $F^N$ . Moreover, the distribution of  $M_N$  degenerates to a point mass on  $z_+ = \sup\{z \in \mathbb{R} : F(z) < 1\}$  (the upper end-point of  $F$ ) when  $N$  goes to infinity. Alternatively, we can borrow some ideas from the central limit theorem and look for suitable asymptotic families of models for  $F^N$ . The following theorem, attributed to [Fisher and Tippett \(1928\)](#) and formalized by [Gnedenko \(1948\)](#) is the key result in univariate Extreme Value Theory, and it can be seen as an extreme value version of the central limit theorem.

**Theorem 1** (Extremal types theorem). Let  $\{Y_1, \dots, Y_N\}$  be a random sample from a distribution  $F$ . Assume that there exist constants  $\{a_N > 0\}$  and  $\{b_N\}$  such that

$$\mathbb{P}\left(\frac{M_N - b_N}{a_N} \leq z\right) \rightarrow G(z), \quad N \rightarrow \infty,$$

where  $G$  is a non-degenerate distribution function. Then,  $G$  takes the form of a *generalized extreme value (GEV) distribution*, which is defined by

$$G(z) = \exp\left[-\left\{1 + \xi \left(\frac{z - \mu}{\sigma}\right)\right\}_+^{-1/\xi}\right], \quad -\infty < z < \infty, \quad (1.2)$$

where  $-\infty < \mu < \infty$ ,  $\sigma > 0$ ,  $-\infty < \xi < \infty$  and  $a_+ = \max\{0, a\}$ .

The distribution in (1.2) has three parameters: a location parameter,  $\mu$ ; a scale parameter,  $\sigma$ ; and a shape parameter,  $\xi$ . The GEV combines into a single expression three different distributions, obtained through different values of the shape parameter: the heavy tailed Fréchet distribution is obtained when  $\xi > 0$ , the light tailed Weibull distribution is obtained when  $\xi < 0$ , and the Gumbel distribution with tails decaying exponentially arises when  $\xi = 0$ , which is interpreted as the limit of (1.2) when  $\xi \rightarrow 0$ . When  $F$  has  $G$  as limiting distribution, we say that  $F$  is in the *domain of attraction* of  $G$ . Distributions lying in the domain of attraction of the Weibull distribution are, for example, the uniform and beta distributions; for the Fréchet distribution, we have the inverse and log gamma distributions; and for the Gumbel distribution, we have the normal, gamma, logistic, log normal, and exponential

distributions (Beirlant *et al.*, 2004, Chapter 2).

Theorem 1 is to block maxima what the central limit theorem is to the mean, in the sense that the GEV provides a model for the limit distribution of block maxima. To apply Theorem 1, we need to block the data into blocks of equal length, and fit the GEV distribution to the set of block maxima. To implement this model for any particular dataset, the choice of block size is critical. If the blocks are too small, the approximation by the limit model in Theorem 1 is not suitable, and leads to biased estimates. On the other hand, few block maxima are generated if the blocks are large, leading to large estimation variance. There is a standard practice of adopting blocks of length one year, but this consideration does not always apply and exploratory techniques become necessary. Parameter estimation is typically carried out through likelihood-based techniques, although different graphical and moment-based techniques have also been proposed (Beirlant *et al.*, 2004, Chapter 5). All in all, the utility and adaptability to complex model-building makes the likelihood techniques a particularly attractive approach. A potential difficulty with the use of likelihood methods concerns the regularity conditions that are required for the usual asymptotic properties associated with the maximum likelihood estimator (MLE). Since the end-points of the GEV distribution are functions of the parameter values (for example, if  $\xi < 0$ , then  $\mu - \sigma/\xi$  is an upper end-point), the GEV distribution does not meet the classical regularity conditions. Smith (1985) studied this problem in detail and found the following results:

- if  $\xi > -0.5$ , MLEs have the usual asymptotic properties;
- if  $-1 < \xi < -0.5$ , MLEs are generally obtainable by solving the score equation, but do not have the usual asymptotic properties;
- if  $\xi < -1$ , MLEs are unlikely to be obtainable by solving the score equation.

Although the last case is worrying, this situation rarely occurs in applications of extreme value modeling, so in practice, the theoretical limitations of the maximum likelihood approach are usually no obstacle. Assessing the goodness-of-fit of an extrapolation considering a GEV model is actually impossible, but some appreciation may be made based on the observed data, e.g., using probability and quantile plots. For more details, see Coles (2001, p. 57).

The result in Theorem 1 can be reformulated in terms of the *block minima*,  $\tilde{M}_N = \min_{1 \leq i \leq N} \{Y_i\}$ ,

giving rise to the so-called *GEV distribution for minima*. Note that if  $Z_i = -Y_i$  for  $i = 1, \dots, N$  and  $M_N = \max_{1 \leq i \leq N} \{Z_i\}$ , then  $\tilde{M}_N = -M_N$ , and the use of Theorem 1 leads to a valid asymptotic distribution for  $\tilde{M}_N$ . Inference and model checking for maxima can be applied straightforwardly. More details on asymptotic models for minima can be found in Coles (2001, p. 92).

As we noticed earlier, an implicit difficulty in any extreme value analysis is the limited amount of data for model estimation, which can lead to large estimation variance. This issue motivates a characterization of extreme value behavior that enables the modeling of data other than just block maxima. In the following paragraphs we present an approach to extreme value analysis when other extreme data, lower than the sample maximum, are available.

### Threshold models

Suppose we have a collection of independent variables  $\{Y_1, \dots, Y_N\}$  with a common distribution function  $F$ , and define extreme values as those  $Y_i$  that exceed some high threshold  $u$ . If we let  $Y$  be an arbitrary term in the  $Y_i$  sequence, then the conditional survival function of the exceedances defined as

$$\mathbb{P}\{Y > u + y | Y > u\} = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0, \quad (1.3)$$

is a suitable description of the stochastic behavior of extreme events. As before, since  $F$  is typically unknown in practice, the distribution in (1.3) is also unknown. As in Theorem 1, we look for an approximate family of distributions to characterize the behavior of (1.3) for high values of the threshold. Such a result is stated in the following theorem.

**Theorem 2.** Let  $\{Y_1, \dots, Y_N\}$  be a random sample from a distribution  $F$  and suppose that  $\{a_N > 0\}$ ,  $\{b_N\}$  are sequences of normalizing constants such that Theorem 1 holds. Then, as  $N \rightarrow \infty$

$$\mathbb{P}\left(\frac{Y_i - b_N}{a_N} > y + u \mid \frac{Y_i - b_N}{a_N} > u\right) \rightarrow \left(1 + \frac{\xi y}{\tilde{\sigma}_u}\right)_+^{-1/\xi}, \quad y > 0. \quad (1.4)$$

The limiting distribution in (1.4) is known as the generalized Pareto (GP) distribution. The shape parameter in (1.4) is the same as the one in (1.2), and the new scale parameter  $\tilde{\sigma}_u > 0$  is a function of the threshold  $u$ , the shape parameter  $\xi$ , the GEV location parameter  $\mu$  and the GEV scale parameter

$\sigma$ , specifically,  $\tilde{\sigma}_u = \sigma + \xi(u - \mu)$ .

Evaluation of the normalizing constants  $\{a_N > 0\}$  and  $\{b_N\}$  used in Theorem 1 and 2 requires knowledge of the exact distributional form of  $F$ . However, since the main goal of Extreme Value Theory is to develop inference for the tails that is independent of the underlying distribution, for purposes of inference, the normalizing constants are usually absorbed into the GEV location and scale parameters. If a GP distribution with parameters  $\tilde{\sigma}_u$  and  $\xi$  is a suitable model for exceedances of  $Y$  over a threshold  $u$ , then the result in (1.4) implies that

$$\mathbb{P}(Y > y) \approx \zeta_u \left\{ 1 + \xi \left( \frac{y - u}{\tilde{\sigma}_u} \right) \right\}^{-1/\xi}, \quad (1.5)$$

with  $\zeta_u = \mathbb{P}(Y > u)$ . Theorem 2 states that if the distribution of the block maxima can be approximated by a GEV distribution, then there is a corresponding approximate distribution for threshold exceedances within the GP family. To identify extreme events in this framework, we need to define a high threshold  $u$ . The issue of threshold choice is analogous to the choice of block size in the block maxima approach, in the sense that it implies a balance between bias and variance. Even though the standard practice is to adopt as low a threshold as possible, subject to the limit model providing a reasonable approximation, we are far from a general criterion. Exploratory techniques and sensitivity analysis are helpful tools in this process. Once the threshold selection is made, the parameters of the GP distribution can be estimated by maximum likelihood, using numerical techniques. Standard errors and confidence intervals may be obtained in the usual way from standard likelihood theory. For model checking, we can use the same graphical techniques as in the asymptotic models for block maxima.

### Point process representation

There are different ways of characterizing the extreme value behavior of a process, and a particular useful formulation is derived from the theory of point processes. The importance of this representation is that it provides an interpretation of extreme value behavior that unifies the two models introduced so far. Although this representation is not explicitly used in later chapters, for convenience of the reader we include a very brief and informal introduction on point processes, and provide the main result of this representation. Further details on point processes can be found in and Coles (2001, Chapter 7).

Let  $\{Y_i : i \in \mathcal{J}\}$  represent the locations of points, indexed by a set  $\mathcal{J}$ , occurring randomly in a state space  $S$ . A *point process*  $P$  counts the number of points in regions of  $S$ :

$$P(A) = \sum_{i \in \mathcal{J}} \mathbf{1}(Y_i \in A), \quad A \subset S.$$

The expected number of points in a set  $A$  is given by the *intensity measure*  $\Lambda(A) = \mathbb{E}\{P(A)\}$ . If the state space  $S$  is Euclidean space or a subset thereof and if the intensity measure  $\Lambda$  has a density function  $\lambda : S \rightarrow [0, \infty)$ , then  $\lambda$  is called the *intensity function of the process*.

A point process  $P$  with intensity measure  $\Lambda$  is said to be a *Poisson process* if:

- For each set  $A$  such that  $\Lambda(A) < \infty$ ,  $P(A)$  is a Poisson random variable with mean  $\Lambda(A)$ .
- For all positive integer  $k$  and all disjoint sets  $A_1, \dots, A_k$ , the random variables  $P(A_1), \dots, P(A_k)$  are independent.

A Poisson process on a (subset of) Euclidean space is called *homogenous* if its intensity function  $\lambda$  is constant,  $\lambda(y) = \lambda$ , and *inhomogenous* otherwise. The Poisson processes are the most common type of point processes, and they are particularly relevant in Extreme Value Theory. The following theorem provides the main result of the point process characterization.

**Theorem 3.** Let  $\{Y_1, \dots, Y_N\}$  be a random sample for which there are sequences of constants  $\{a_N > 0\}$  and  $\{b_N\}$  such that Theorem 1 holds. Let  $z_-$  and  $z_+$  be the lower and upper endpoints of the GEV distribution in (1.2), respectively. Then, the sequence of point processes

$$P_N = \left\{ \left( \frac{i}{N+1}, \frac{Y_i - b_N}{a_N} \right) : i = 1, \dots, N \right\}$$

converges on regions of the form  $(0, 1) \times [u, \infty)$ , for any  $u > z_-$ , to a Poisson process, with intensity measure on  $A = [t_1, t_2] \times [z, z_+)$  given by

$$\Lambda(A) = (t_2 - t_1) \left\{ 1 + \xi \left( \frac{z - u}{\sigma} \right) \right\}^{-1/\xi}.$$

A main contribution of this thesis is the development of methodologies for modeling nonstationarity in the extremal dependence structure. Therefore, a revision of terminology and methods developed for

nonstationary sequences is essential. We devote the following section to a revision of the main results for univariate nonstationary processes.

### 1.1.2 Extremes for nonstationary sequences

Roughly speaking, we say that a random process is nonstationary when its probabilistic characteristics change over time. Such departures from the simple i.i.d. assumption made in the derivation of the extreme value characterization, question the usefulness of the standard models. Even though under certain circumstances we are able to apply the usual extreme value limiting models in the presence of temporal dependence (Coles, 2001, p. 92), no such general theory has been provided for nonstationary processes. In this setting, the usual approach is to use standard extreme value models as building blocks, and then make improvements using different statistical techniques.

Consider a GEV model for  $\{Y_t\}_{t \geq 1}$  with temporal structure in the location and scale parameters, i.e.,

$$Y_t \sim \text{GEV}(\mu(t), \sigma(t), \xi), \quad t \geq 0. \quad (1.6)$$

The extreme value shape parameter is difficult to estimate with precision, so it is usually unrealistic to try modeling  $\xi$  as a smooth function of time. Variations through time in the observed process may be modeled as a linear trend in the location parameter:

$$\mu(t) = \beta_0 + \beta_1 t,$$

for parameters  $\beta_0$  and  $\beta_1$ , or through a nonstationary scale parameter, e.g.,

$$\sigma(t) = \exp(\beta_0 + \beta_1 t),$$

where we select an exponential link function to ensure the positivity of  $\sigma$ .

There is a unity of structure in the above examples. In each case, the extreme value parameters can

be written in the form

$$\theta(t) = h(\mathbf{Y}^T \boldsymbol{\beta}), \quad (1.7)$$

where  $\theta$  denotes either  $\mu$  or  $\sigma$ ,  $h$  is a specific link function,  $\boldsymbol{\beta}$  is a vector of regression parameters, and  $\mathbf{Y}$  is a vector of covariates. We can see some similarities between the class of models implied by (1.7) and the generalized linear models (GLMs) introduced by [Nelder and Wedderburn \(1972\)](#). The theory for GLMs is well developed and estimating algorithms are routinely provided in statistical softwares. Unfortunately, standard results or computational tools are not directly transferable to the extreme value context. The main difference is that the GLM family is restricted to distributions that are within the exponential family of distributions; the standard extreme value models generally fall outside of this family. Nonetheless, (1.7) applied to any of the parameters in a extreme value model provides a broad and attractive family for representing nonstationarity in extreme value datasets.

Inference for nonstationary GEV models such as (1.6) can be conducted as follows. Parameter estimation may be carried out through maximum likelihood, due to its adaptability to changes in model structure. If  $y_1, \dots, y_m$  is a random sample from (1.6) and  $\boldsymbol{\beta}$  denotes the associated complete vector of parameters, then the likelihood function is

$$L(\boldsymbol{\beta}) = \prod_{t=1}^m g(y_t; \mu(t), \sigma(t), \xi), \quad (1.8)$$

where  $g(\cdot; \mu(t), \sigma(t), \xi)$  denotes the GEV density function with parameters  $\mu(t), \sigma(t)$  and  $\xi$ . Numerical techniques can be used to maximize (1.8), yielding the maximum estimator of  $\boldsymbol{\beta}$ . Approximate standard errors and confidence intervals follow in the usual way from the observed information matrix, which can also be evaluated numerically. For the model choice, the basic principle is parsimony, i.e., obtaining the simplest model possible, that explains as much of the variation in the data as possible. With the possibility of modeling any combination of the extreme value model parameters as functions of time, there is a large catalogue of models to choose from. Nested models can be compared using the deviance statistic, whereas non-nested models can be tested using a variety of modifications to the deviance-based criterion (see for instance, [Cox and Hinkley, 1974](#)). Finally, model diagnostics can be

conducted similarly to the stationary case, although generally it is only possible to apply such diagnostic checks to a standardized version of the data, conditional on the fitted parameter values. For more details and examples about inference on extremes of nonstationary sequences, see [Coles \(2001, p. 108\)](#).

Although univariate extreme value theory offers a handy framework to develop inference for extreme of a single process, many problems involving extreme events are inherently multivariate. We now turn our attention to the bivariate setting, on which this thesis is devoted.

### 1.1.3 Bivariate extreme value theory

The study of bivariate extremes may be split into two components: the marginal distributions and the dependence structure. We first model marginal distributions using univariate techniques and then, after a transformation standardizing the margins to a common scale, we deal with the dependence structure. Here we restrict our attention to the modeling of extremal dependence, but we emphasize that marginal distribution modeling is an essential step to be able to characterize extremal dependence.

#### Asymptotic characterization of componentwise maxima

Let  $\{(Y_{i,1}, Y_{i,2})\}_{i=1}^N$  be a collection of independent and identically distributed vectors with distribution function  $F(y_1, y_2)$ . Without loss of generality, assume that  $F(y_1, y_2)$  has unit Fréchet marginal distributions,  $F_1(y) = F_2(y) = \exp(-1/y)$ ,  $y > 0$ . If

$$M_{1,N} = \max_{1 \leq i \leq N} \{Y_{i,1}\} \quad \text{and} \quad M_{2,N} = \max_{1 \leq i \leq N} \{Y_{i,2}\}, \quad (1.9)$$

we say that  $\mathbf{M}_N = (M_{1,N}, M_{2,N})$  is the vector of componentwise maxima. The asymptotic theory of bivariate extremes begins by noticing that, for  $y > 0$ ,  $\mathbb{P}(M_{1,N}/N \leq y) = \mathbb{P}(M_{2,N}/N \leq y) = \exp(-1/y)$  and therefore, we should consider the re-scaled vector  $(M_{1,N}^*, M_{2,N}^*) = N^{-1} (M_{1,N}, M_{2,N})$ . The following result provides a bivariate analog to [Theorem 1](#), characterizing the distribution of  $(M_{1,N}^*, M_{2,N}^*)$  as  $N$  goes to infinity.

**Theorem 4.** Let  $\{(Y_{i,1}, Y_{i,2})\}_{i=1}^N$  be a sequence of independent identically distributed random vectors

whose distribution function has unit Fréchet marginal distributions. Let  $(M_{1,N}^*, M_{2,N}^*) = N^{-1}(M_{1,N}, M_{2,N})$ , with  $M_{1,N}$  and  $M_{2,N}$  defined as in (1.9). Then if

$$\mathbb{P}(M_{1,N}^* \leq y_1, M_{2,N}^* \leq y_2) \rightarrow G(y_1, y_2), \quad N \rightarrow \infty, \quad (1.10)$$

where  $G$  is a non-degenerate distribution function,  $G$  has the form

$$G(y_1, y_2) = \exp \{-V(y_1, y_2)\},$$

where

$$V(y_1, y_2) = 2 \int_{[0,1]} \max\left(\frac{w}{y_1}, \frac{1-w}{y_2}\right) H(dw), \quad (1.11)$$

and  $H$  is a measure satisfying the following moment constraints

$$\int_{[0,1]} H(dw) = 1, \quad \int_{[0,1]} wH(dw) = \frac{1}{2}. \quad (1.12)$$

The family of distributions that arises as limit in Equation (1.10) is called the class of bivariate extreme value distributions. The measure  $H$  is called spectral measure, and is a probability distribution on  $[0,1]$ . The function  $V$  is called exponent measure. Specification of either  $V$  or  $H$  would characterizes the dependence structure of the limiting distribution  $G$ . As in the univariate case, if  $F$  has  $G$  as limiting distribution, we say that  $F$  is in the *domain of attraction* of  $G$ . If  $H$  is absolutely continuous with density  $h$ , the expression in (1.11) is simply

$$V(y_1, y_2) = 2 \int_0^1 \max\left(\frac{w}{y_1}, \frac{1-w}{y_2}\right) h(w)dw.$$

In this dissertation, we assume that  $H$  is absolutely continuous. However, bivariate extreme value distributions can also be generated by measures  $H$  that are not absolutely continuous. Two limiting cases are discussed in the following examples.

**Example 1** (Independent variables). if  $H$  is a measure that puts mass 0.5 on  $w = 0$  and  $w = 1$ , then

$V(y_1, y_2) = y_1^{-1} + y_2^{-1}$  and

$$G(y_1, y_2) = \exp\{-(y_1^{-1} + y_2^{-1})\} = \exp(-y_1^{-1}) \exp(-y_2^{-1}), \quad y_1, y_2 > 0. \quad (1.13)$$

This situation is referred to as asymptotic independence, a boundary case that will be studied further in this chapter.

**Example 2** (Complete dependent variables). If  $H$  is a measure that puts mass 1 on  $w = 0.5$ , then  $V(y_1, y_2) = \max(y_1^{-1}, y_2^{-1})$  and

$$G(y_1, y_2) = \exp\{-\max(y_1^{-1}, y_2^{-1})\}, \quad y_1, y_2 > 0. \quad (1.14)$$

In this case,  $G$  is the distribution function of variables that are completely dependent:  $Y_1 = Y_2$  with probability 1.

Although Theorem 4 is stated for random vectors with unit Fréchet margins, the result applies to any other choice of marginal distributions. This particular choice is not so important from an asymptotic point of view, but some useful properties can be more easily studied from a specific selection. Popular choices include exponential, Gumbel, uniform, and Weibull distributions. Examples of other margins employed in the literature can be found in [Beirlant \*et al.\* \(2004, p. 271\)](#).

The importance of the spectral measure  $H$  is that it can be used to describe the extremal dependence, and this can be understood through a pseudo-polar transformation, where we map  $\{Y_{i,1}, Y_{i,2}\}_{i=1}^N$  to a pseudo-angular, pseudo-radial scale

$$(W_i, R_i) = \left( \frac{Y_{i,1}}{Y_{i,1} + Y_{i,2}}, Y_{i,1} + Y_{i,2} \right), \quad i = 1, \dots, N.$$

If  $W$  and  $R$  are general terms of  $\{W_i\}_{i=1}^N$  and  $\{R_i\}_{i=1}^N$  respectively, [de Haan and Resnick \(1977\)](#) showed that the spectral measure is the asymptotic distribution of the ratios  $W$  given that the sum  $R$  is large, i.e.,

$$P(W \in \cdot | R > t) \rightarrow H(\cdot), \quad t \rightarrow \infty. \quad (1.15)$$

Further details on the effect of the transformation from Cartesian to pseudo-polar coordinates, will be

studied later in this chapter.

It can be shown that the convergence in (1.15) is equivalent to the weak convergence of  $(M_{1,N}^*, M_{2,N}^*)$  to  $G$ , and therefore, it is not unreasonable to assert that  $H$  provides relevant information on extremal dependence.

### Bivariate threshold excess model

As in the univariate case, there is an alternative representation of bivariate extremes that avoids the wastefulness of data implied by the study of componentwise block maxima. Again, let  $\{(Y_{i,1}, Y_{i,2})\}_{i=1}^N$  be a collection of independent and identically distributed vectors with distribution function  $F(y_1, y_2)$ . The goal is to find a family of distributions which approximates  $F$  on regions of the form  $y_1 > u_1$ ,  $y_2 > u_2$ , for high enough thresholds  $u_1$  and  $u_2$ . For these thresholds, the marginal distributions of  $F$  have an approximation of the form (1.5), with parameters  $(\zeta_1, \sigma_1, \xi_1)$  and  $(\zeta_2, \sigma_2, \xi_2)$  respectively. Let

$$\tilde{Y}_j = - \left[ \log \left\{ 1 - \zeta_j \left( 1 + \frac{\xi_j (Y_j - u_j)}{\sigma_j} \right)^{-1/\xi_j} \right\} \right]^{-1}, \quad j = 1, 2. \quad (1.16)$$

Then,  $(\tilde{Y}_1, \tilde{Y}_2)$  has distribution function  $\tilde{F}$  whose margins are approximately standard Fréchet for  $Y_j > u_j$ ,  $j = 1, 2$ . Note that if  $\tilde{y}_1$  and  $\tilde{y}_2$  are defined in terms of  $y_1$  and  $y_2$  as in (1.16), then  $\tilde{F}(\tilde{y}_1, \tilde{y}_2) = F(y_1, y_2)$  and, for large enough  $N$ ,

$$\tilde{F}(\tilde{y}_1, \tilde{y}_2) = \{\tilde{F}^N(\tilde{y}_1, \tilde{y}_2)\}^{1/N} \approx [\exp\{-V(\tilde{y}_1/N, \tilde{y}_2/N)\}]^{1/N} = \exp\{-V(\tilde{y}_1, \tilde{y}_2)\}.$$

It follows that

$$F(y_1, y_2) \approx G(y_1, y_2) = \exp\{-V(\tilde{y}_1, \tilde{y}_2)\}, \quad y_1 > u_1, \quad y_2 > u_2. \quad (1.17)$$

In other words, for large enough thresholds  $u_1$  and  $u_2$ , we find that  $G(y_1, y_2)$  is a suitable family which approximates an arbitrary joint distribution  $F(y_1, y_2)$  on regions of the form  $y_1 > u_1$ ,  $y_2 > u_2$ .

If a bivariate pair exceeds a specified threshold in just one of its components, then the model in (1.17) is not appropriate. In this case, it is necessary to adopt methods like censored-likelihood (Ledford and Tawn, 1996; Smith, 1994; Smith *et al.*, 1997). General aspects of the difficulties in this setting can be

found in Coles (2001, p. 155). More details on bivariate threshold exceedances can be found in Beirlant *et al.* (2004, Sections 8.3 and 9.4).

### Point process representation

As with the univariate approach, there is an alternative point process representation for bivariate extremes, which can also be generalized for higher dimensions. We summarized the point process characterization in the following theorem, which gives an interpretation of the spectral measure  $H$ .

**Theorem 5.** Let  $\{(y_{i,1}, y_{i,2})\}_{i=1}^N$  be a sequence of independent bivariate observations from a distribution with unit Fréchet margins that satisfies the convergence result in theorem (4). Let  $\{P_N\}$  be a sequence of point processes defined by

$$P_n = \{(N^{-1}Y_{1,1}, N^{-1}Y_{1,2}), \dots, (N^{-1}Y_{N,1}, N^{-1}Y_{N,2})\}.$$

Then,  $P_n \xrightarrow{d} P$  on regions bounded from the origin  $(0,0)$ , where  $P$  is a non-homogeneous Poisson process on  $(0, \infty) \times (0, \infty)$ . Moreover, letting

$$r = y_1 + y_2 \quad \text{and} \quad w = \frac{y_1}{y_1 + y_2},$$

the intensity function of  $P$  is

$$\lambda(r, w) = 2 \frac{dH(w)}{r^2}. \tag{1.18}$$

To interpret the later result, note that  $r$  is a measure of distance from the origin and  $w$  measures angle on a  $(0, 1]$  scale. Equation (1.18) implies that the intensity of the limiting process  $P$  factorizes across radial and angular components. In other words, the angular spread of points of  $P$  is determined by  $H$ , and is independent of radial distance. If  $H$  has density  $h$ , then  $h$  determines the relative frequency of event of different relative size. If extremes are near-independent, then  $h(w)$  is large close to  $w = 0$  and  $w = 1$ , and small elsewhere. In contrast, if dependence is very strong,  $h(w)$  is large close to  $w = 1/2$ . The general result in (1.18) allows to obtain the two limiting cases studied in Examples 1 and 2.

### Pickands dependence function and the extremal coefficient

Apart from the spectral measure  $H$ , various alternative expressions to describe the dependence structure of  $G$  have been proposed in the literature. One of such expressions is given by the Pickands dependence function (Pickands, 1981), defined as

$$A(w) = 1 - w + 2 \int_0^w H(u) du, \quad w \in [0, 1]. \quad (1.19)$$

It can be shown that  $G$  is completely determined by its margins, say  $G_1$  and  $G_2$ , and its Pickands dependence function,  $A$ , through

$$G(y_1, y_2) = \exp \left[ \log \{G_1(y_1)G_2(y_2)\} A \left\{ \frac{\log(G_2(y_2))}{\log(G_1(y_1)G_2(y_2))} \right\} \right], \quad y_1, y_2 > 0.$$

If  $G_1$  and  $G_2$  are unit Fréchet distributions, the above expression reduces to

$$G(y_1, y_2) = \exp \left\{ - (y_1^{-1} + y_2^{-1}) A \left( \frac{y_1}{y_1 + y_2} \right) \right\}, \quad y_1, y_2 > 0.$$

By the definition in (1.19),  $A(0) = A(1) = 1$ . In addition, we have that  $\max\{1 - w, w\} \leq A(w) \leq 1$  and, most importantly,  $A(\cdot)$  is convex within this region. The lower bound  $A(w) = \max\{1 - w, w\}$  corresponds to complete dependent variables, with a measure  $H$  that puts mass 1 on  $w = 0.5$  (see Example 2). The upper bound  $A(w) = 1$  corresponds to independent variables, with a measure that puts mass 0.5 on  $w = 0$  and  $w = 1$  (see Example 1). The class of dependence functions is a convex set, which means that if  $A_1, \dots, A_p$  are dependence functions, then

$$A(w) = \sum_{i=1}^p \pi_i A_i(w)$$

is a dependence function for  $\pi_i \geq 0$  ( $i = 1, \dots, p$ ) and  $\sum_{i=1}^p \pi_i = 1$ . Two interesting properties of the variables  $Y_1$  and  $Y_2$  can be formulated in terms of  $A(w)$ :

1. It can be shown that the variables are exchangeable if and only if  $A(w)$  is symmetric about 0.5,
2. the correlation between  $(Y_1, Y_2)$  is  $\rho = \int_0^1 \frac{dw}{\{A(w)\}^2} - 1$ , and is always nonnegative because  $A(w) \leq 1$ .

Although the spectral measure and the Pickands dependence function are able to describe the dependence structure of a bivariate extreme value distribution, their infinite-dimensional nature makes them difficult to handle. Moreover, from an exploratory point of view, it is useful to account for summary measures that can easily capture the main properties of the extremal behavior, to help us understand the overall dependence strength. One such measure that has gained popularity is the extremal coefficient

$$C = 2A(1/2), \tag{1.20}$$

which is bounded in  $1 \leq C \leq 2$ . Independence occurs only with  $C = 2$ , whereas complete dependence occurs only with  $C = 1$ .

The problem of estimating a bivariate extreme value distribution can be formulated as how to estimate the Pickands dependence function. A wide range of nonparametric estimators have been proposed to estimate this function and, as a consequence, the extremal coefficient. Pickands itself proposed an estimator (Pickands, 1981) that was later improved by Deheuvels (1991). A more recent proposal can be found in Capéraà *et al.* (1997) and Gudendorf and Segers (2011, 2012). Examples of Pickands dependence functions and extremal coefficients for some parametric models will be given later in this chapter.

### Asymptotic independence and asymptotic complete dependence

There are two limiting cases in the class of dependence structures of bivariate extreme value distributions, which arise either when the extremes of marginal variables are completely dependent or rather independent. Sibuya (1960), defined the latter limiting case as follows: suppose that  $(Y_1, Y_2)$  is a bivariate random variable with identically distributed marginals. We say that the pair  $(Y_1, Y_2)$  is *asymptotically independent* if

$$\lim_{\tau \rightarrow \tau^+} \mathbb{P}(Y_1 > \tau | Y_2 > \tau) = 0, \tag{1.21}$$

where  $\tau^+ = \sup\{\tau \in \mathbb{R} : \mathbb{P}(Y_j \leq \tau) < 1, j = 1, 2\}$ . If the limit in (1.21) is a nonzero constant, the corresponding limiting componentwise maxima are then dependent, and we say that  $Y_1$  and  $Y_2$  are *asymptotically dependent*. Examples 1 and 2 are clear reflects of these two cases.

Asymptotic dependence is interesting from a theoretical point of view. Within this class, the assump-

tions made so far in this section are reasonable, and Theorem 4 characterizes in a reasonable way, the probabilities of joint extreme events. Asymptotic independence is particularly important in applications, mostly because many bivariate distributions lie in the domain of attraction of a bivariate extreme value distribution with independent margins (Sibuya, 1960). This is the case, for example, for a random vector  $\{(Y_{i,1}, Y_{i,2})\}_{i=1}^N$  with bivariate normal distribution, and correlation coefficient  $\rho < 1$ . Under asymptotic independence, fitting a model based on Theorem 4 will likely overestimate dependence on extrapolation, since there is mis-placed assumption that the most extreme marginal events may occur simultaneously. This makes the class of bivariate extreme value distributions rather inappropriate for modeling data that exhibit association that gradually disappears at more and more extreme levels. Statistical models for the general class of distributions for which (1.21) is true, have been developed comparatively recently. Important contributions include in Ledford and Tawn (1996), Bortot and Tawn (1998), and Bruun and Tawn (1998).

After all these examinations, we would like to easily discern whether a bivariate distribution belongs to the class of asymptotically independent distributions. To do so, we turn our attention to a number of coefficients that are able to produce the following:

- A measure useful in distinguishing asymptotic dependence from asymptotic independence.
- A measure of dependence strength within the class of asymptotically dependent distributions.
- A measure of extremal dependence within the class of asymptotically independent distributions.

- **Coefficients of extremal dependence: asymptotic dependence**

The limit in (1.21) can be generalized to the case where the marginal distribution functions,  $F_1$  and  $F_2$ , are non-identical. Let

$$\chi = \lim_{u \rightarrow 1} \mathbf{P}(F_1(Y_1) > u | F_2(Y_2) > u). \quad (1.22)$$

Loosely stated, (1.22) is the probability of one variable being extreme given that the other is extreme. We already know that the variables are asymptotically independent when  $\chi = 0$ , and asymptotically dependent when  $0 < \chi \leq 1$ . Furthermore, it is possible to show that if  $F_1$  and  $F_2$  are in the domain of attraction of univariate extreme value distributions  $G_1$  and  $G_2$  respectively, then  $\chi = 0$  if and only if  $G$  is in the domain of attraction of the bivariate extreme value distribution

$G(y_1, y_2) = G_1(y_1)G_2(y_2)$ . Within the class of asymptotically dependent variables, the value of  $\chi$  increases with increasing degree of dependence at extreme levels.

Although  $\chi$  is a natural summary of extremal dependence, it is more convenient to obtain  $\chi$  as the limit of an alternative (but asymptotically equivalent) function. If

$$\chi(u) = 2 - \frac{\log \mathbb{P}(F_1(Y_1) < u, F_2(Y_2) < u)}{\log(u)}, \quad 0 < u < 1, \quad (1.23)$$

then, it can be shown that for  $u \rightarrow 1$

$$\chi(u) = 2 - \frac{1 - \mathbb{P}(F_1(Y_1) < u, F_2(Y_2) < u)}{1 - u} + o(1) = \mathbb{P}(F_2(Y_2) > u | F_1(Y_1) > u) + o(1),$$

whence  $\chi = \lim_{u \rightarrow 1} \chi(u)$ . The function  $\chi(u)$  is bounded by

$$2 - \frac{\log(\max\{2u - 1, 0\})}{\log u} \leq \chi(u) \leq 1,$$

and is convenient not only because it provides the limit  $\chi$ , but also because it can be interpreted as a quantile-dependent measure of dependence. In particular, the sign of  $\chi(u)$  determines whether the variables are positively or negatively associated at the quantile level  $u$ . For any distribution falling in the class of bivariate extreme value distributions,  $\chi(u) = 2 - C$ , where  $C$  is the extremal coefficient defined in (1.20). Therefore,  $\chi(u)$  is constant in  $u$  and its estimate provides a diagnostic check for membership of the bivariate extreme value class. The simplicity of the formulation of  $\chi(u)$  is a property of the bivariate extreme value distributions and does not always apply to other distributions. Examples of dependence models where  $\chi(u)$  is non-trivial can be found in [Coles et al. \(1999\)](#).

- **Coefficients of extremal dependence: asymptotic independence**

We already highlighted the importance of the class of asymptotically independent distributions in bivariate extreme modeling. Within this class, it is still possible to distinguish different degrees of dependence at finite levels. By definition,  $\chi = 0$ , so  $\chi$  is unable to provide information on the relative strength of dependence for such models. To overcome this situation, we define a second

dependent measure,  $\bar{\chi}$ , based on a comparison of joint and marginal survivor functions of  $F_1(Y_1)$  and  $F_2(Y_2)$ . Let

$$\bar{\chi}(u) = \frac{2 \log \mathbf{P}(F_1(Y_1) > u)}{\log \mathbf{P}(F_1(Y_1) > u, F_2(Y_2) > u)} - 1 = \frac{2 \log(1 - u)}{\log \mathbf{P}(F_1(Y_1) > u, F_2(Y_2) > u)} - 1, \quad 0 < u < 1. \quad (1.24)$$

Then  $-1 < \bar{\chi}(u) \leq 1$  for all  $0 < u < 1$ . Analogous to  $\chi$ , and with the purpose of describing extremal characteristics, we define

$$\bar{\chi} = \lim_{u \rightarrow 1} \bar{\chi}(u).$$

For asymptotically dependent variables  $\bar{\chi} = 1$ , and for asymptotically independent variables,  $-1 < \bar{\chi} < 1$ , and  $\bar{\chi}$  provides a limiting measure that increases with dependence strength.

As a result, the pair  $(\bar{\chi}, \chi)$  provides informative and complementary information about the form of extremal dependence. It can be used as a summary of extremal dependence in the following sense:

- If  $\bar{\chi} = 1$  and  $0 < \chi \leq 1$ , the variables are asymptotically dependent and  $\chi$  is a measure of strength of dependence within the class of asymptotically dependent distributions.
- If  $-1 < \bar{\chi} < 1$  and  $\chi = 0$ , the variables are asymptotically independent and  $\bar{\chi}$  is a measure of strength of dependence within the class of asymptotically independent distributions.

- **The coefficient of tail dependence**

This coefficient was introduced by [Ledford and Tawn \(1996\)](#) and is useful in distinguishing asymptotic dependence from asymptotic independence and, within the class of asymptotically independent distributions, positive from negative association. Let  $Z_j = -1/\log F_j(Y_j)$ , for  $j = 1, 2$ . It is easy to check that  $Z_1$  and  $Z_2$  have unit Fréchet distributions. To introduce this coefficient, [Ledford and Tawn \(1996\)](#) assume that the joint survivor function of  $Z_1$  and  $Z_2$  is a regularly varying function, that is:

$$\mathbf{P}(Z_1 > z, Z_2 > z) = \mathcal{L}(z)z^{-1/\eta}, \quad z > 0,$$

and  $\mathcal{L}$  is a *slowly varying function*, which means that  $\lim_{z \rightarrow \infty} \mathcal{L(xz)/\mathcal{L}(z) = 1$  for all fixed  $x > 0$ . The number  $\eta$  is a constant, called the *coefficient of tail dependence*, and is bounded in  $0 < \eta \leq 1$ . It can be shown that

$$\lim_{z \rightarrow \infty} \mathcal{L}(z)z^{1-1/\eta} = \chi, \quad \lim_{u \rightarrow 1} \bar{\chi}(u) = \bar{\chi} = 2\eta - 1,$$

provided the limits exist. As a consequence:

- If  $\eta = 1$  and  $\lim_{z \rightarrow \infty} \mathcal{L}(z) = c$  for some  $0 < c \leq 1$ , then  $\bar{\chi} = 1$  and the variables are asymptotically dependent of degree  $\chi = c$ .
- If  $0 < \eta < 1$  or if  $\eta = 1$  and  $\lim_{z \rightarrow \infty} \mathcal{L}(z) = 0$ , then  $\chi = 0$  and the variables are asymptotically independent of degree  $\bar{\chi} = 2\eta - 1$ . Within this class, three types of independence can be identified according to the sign of  $\bar{\chi}$ :
  - \* If  $1/2 < \eta < 1$  or  $\eta = 1$  and  $\lim_{z \rightarrow \infty} \mathcal{L}(z) = 0$ , there is positive association, which means that observations for which both  $Z_1$  and  $Z_2$  exceed a large threshold  $z$  occur more frequently than under exact independence.
  - \* If  $\eta = 1/2$ , extremes of  $Z_1$  and  $Z_2$  are near independent and even exactly independent in the case  $\mathcal{L}(z) = 1$ .
  - \* If  $0 < \eta < 1/2$ , there is negative association, which means that observations for which both  $Z_1$  and  $Z_2$  exceed a large threshold  $z$  occur less frequently than under exact independence.

To summarize, the degree of dependence between large values of  $Z_1$  and  $Z_2$  is determined by  $\eta$ , with increasing values of  $\eta$  corresponding to stronger association. For a given  $\eta$ , the relative strength of dependence is characterized by  $\mathcal{L}$ .

### Modeling the spectral measure

As we can observe from Equation (1.11), any measure  $H$  on  $[0, 1]$  satisfying the mean constraint (1.12) gives rise to a valid limit in (1.10). In other words, the limit family has no finite parametrization, and the use of nonparametric methods seems a natural approach. Nonparametric techniques to estimate the

spectral measure are a keystone of this dissertation, and the current models available in the literature are presented at the end of this section. Nevertheless, to understand the historical development of bivariate extreme value modeling, we believe it is necessary to begin with an outline of some parametric alternatives proposed in the literature.

- **Parametric models**

The parametric approach is based in obtaining a small subset of the class of limit distributions for  $G$ , which can flexibly adapt itself to the data at hand. This is carried out using parametric sub-families of distributions for  $H$  that leads to sub-families of distributions for  $G$ . Building suitable models is not an easy task, since we need parametric families whose mean is parameter-free, and for which the exponent measure is tractable. In the following examples, we describe two parametric models that have been widely studied in the literature. We characterize their dependence structure, as well as summary measures for extremal dependence studied before in this chapter. We will refer to these two models in Chapters 2 and 3 when we conduct simulation studies. Other parametric models as well as model construction methods can be found in [Beirlant \*et al.\* \(2004, p. 300\)](#).

**Example 3** (Logistic family). The logistic model ([Coles, 2001, p. 146](#)) is a popular model, based on the spectral density

$$h(w) = \frac{1}{2} \left( \frac{1}{\alpha} - 1 \right) \{w(1-w)\}^{-1-1/\alpha} \{w^{-1/\alpha} + (1-w)^{-1/\alpha}\}^{\alpha-2}, \quad w \in (0, 1), \quad (1.25)$$

where  $\alpha \in (0, 1]$  is the dependence parameter. The closer  $\alpha$  is to 0, the higher the level of extremal dependence, while the closer  $\alpha$  is to 1, the closer we get to independence. The mean constraint in (1.12) is satisfied since the model is symmetric about  $w = 0.5$ . The bivariate extreme value distribution induced by (3.6) is

$$G(y_1, y_2) = \exp \left\{ - \left( y_1^{-1/\alpha} + y_2^{-1/\alpha} \right)^\alpha \right\}, \quad y_1, y_2 > 0,$$

and as we can see, the model is exchangeable. The main reason for the popularity of the logistic family is its tractability; it covers all levels of dependence from independence to perfect

dependence, and this can be understood by noticing that

$$\lim_{\alpha \rightarrow 1} G(y_1, y_2) = \exp\{-(y_1^{-1} + y_2^{-1})\}, \quad \lim_{\alpha \rightarrow 0} G(y_1, y_2) = \exp\{-\max(y_1^{-1}, y_2^{-1})\}.$$

This can be better illustrated through the Pickands dependence function

$$A(w) = \{(1 - w)^{1/\alpha} + w^{1/\alpha}\}^\alpha, \quad w \in (0, 1),$$

which is displayed in Figure 1-1 (a) for three values of the dependence parameter. The closer  $A(w)$  is to its lower bound  $\max\{1 - w, w\}$ , the closer the model is to complete dependence ( $\alpha$  close to 0), while the closer  $A(w)$  is to its upper bound 1, the closer the model is to independence ( $\alpha$  close to 1). The extremal coefficient is  $C = 2^\alpha$  and the bivariate survivor function corresponding to the unit Fréchet margins  $Z_j = -1/\log F_j(Y_j)$ ,  $j = 1, 2$ , satisfies

$$\mathbb{P}(Z_1 > z, Z_2 > z) = (2 - 2^\alpha)z^{-1} + (2^{2\alpha-1} - 1)z^{-2} + o(z^{-2}), \quad z \rightarrow \infty.$$

If  $0 < \alpha < 1$ , we find a coefficient of tail dependence  $\eta = 1$  and a slowly varying function  $\mathcal{L}$  converging to  $\chi = 2 - 2^\alpha$ .

**Example 4** (Dirichlet model). The Dirichlet model proposed by [Coles and Tawn \(1991\)](#) is an asymmetric model, based on the spectral density

$$h(w) = \frac{ab\Gamma(a + b + 1)(aw)^{a-1}\{b(1 - w)\}^{b-1}}{2\Gamma(a)\Gamma(b)\{aw + b(1 - w)\}^{a+b+1}}, \quad w \in (0, 1), \quad (1.26)$$

where  $a, b > 0$ . The Dirichlet model is symmetric only in the case  $a = b$ . The corresponding bivariate extreme value distribution is given by

$$G(y_1, y_2) = \exp \left[ -y_1^{-1} \{1 - B(q; a + 1, b)\} - y_2^{-1} B(q; a, b + 1) \right], \quad y_1, y_2 > 0,$$

where  $q = ay_1/(ay_1 + by_2)$  and  $B(q; a, b)$  is the regularized incomplete beta function. If  $(a, b) =$

(1, 1), the Pickands dependence function can be easily expressed as

$$A(w) = 1 - w(1 - w), \quad w \in (0, 1),$$

with an extremal coefficient  $C = 3/2$ . The Pickands dependence function inherits the asymmetry of the model, and is displayed in Figure 1-1 (b) for three different pairs of parameters.

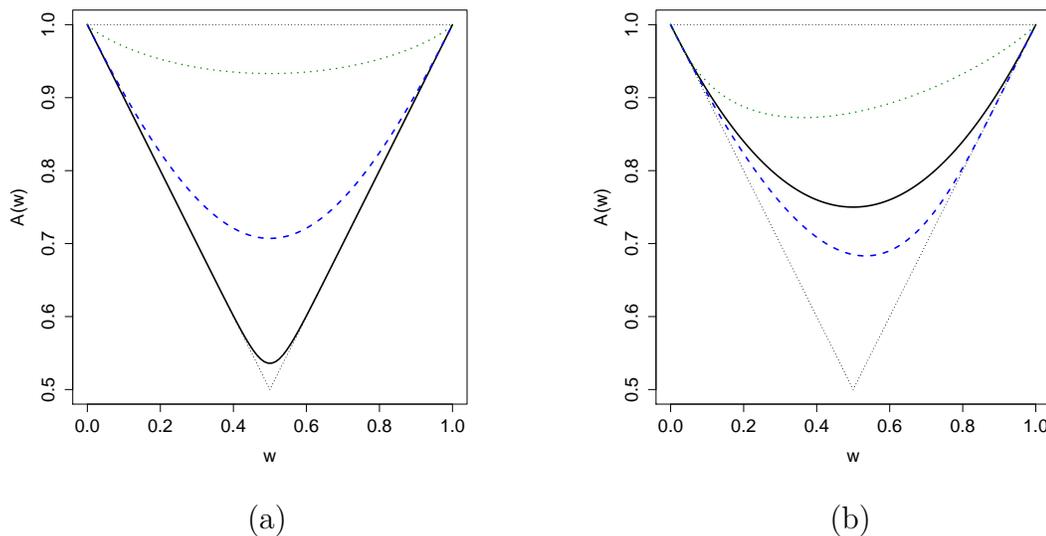


Figure 1-1. Pickands dependence functions: (a) logistic model with  $\alpha = 0.1$  (black solid line),  $\alpha = 0.5$  (blue dashed line),  $\alpha = 0.9$  (green dotted line); (b) Dirichlet model with  $(a, b) = (1, 1)$  (black solid line),  $(a, b) = (100, 1)$  (blue dashed line) and  $(a, b) = (0.1, 10)$  (green dotted line).

Although some parametric models may offer a good balance between model flexibility and analytical tractability, they are far from covering the general class of bivariate extreme value distributions. Moreover, the simplicity of the models can lead to model mis-specification. As mentioned before, the nonparametric nature of  $H$  makes nonparametric methods a perhaps more natural and reasonable approach. Nevertheless, it is not straightforward to constrain nonparametric estimators to satisfy functional constraints of the form (1.12). In the next section, we study three nonparametric estimators of the spectral measure that have been proposed in the literature. More details on nonparametric estimation of the spectral measure and on estimation of extremal dependence, can be found in Kiriliouk et al. (2015).

• **Nonparametric estimation**

As mentioned before, to understand the tail dependence structure of a random vector, we model the joint tail of a bivariate distribution by modeling the marginal distributions and the dependence structure separately. To estimate dependence at high levels, the spectral measure as defined in (1.15) is particularly useful. In practice, we do not observe the pseudo-angles  $W_i$ , but we can construct proxies by setting

$$\widehat{R}_i = \widehat{Y}_{i,1}^* + \widehat{Y}_{i,2}^*, \quad \widehat{W}_i = \frac{\widehat{Y}_{i,1}^*}{\widehat{Y}_{i,1}^* + \widehat{Y}_{i,2}^*},$$

where  $\widehat{Y}_{i,1}^* = -1/\log \widehat{F}_1(Y_{i,1})$  and  $\widehat{Y}_{i,2}^* = -1/\log \widehat{F}_2(Y_{i,2})$  and  $\widehat{F}_1 = \widehat{F}_{1,N}$ ,  $\widehat{F}_2 = \widehat{F}_{2,N}$  are estimators of the marginal distribution functions  $F_1$  and  $F_2$  respectively. A robust choice for  $\widehat{F}_X$  and  $\widehat{F}_Y$  is the pair of univariate empirical distribution functions, normalized by  $N + 1$  rather than by  $N$  to avoid division by zero. In this case,  $\widehat{Y}_{i,1}^*$  and  $\widehat{Y}_{i,2}^*$  are functions of the ranks. For a high enough threshold  $t_N$ , the collection of angles  $\{\widehat{W}_i : i \in I_N\}$  with  $I_N = \{i = 1, \dots, N : \widehat{R}_i > t_N\}$  can be regarded as an approximate sample from the spectral measure  $H$ . Parametric or nonparametric inference on  $H$  may then be based upon the sample  $\{\widehat{W}_i : i \in I_N\}$ .

The spectral measure lies at the basis of nonparametric techniques and in this section, we study three nonparametric estimators proposed in the literature. All three are of the form

$$\tilde{H}_l(w) = \sum_{i \in I_N} \tilde{p}_{l,i} \mathbb{1}\{\widehat{W}_i \leq w\}, \quad w \in [0, 1], \quad l \in \{1, 2, 3\},$$

and the estimators distinguish themselves in the way the weights  $\tilde{p}_{l,i}$  are defined.

For the *empirical spectral measure* (Einhmahl *et al.*, 2001), the weights are  $\tilde{p}_{1,i} = 1/n$ , where  $n = |I_N|$  denotes the cardinality of the set  $I_N$ , i.e., the number of extreme observations. The estimator  $\tilde{H}_1$  becomes an empirical version of (1.15), i.e.,

$$\tilde{H}_1(w) = \frac{1}{n} \sum_{i \in I_N} \mathbb{1}\{\widehat{W}_i \leq w\}, \quad w \in [0, 1].$$

This estimator does not necessarily satisfy the moment constraints in (1.12) and this is the

motivation for the two other estimators, where the moment constraint is enforced by requiring that  $\sum_{i \in I_N} \widehat{W}_i \tilde{p}_{l,i} = 1/2$ , for  $l = 2, 3$ .

For the *maximum empirical likelihood estimator* (Einhmahl and Segers, 2009),  $\tilde{H}_2$  has probability masses  $\tilde{p}_{2,i}$  solving the optimization problem

$$\begin{aligned} & \max_{p \in \mathbb{R}_+^n} \sum_{i \in I_N} \log p_{2,i} \\ & \text{s.t. } \sum_{i \in I_N} p_{2,i} = 1, \quad \sum_{i \in I_N} \widehat{W}_i p_{2,i} = 1/2. \end{aligned} \tag{1.27}$$

We can see that, by construction, the weights  $\tilde{p}_{2,i}$  satisfy the moment constraints and is implicitly assumed that  $\tilde{p}_{2,i} > 0$ . The optimization problem in (1.27) can be solved by the method of Lagrange multipliers. In the nontrivial case where  $1/2$  is in the convex hull of  $\{\widehat{W}_i : i \in I_N\}$ , the solution is given by

$$\tilde{p}_{2,i} = \frac{1}{n} \frac{1}{1 + \lambda \{\widehat{W}_i - 1/2\}}, \quad i \in I_N,$$

where  $\lambda \in \mathbb{R}$  is the Lagrange multiplier associated to the second equality constraint in (1.27), defined implicitly as the solution to the equation

$$\frac{1}{n} \sum_{i \in I_N} \frac{\widehat{W}_i - 1/2}{1 + \lambda(\widehat{W}_i - 1/2)} = 0,$$

(see, for instance, Owen, 2001, Chapter 3). Then, the maximum empirical likelihood estimator is

$$\tilde{H}_2(w) = \sum_{i \in I_N} \frac{1}{n \{1 + \lambda(\widehat{W}_i - 1/2)\}} \mathbb{1}\{\widehat{W}_i \leq w\}, \quad w \in [0, 1].$$

Computation of  $\tilde{H}_2$  is not simple since its expression is not free of Lagrange multipliers. This also makes asymptotic theory less manageable. This complexity motivates the introduction of the next estimator.

The *Euclidean likelihood estimator* (de Carvalho *et al.*, 2013) satisfies the moment constraints in (1.12) and is related with the maximum empirical likelihood estimator. The probability masses

$\tilde{p}_{3,i}$  solve the optimization problem

$$\begin{aligned} \max_{p \in \mathbb{R}_+^n} & -\frac{1}{2} \sum_{i \in I_N} \log(np_{3,i} - 1)^2 \\ \text{s.t.} & \sum_{i \in I_N} p_{3,i} = 1, \quad \sum_{i \in I_N} \widehat{W}_i p_{3,i} = 1/2. \end{aligned} \quad (1.28)$$

This quadratic optimization problem with linear constraints can be solved explicitly with the method of Lagrange multipliers, yielding

$$\tilde{p}_{3,i} = \frac{1}{n} \{1 - (\overline{W} - 1/2) S_W^{-2} (\widehat{W}_i - \overline{W})\}, \quad i \in I_N, \quad (1.29)$$

where  $\overline{W}$  and  $S_W^2$  denote the sample mean and sample variance of  $\widehat{W}_i$ ,  $i \in I_N$ , respectively, that is,

$$\overline{W} = \frac{1}{n} \sum_{i \in I_N} \widehat{W}_i, \quad S_W^2 = \frac{1}{n} \sum_{i \in I_N} (\widehat{W}_i - \overline{W})^2.$$

The Euclidean likelihood estimator can then be written as

$$\tilde{H}_3(w) = \sum_{i \in I_N} \frac{1}{n} \{1 - (\overline{W} - 1/2) S_W^{-2} (\widehat{W}_i - \overline{W})\} \mathbb{1}\{\widehat{W}_i \leq w\}, \quad w \in [0, 1].$$

We can see that  $\tilde{H}_3$  is especially convenient as the weights  $\tilde{p}_{3,i}$  are given explicitly. The weights  $\tilde{p}_{3,i}$  could be negative, but this does usually not occur as the weights are all nonnegative with probability tending to one.

It is shown in [Einmahl and Segers \(2009\)](#) and in [de Carvalho \*et al.\* \(2013\)](#) that  $\tilde{H}_2$  and  $\tilde{H}_3$  are more efficient than  $\tilde{H}_1$ . Moreover, asymptotically, there is no difference between the maximum empirical likelihood or maximum Euclidean likelihood estimators.

**Example 5.** Recall the bivariate logistic model from [Example 3](#):

$$G(y_1, y_2) = \exp \left\{ - \left( y_1^{-1/\alpha} + y_2^{-1/\alpha} \right)^\alpha \right\}, \quad y_1, y_2 > 0.$$

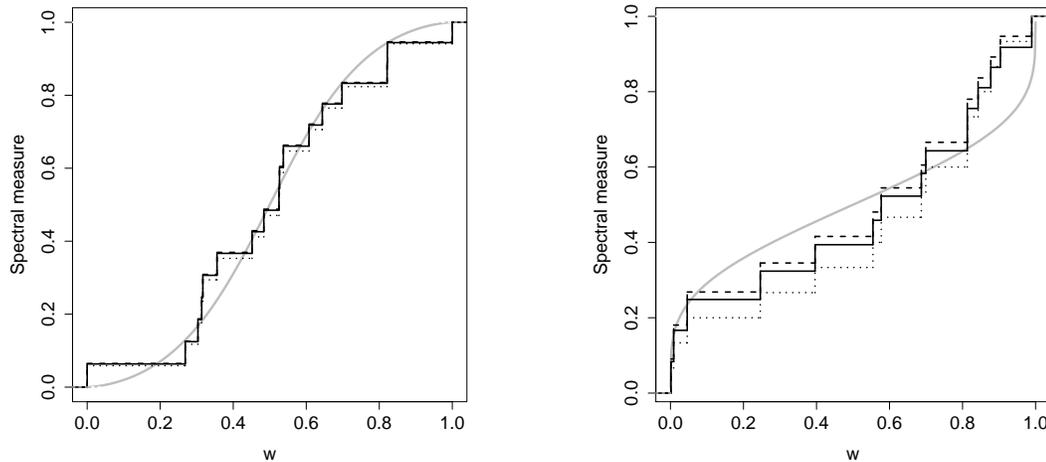


Figure 1-2. Trajectories of the empirical measure (dotted line), the maximum empirical likelihood estimator (dashed line), and the maximum Euclidean likelihood estimator (solid line). The solid grey line corresponds to the true spectral measure of the bivariate logistic model with  $\alpha = 0.4$  (left) and  $\alpha = 0.8$  (right).

For this model, Figure 1-2 shows typical trajectories of the three nonparametric estimators. We can see that the performance of the empirical spectral measure is better for lower values of  $\alpha$  (i.e., when the model is closer to complete dependence), although in both cases the maximum Euclidean and empirical likelihood estimators perform better. We can also see the closeness of the maximum Euclidean and empirical likelihood estimators.

By construction, the three estimators of the spectral measure are discrete. A smooth version which still obeys the moment constraint in (1.12) can be obtained by smoothing the maximum Euclidean or empirical likelihood estimator using kernel smoothing techniques (Chen, 1997; Hall and Presnell, 1988), although some care is needed since the spectral measure is defined on a compact interval. For the Euclidean empirical likelihood estimator, de Carvalho *et al.* (2013) propose a smooth estimator by combining beta distributions with the weights (1.29). To ensure that the estimated measure obeys the marginal moment constraint, it is imposed that the mean of each smoother equals the observed pseudo-angles. For a sample of  $n$  pseudo-angles, the *smooth*

*Euclidean spectral density* is defined as

$$\tilde{h}(w) = \sum_{i=1}^n \tilde{p}_{3,i} \beta\{w; W_i \nu, (1 - W_i) \nu\}, \quad w \in (0, 1),$$

where  $\nu > 0$  is a concentration parameter, asymptotically inversely proportional to the variance of the kernel beta, and with the main role of controlling the amount of smoothing. The corresponding smoothed spectral measure is defined as

$$\tilde{H}(w) = \int_0^w \tilde{h}(v) dv = \sum_{i=1}^n \tilde{p}_{3,i} B\{w; W_i \nu, (1 - W_i) \nu\}, \quad w \in (0, 1),$$

where  $B(\cdot; a, b)$  is the regularized incomplete beta function, with  $a, b > 0$ . Note that since

$$\int_0^1 w \tilde{h}(w) dw = \sum_{i=1}^n \tilde{p}_{3,i} \left\{ \frac{\nu W_i}{\nu W_i + (1 - W_i) \nu} \right\} = \sum_{i=1}^n \tilde{p}_{3,i} W_i = 1/2,$$

the mean constraint is satisfied.

Plug-in estimators can be immediately constructed using the smooth Euclidean spectral density and measure. For the Pickands dependence function and the bivariate extreme value distribution the estimates are

$$\begin{aligned} \tilde{A}(w) &= 1 - w + 2 \sum_{i=1}^n \tilde{p}_{3,i} \int_0^w B\{u; W_i \nu, (1 - W_i) \nu\} du, \quad w \in (0, 1), \\ \tilde{G}(y_1, y_2) &= \exp \left\{ -2 \sum_{i=1}^n \tilde{p}_{3,i} \int_0^1 \max \left( \frac{u}{y_1}, \frac{1-u}{y_2} \right) \beta\{u; W_i \nu, (1 - W_i) \nu\} du \right\}, \quad y_1, y_2 > 0. \end{aligned}$$

#### 1.1.4 Regression analysis in bivariate extremes

The aim of regression analysis is the description of a variable of primary interest (the response) in terms of a set of covariates, and this is carried out through the construction of suitable mathematical models. The regression model has been widely studied over the years and linear regression analysis is one of the oldest and most used statistical techniques. The general linear model links the dependent variable to the covariates in an approximate linear way, making the response distribution depend on

the covariates through its mean. Many extensions of this model have been proposed over the past years, unified by the GLM models, which provide an accessible framework for this kind of analysis.

When dealing with heavy-tailed distributions, the population moments may not be finite, and the techniques for GLMs cannot be used for statistical analysis. Further, from an extreme value point of view, the main interest is in describing conditional tail characteristics rather than modeling conditional means.

The use of covariate information allow us to combine data sets originating from different sources, resulting in opportunities for better point estimates and improved inference. In the univariate setting, the method of block maxima fits the GEV to a sample of maxima, taking one or more of the GEV parameters as a function of the covariates and the regression coefficients, similarly to the approach in nonstationary sequences. Inference can be carried out also in a similar way to the nonstationary sequences case, and goodness-of-fit assessment only requires a suitable transformation of the sample of maxima to obtain identically distributed residuals to produce several kinds of diagnostic plots. Regression for peaks over threshold models were introduced by [Davison and Smith \(1990\)](#). Here, generalized Pareto distribution-based regression models are fitted to exceedances over a high threshold. Similar to the approach followed with the GEV, the idea is to extend the generalized Pareto distribution to a regression model by taking any of the parameters as a function of the covariates and the regression coefficients.

Although the regression terminology has not been used in the bivariate setting, in the literature there are two articles dealing with extremal dependence structures where covariate can be incorporated. The first article correspond to the *density ratio model* introduced by [de Carvalho and Davison \(2014\)](#). This is a semiparametric model for the case where several multivariate extremal distributions are linked through the action of a covariate on an unspecified baseline distribution. The model links the spectral distributions through a known weight function modulated by so-called tilting parameters. Specifically, let  $H_0, \dots, H_K$  be the measures of interest, and assume that they are linked through an exponential tilt, i.e.,

$$\frac{dH_k(w)}{dH_0(w)} = \exp\{\alpha_k + \beta_k c(w)\}, \quad k = 0, \dots, K. \quad (1.30)$$

Here,  $c(w)$  is a known distortion function and  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)^T$  and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)^T$  are the tilting parameters. Specifically,  $\boldsymbol{\beta}$  measures the dependence with respect to the baseline distribution  $H_0$ , and  $\boldsymbol{\alpha}$  is related to  $\boldsymbol{\beta}$  through the normalization and moments constraints

$$\int_0^1 \exp\{\alpha_k + \beta_k c(w)\} dH_0(w) = 1, \quad \int_0^1 w \exp\{\alpha_k + \beta_k c(w)\} dH_0(w) = 1/2, \quad (1.31)$$

for  $k = 0, \dots, K$ . Inference for this model can be carried out through an empirical likelihood approach, based on the combined sample  $\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} = \{w_{0,1}, \dots, w_{0,n_0}, \dots, w_{K,1}, \dots, w_{K,n_K}\}$ . The log likelihood obtained from the model specification in (1.30) is

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, H_0) = \sum_{i=1}^n \log p_i + \sum_{k=0}^K \sum_{j=1}^{n_k} \{\alpha_k + \beta_k c(\mathbf{w}_{k,j})\}, \quad (1.32)$$

where  $p_i = dH_0(\mathbf{v}_i)$  denotes the sizes in the jump of the baseline spectral distribution function  $H_0$  at the observed  $\mathbf{v}_i$ . Empirical likelihood estimation of the spectral density ratio model involves maximizing (1.32) subject to the empirical versions of constraints (1.31), i.e.,

$$\begin{aligned} p_i &\geq 0, \quad \mathbf{v}_i \in (0, 1), \\ \sum_{i=1}^n p_i &= 1, \quad \sum_{i=1}^n p_i \mathbf{v}_i = 1/2, \\ \sum_{i=1}^n p_i \exp\{\alpha_k + \beta_k c(\mathbf{v}_i)\} &= 1, \quad \sum_{i=1}^n p_i \mathbf{v}_i \exp\{\alpha_k + \beta_k c(\mathbf{v}_i)\} = 1/2. \end{aligned} \quad (1.33)$$

This yields the estimator

$$\widehat{H}_k(w) = \sum_{i=1}^n \frac{I(v_i \leq w) \exp\{\widehat{\alpha}_k + \widehat{\beta}_k c(v_i)\}}{\rho_0 + (v_i - 1/2) + \sum_{k=1}^K [\exp\{\widehat{\alpha}_k + \widehat{\beta}_k c(v_i)\} \{\rho_k + \widehat{\lambda}_k (v_i - 1/2)\}]},$$

where  $n = \sum_{k=1}^K n_k$ ,  $\rho_k = n_k/n$  and  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_K)^T$  denotes the Lagrange multipliers corresponding to the constraints in (1.31).

Since the estimate of  $H_k$  is based on  $n = \sum_{k=0}^K n_k$ , the model of [de Carvalho and Davison \(2014\)](#) allows for borrowing strength across samples, but as posed by [de Carvalho \(2015\)](#), their approach requires a substantial computational investment; in particular, inference entails intensive constrained optimiza-

tion problems so that estimates of  $H_k$  obey the empirical versions of the normalization and moment constraints in (1.33).

A second article that includes covariates in the modeling of extremal dependence structures, correspond to the nonstationary max-stable dependence structures developed by [Huser and Genton \(2014\)](#). The authors propose a sensible nonstationary dependence model for the context of spatial extremes, where covariates can be incorporated. The model combines the nonstationary extremal  $t$  model ([Nikoloulopoulos et al., 2009](#); [Opitz, 2013](#); [Thibaud and Opitz, 2014](#)), which is a model proposed in the literature for modeling spatial extremes within a copula framework, with nonstationary correlation functions, specifically, the kernel convolution approach advocated by [Paciorek \(2006\)](#). To incorporate covariate effects in the dependence structure at extreme levels, the authors propose modeling the covariance matrices using the [Paciorek \(2006\)](#) approach, as functions of vector of covariates associated with each location and vector of parameter measuring the importance of covariates. Since the full likelihood for max-stable processes is intractable, the authors conduct inference using pairwise likelihoods ([Lindsay, 1988](#); [Varin et al., 2011](#)). The nonstationary max-stable models fitted by the authors have proved to provide a better fit with respect to the traditional stationary and isotropic max-stable counterpart, although their attention is focused on componentwise maxima instead of peaks over threshold. This somehow less efficient approach has its foundations in the complications inherent to threshold models (threshold selection, modeling of temporal dependence, etc.) and the additional difficulties derived from the nonstationarity context. We choose not to include further details on the [Huser and Genton \(2014\)](#) model, since the spatial context is beyond the scope of this thesis, but we encourage the reader to explore this interesting approach.

The methodologies developed in this dissertation are restricted to extremes of bivariate data, but theoretically, the main results can be extended to a  $d$ -dimensional random sample, with  $d > 2$ . The feasibility of these extensions is discussed in Chapter 4. To have some insights on multivariate extremes, the following section is devoted to the multivariate version of the domain of attraction problem, which is to find a suitable family of multivariate distributions that approximate the distribution of properly normalized componentwise maxima.

### 1.1.5 Multivariate extreme value theory

Let  $\{\mathbf{Y}_i\}_{i=1}^N$  be independent and identically distributed vectors of continuous random variables on  $\mathbb{R}^d$ , with distribution function  $F$ . Without loss of generality, suppose that  $F$  has unit Fréchet marginal distributions. Pickands (1981) asserts that the limiting distribution of the componentwise standardized maximum  $\mathbf{M}_N = N^{-1} \max\{\mathbf{Y}_1, \dots, \mathbf{Y}_N\}$ , which corresponds to the multivariate standardized version of (1.9), is a  $d$ -dimensional extreme value distribution:

$$G(\mathbf{y}) = \exp\{-V(\mathbf{y})\} = \exp\left\{-d \int_{\Delta_d} \max\left(\frac{w_1}{y_1}, \dots, \frac{w_d}{y_d}\right) dH(\mathbf{w})\right\}, \quad \mathbf{y} = (y_1, \dots, y_d) \in [0, \infty)^d. \quad (1.34)$$

Here  $V$  is the exponent measure, and  $H$  is the spectral distribution, defined on the unit simplex in  $\mathbb{R}^d$ , i.e.,  $\Delta_d = \{\mathbf{w} \in [0, \infty)^d : \sum_{j=1}^d w_j = 1\}$ .  $H$  satisfies the normalization and moment constraints

$$\int_{\Delta_d} dH(\mathbf{w}) = 1, \quad \int_{\Delta_d} \mathbf{w} dH(\mathbf{w}) = d^{-1} \mathbf{1}_d, \quad (1.35)$$

where  $\mathbf{1}_d$  is the  $d$ -vector of ones. As in the bivariate case, the distribution  $H$  determines the interaction between joint extremes, and this can be understood through the pseudo-polar transformation  $\mathbf{W}_i = R_i^{-1} \mathbf{Y}_i$ ,  $R_i = \sum_{j=1}^d W_{i,j}$ . It can be shown that  $\mathbf{W}_i$  has measure  $H$  on  $\Delta_d$  conditional on  $R_i \rightarrow \infty$ .

The limiting cases of independence and dependence in the  $d$ -dimensional setting have an analogous interpretation to the bivariate setting (examples 1 and 2): extremal independence corresponds to a spectral distribution placing equal masses  $d^{-1}$  at the vertices of  $\Delta_d$ , yielding  $G(\mathbf{y}) = \exp\left\{-\sum_{j=1}^d y_j^{-1}\right\}$ , whereas perfect extremal dependence corresponds to a spectral distribution having mass only at the barycenter  $d^{-1} \mathbf{1}_d$  of  $\Delta_d$ , and hence  $G(\mathbf{y}) = \exp\left\{-1/\min(y_1, \dots, y_d)\right\}$ .

Then,  $G$  is the limit distribution of properly normalized componentwise maxima of an independent sample from  $F$ , as the sample size tends to infinity. The margins of  $G$ ,  $G_j$ , for  $j = 1, \dots, d$ , are univariate extreme value distribution functions. They have the corresponding margins,  $F_j$ , of  $F$  in their respective domains of attraction. Alternative representations of the dependence structure of  $G$  are sometimes handy, and this is the case of the *stable tail dependence function*  $l$ , which can be defined

for  $\mathbf{v} = (v_1, \dots, v_d)'$  in terms of the spectral measure  $H$  through

$$l(\mathbf{v}) = V\left(\frac{1}{v_1}, \dots, \frac{1}{v_d}\right) = d \int_{\Delta_d} \max(w_1 v_1, \dots, w_d v_d) H(d\mathbf{w}), \quad \mathbf{v} \in [0, \infty)^d.$$

The partial derivatives of  $l$  can be used to compute the densities of the spectral measure  $H$  on the  $2^d$  faces of the unit simplex  $\Delta_d$ .

As in the bivariate case, there is a similar result in terms of the limit distribution of excesses over a high multivariate threshold, which is beyond the scope of this dissertation. See (Beirlant *et al.*, 2004, p. 277) for a concise description of this representation. More details on multivariate extremes value theory and applications can be found in (Beirlant *et al.*, 2004, Chapters 8 and 9). Segers (2012) provides a comprehensive account of the ways in which models for multivariate extremes can be described. Further reading may be found in Kotz and Nadarajah (2000), Coles (2001), Drees (2001), Reiss and Thomas (2001), and Fougères (2004).

## 1.2 Motivation

The motivation for the developments of this thesis is twofold and is summarized in the following sections.

### 1.2.1 Nonstationarity in bivariate extreme value distributions

Modeling nonstationarity in marginal distributions has been widely studied in applied extreme value modeling (Coles, 2001, Chapter 6), but little effort has been made to extend this idea to more complex settings. The simplest approach was popularized long ago by Davison and Smith (1990), and it is based on indexing the location and scale parameters of the generalized extreme value distribution by a predictor,  $x \in \mathcal{X}$ ,

$$G(y; \mu_x, \sigma_x, \xi) = \exp[-\{1 + \xi(y - \mu_x)/\sigma_x\}_+^{-1/\xi}]. \quad (1.36)$$

In the bivariate setting, much work has been done on developing dependence modeling frameworks, but to the best of our knowledge, there is a scarcity of results in the nonstationary setting related

to extreme dependence structures changing according to a predictor. Some related results have been provided by [Eastoe \(2009\)](#), who introduces a conditionally independent hierarchical model, [Jonathan \*et al.\* \(2014\)](#), who develop methodology for including covariates in the model of [Heffernan and Tawn \(2004\)](#), and the models of [de Carvalho and Davison \(2014\)](#) and [Huser and Genton \(2014\)](#) discussed above. However, none of these results consider the evolution of the dependence structure as their modeling framework.

The gap between the developments in nonstationary marginal distributions and bivariate distributions is a problem that needs to be addressed. In particular, it is necessary to find a bivariate analogous to the [Davison and Smith \(1990\)](#) setting in (1.36). This analogy comes on the heels of what we call *predictor-dependent spectral measure* ([de Carvalho, 2015](#)) and is the basis of our modeling. It is also necessary to produce suitable methodologies to estimate the predictor-dependent spectral distribution, in the sense that satisfy the functional constraint (1.12). These methodologies should be balanced in a way that we account for desirable theoretical properties and easy and efficient computational implementation. To address these issues, we study nonstationary from two prospects. First, we consider a setting where several extreme observations are related to a unique value of the predictor. This setting is motivated by a temperature analysis case study, where the dependence between extreme air temperatures under the forest canopy and in a nearby open field at 14 sites in Switzerland is being analyzed. In this case, the altitude of the 14 sites is used as predictor, and the aim is to assess how extremely high temperatures in the open are related to those under the canopy. The second prospect is related to pairs of observations, one corresponding to the predictor value and the other to the extreme observation. This setting is particularly suited to assess temporal changes in extremal dependence and is motivated by the work of [Poon \*et al.\* \(2003, 2004\)](#), who studied the dependence between stock market returns in the US, UK, France, Germany, and Japan. Considering only the European markets, they found that there was evidence for relatively strong left tail dependence, but most importantly, the results in [Poon \*et al.\* \(2003\)](#) suggest that dependence is not stationary in time. In this case, our main focus is to explore this nonstationarity using a full model for the time-varying dependence structure.

### 1.2.2 Nonparametric estimation of the spectral density

A general review of applications of extreme values, from financial risk (Poon *et al.*, 2003) to environmental studies (Eastoe and Tawn, 2009), reveals the need to augment the literature on modeling the spectral density. The application of bivariate extremes involves a number of choices that needs to be made in order to analyze the data: Should we use parametric or nonparametric modeling? Should we focus on block maxima or threshold exceedances? Do we account for asymptotic dependence or independence? The latest developments have highlighted the difficulty in finding methodologies that are universally applicable and able to answer to all of these questions automatically. Nevertheless, this is not a reason to cease the search for better approaches or improvement of the already existing methods in any of the choices mentioned before. Within the framework of asymptotic dependence, we propose a fully nonparametric approach which is advantageous since neither the form of the bivariate distribution nor the form of dependence on the covariate is assumed. Thereby, we augment the statistical methods for modeling bivariate extreme values, allowing the incorporation of covariate information. This results in the possibility to combine data sets originating from different sources, turning in opportunities for better point estimates and improved inference.

## 1.3 Thesis outline

The work of this dissertation can be divided in two parts that have been developed inside a nonstationary bivariate extreme value setting. Particularly, in the context of covariate or predictor-dependent bivariate extremes. We present each part in individual chapters that can be read separately, since they are self-contained in terms of notation, definitions, and results. For convenience of the reader, some parts mentioned in the introduction may be repeated on later chapters.

Chapter 2 is devoted to developing our first predictor-dependent model, that allows us to assess how extremal dependence can change over different values of a discrete predictor. For estimating our model we propose two double kernel estimators (the second been a reweighed version of the first), which can be seen as an extension of the Nadaraya–Watson estimator where the usual scalar responses are replaced by mean-constrained densities on the unit simplex. We provide numerical experiments that

show that our methods require a much lower computational investment than its closer competitor.

In Chapter 3 we propose our second predictor-dependent model, which is suitable to assess temporal changes in extremal dependence. We consider as a substantial case study the evolution of extremal dependence in some major European stock markets. We provide evidence that lead us to sustain that the dependence is not stationary in time, and we explore this nonstationarity in a more complete manner than has been done before. From a theoretical point of view, we introduce the notion of predictor-dependent spectral measure, propose a nonparametric estimator and settled conditions for desirable asymptotic properties. Details on practical implementation are also provided.

In Chapter 4 we discuss limitations of proposed models and comment on possible directions of future research.

# 2

## Spectral density regression for bivariate extremes

---

### 2.1 Introduction

Modeling nonstationarity in marginal distributions has been the focus of much recent literature in applied extreme value modelling. The simplest approach was popularized long ago by [Davison and Smith \(1990\)](#), and it is based on indexing the location and scale parameters of the generalized extreme value distribution by a predictor, say by considering

$$G_{(\mu_x, \sigma_x, \xi)}(y) = \exp[-\{1 + \xi(y - \mu_x)/\sigma_x\}_+^{-1/\xi}]. \quad (2.1)$$

See also [Coles \(2001, Ch. 6\)](#), [Chavez-Demoulin and Davison \(2005\)](#), [Eastoe and Tawn \(2009\)](#), and [Chavez-Demoulin et al. \(2015\)](#), for related approaches.

In areas such as environmental impact assessment or financial risk management, one is often concerned in assessing how extreme outcomes of two or more variables are related, and the mathematical basis for such modeling is that of statistics of bivariate extremes. In such contexts, extremal dependence is often interpreted as a synonym of risk, and when modeling bivariate extremes we are naturally led to the bivariate extreme value distribution. It is well known that the bivariate extreme value distribution, depends on an infinite-dimensional parameter ( $H$ ) ([Coles, 2001](#), Theorem 8.1), and it can be written as

$$G_H(y_1, y_2) = \exp \left\{ -2 \int_0^1 \max \left( \frac{w}{y_1}, \frac{1-w}{y_2} \right) dH(w) \right\}, \quad (2.2)$$

for  $y_1, y_2 > 0$ , where  $H$  is the so-called spectral distribution function, which is a distribution function

on  $[0, 1]$  obeying the moment constraint

$$\int_0^1 w \, dH(w) = 1/2. \quad (2.3)$$

Roughly speaking, the more mass  $H$  puts close to  $1/2$  the higher the level of extremal dependence, whereas the more mass  $H$  puts close to  $0$  and  $1$  the more independent the extremes are. Since the object of interest in bivariate extremes is intrinsically nonparametric, nonparametric methods have become a natural tool for estimation. A survey on nonparametric estimation of extremal dependence can be found in [Kiriliouk et al. \(2015\)](#).

And how to model ‘nonstationary bivariate extremes’ if one must? Surprisingly, by comparison to the marginal case, approaches to modelling nonstationarity in the extremal dependence structure have received relatively little attention. However, in many settings of applied interest, it seems natural to regard risk from a covariate-adjusted viewpoint, allowing for extremal dependence to increase/decrease according to a covariate. But to develop ideas of covariate-adjusted risk using statistics of bivariate extremes we need to allow for nonstationary extremal dependence structures, so to assess the dynamics governing extremal dependence of pairs of variables of interest.

In this chapter we discuss methods for modeling nonstationary extremal dependence structures. Our approach can be regarded as an analogue to the bivariate setting of the Davison–Smith approach in (2.1), and it is based on indexing the parameter of the bivariate extreme value distribution ( $H$ ) with a covariate, i.e. considering  $\{H_x : x \in \mathcal{X} \subset \mathbb{R}\}$ , and taking

$$G_{H_x}(y_1, y_2) = \exp \left\{ -2 \int_0^1 \max \left( \frac{w}{y_1}, \frac{1-w}{y_2} \right) dH_x(w) \right\}, \quad (2.4)$$

for  $y_1, y_2 > 0$ . Obviously, the  $H_x$ —to which we refer as predictor-dependent spectral measures—will need to obey the moment constraints (2.3), for every  $x$ , so that  $G_{H_x}$  is a valid bivariate extreme value distribution.

A main goal of this chapter is on modeling families of spectral densities indexed by a covariate, and we refer to our approach as a spectral ‘density regression.’ In terms of estimation, we propose a nonparametric estimator, that have connections with the Nadaraya–Watson estimator ([Nadaraya, 1964](#);

Watson, 1964). While ‘density regression,’ could sound like a misnomer, we underscore that similar terminology has been used on related topics for referring to contexts where the interest is in estimating a predictor-dependent family of densities; see Dunson *et al.* (2007). A related approach to the one discussed here has been recently proposed by de Carvalho and Davison (2014) who introduced a model for the case where several bivariate extremal distributions are linked through the action of a covariate. A challenge with their model is however that inference entails intensive constrained optimization problems. In comparison with de Carvalho and Davison (2014) approach, our model avoids the need of specifying a tilting function, it allows for straightforward extrapolation to unobserved covariate values, it allows for estimation of covariate-adjusted spectral densities (and not only spectral measures), and it is computationally straightforward. Another related approach is that of Huser and Genton (2014) who use nonstationary max-stable dependence structures to develop nonstationary models for spatial extremes in which covariates can be incorporated.

In Section 2.2 we introduce spectral density regression, propose a method for inference and estimation, and give details on computational implementation. A simulation study is conducted in Section 2.3, while an application to extreme forest temperatures is given in Section 2.4. Section 2.5 offers conclusions. The Appendix includes additional empirical reports.

## 2.2 Spectral density regression model

### 2.2.1 Bivariate statistics of extremes and $K$ -sample setting

Let  $\{(Y_{i,1}, Y_{i,2})\}_{i=1}^N$  be a sequence of independent identically distributed random vectors with unit Fréchet marginal distributions,  $F_1(y) = F_2(y) = \exp(-1/y)$ , for  $y > 0$ . The underlying theory for modeling bivariate extremes is based in the so-called Pickands’ (1981) representation theorem, a convergence result which provides the limiting distribution of the componentwise standardized maximum,

$$(M_{1,N}, M_{2,N}) = N^{-1} \left( \max_{i=1, \dots, N} \{Y_{i,1}\}, \max_{i=1, \dots, N} \{Y_{i,2}\} \right).$$

Pickands (1981) established that

$$P(M_{1,N} \leq y_1, M_{2,N} \leq y_2) \rightarrow G_H(y_1, y_2), \quad (2.5)$$

as  $N \rightarrow \infty$ , where  $y_1, y_2 > 0$ , provided the limit exists and is non-degenerate; see also Coles (2001, Theorem 8.1). Here  $G$  is the bivariate extreme value distribution defined in (2.2) and is in one to one correspondence with  $H$ , the spectral distribution function that is mean-constrained according to (2.3). The spectral measure  $H$  provides relevant information on extremal dependence, and can be used to describe the extremal dependence structure of the random vector  $(Y_1, Y_2)$ . This can be understood through a pseudo-polar transformation, where we map  $(Y_{1,1}, Y_{1,2}), \dots, (Y_{N,1}, Y_{N,2})$ , to pseudo-angular and radial variates

$$(W_i, R_i) = \left( \frac{Y_{i,1}}{Y_{i,1} + Y_{i,2}}, Y_{i,1} + Y_{i,2} \right), \quad i = 1, \dots, N.$$

de Haan and Resnick (1977) showed that  $W_i$  has measure  $H$  on  $[0,1]$  conditional on  $R_i \rightarrow \infty$ . If  $W$  and  $R$  are general terms of the sequence  $\{(W_i, R_i)\}_{i=1}^N$ , the latter result tells us that when the radius  $R$  is large, the pseudo-angle  $W$  is approximately distributed according to  $H$ , and approximately independent of  $R$ . The limiting cases of the distribution  $H$  are given by asymptotic independence, whereby all mass is placed at the boundaries of  $[0, 1]$ , giving  $G(y_1, y_2) = \exp\{-(y_1^{-1} + y_2^{-1})\}$ , and by complete dependence, whereby all mass is placed at the centre of the interval, yielding  $G(y_1, y_2) = \exp\{-\max(y_1^{-1}, y_2^{-1})\}$ . We refer to situations where  $H$  has mass away from the vertices as asymptotic dependence, and this will be the framework of our modeling. Throughout we assume that  $H$  is absolutely continuous with spectral density  $h(w) = dH(w)/dw$ .

Given a sample  $(Y_{1,1}, Y_{1,2}), \dots, (Y_{N,1}, Y_{N,2})$ , we may construct proxies for the unobservable pseudo-angles  $W_i$  by setting

$$(W_i, R_i) = \left( \frac{\hat{Y}_{i,1}}{\hat{Y}_{i,1} + \hat{Y}_{i,2}}, \hat{Y}_{i,1} + \hat{Y}_{i,2} \right), \quad i = 1, \dots, N.$$

where

$$\hat{Y}_{i,1} = -1/\{\log \hat{F}_1(Y_{i,1})\}, \quad \hat{Y}_{i,2} = -1/\{\log \hat{F}_2(Y_{i,2})\},$$

and where  $\hat{F}_1$  and  $\hat{F}_2$  are estimators of the marginal distribution functions  $F_1$  and  $F_2$ . A robust choice for  $\hat{F}_1$  and  $\hat{F}_2$  is the pair of univariate empirical distribution functions, normalized by  $N+1$  rather than

by  $N$  to avoid division by zero. For a high enough threshold  $u$ , the collection of angles  $\{W_i : R_i > u\}$  can be regarded as an approximate sample from the spectral measure  $H$ . Parametric or nonparametric inference on  $H$  may then be based upon the sample  $\{W_i : R_i > u\}$ .

Similarly to [de Carvalho and Davison \(2011, 2014\)](#), below we work under the so-called *K-population setting for bivariate extremes*. Indeed, our applied setting of interest in [Section 2.4](#) is one where the raw data consists of  $K$  pairs,

$$(Y_{1,1,k}, Y_{1,2,k}), \dots, (Y_{N_k,1,k}, Y_{N_k,2,k}), \quad k = 1, \dots, K,$$

plus a covariate  $x_k$ , and by applying similar principles as discussed above, we end up with  $K$  samples of pseudo-angles, i.e we work with

$$\mathbf{w}_k = (W_{1,k}, \dots, W_{n_k,k}).$$

The  $\mathbf{w}_k$  should be regarded as an approximate sample of  $n_k$  pseudo-angles from the spectral measure corresponding to the  $k$ th population,  $H_k$ . Thus, in the  $K$ -sample setting for bivariate extremes, data are of the type  $\{(x_k, \mathbf{w}_k)\}_{k=1}^K$ . We assume that  $H_k$  is absolutely continuous with spectral density  $h_k = dH_k/dw$ . The combined sample size is denoted by  $n = n_1 + \dots + n_K$ .

### 2.2.2 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 indexed by a covariate  $x \in \mathcal{X} \subseteq \mathbb{R}$ . Analogously, we say that the family

$$\{H_x : x \in \mathcal{X}\}$$

is a set of *pd spectral measures* if

$$\int_0^1 dH_x(w) = 1, \quad \int_0^1 w dH_x(w) = \frac{1}{2}, \quad x \in \mathcal{X}. \quad (2.6)$$

Pd spectral measures allow us to assess how extremal dependence evolves over a certain covariate  $x$ , i.e., they allow us to model nonstationary extremal dependence structures; further details on pd spectral

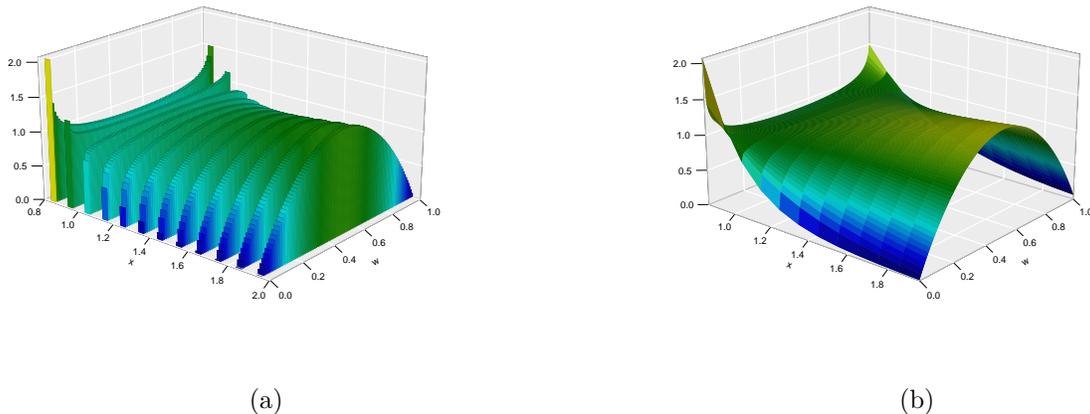


Figure 2-1. (a) Histograms for set of pseudo-angles and (b) interpolated spectral surface. Both figures were generated from  $h_x(w) = \beta(w; x, x)$  with  $x$  taking values in a grid between 0.85 and 2.

measures can be found in [de Carvalho \(2015, Section 2.3\)](#).

Suppose  $H_x$  is absolutely continuous for all  $x \in \mathcal{X}$ . We define the pd spectral density as  $h_x(w) = dH_x(w)/dw$ , and following [de Carvalho \(2015\)](#) we refer to the set

$$\{h_x(w) : w \in [0, 1], x \in \mathcal{X}\}$$

as the *spectral surface*. Spectral surfaces can be readily constructed from parametric models for the spectral density; see, for instance, [Coles \(2001, Section 8.2.1\)](#). Examples of spectral surfaces can be found in [Figures 2-1\(b\), 2-3, and 2-8](#).

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 in [\(2.4\)](#) can be regarded as an analogue of the Davison–Smith paradigm in [\(2.1\)](#), but for the bivariate setting.

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 challenging to construct nonparametric estimators able to yield valid pd spectral densities. Indeed, any such estimator,  $\hat{h}_x$ , needs to obey the moment constraint, i.e.,  $\int_0^1 w \hat{h}_x(w) dw = 1/2$ , for all  $x \in \mathcal{X}$ . In the next section we introduce one such estimator.

### 2.2.3 Double kernel estimator

Figure 2-1 resumes key ideas underlying the construction of our estimator. Figure 2-1(a) shows histograms for sets of pseudo-angles generated from  $h_x(w) = \beta(w; x, x)$  with  $x$  taking values in a grid between 0.85 and 2. For each value of  $x$  in the grid  $\{x_k\}$ , we would like to estimate the associated spectral density, and then interpolate for unobserved values of  $x$ , as shown in Figure 2-1(b).

Suppose that we have a method to compute  $\tilde{h}_k$ , the spectral densities estimate at every  $x_k$ ; in Figure 2-1(a)  $\tilde{h}_k$  would correspond to the histogram estimates, but for reasons that will become obvious below we will not work with these estimates. Estimation of the pd spectral density on the basis of data available on the  $K$ -population setting,  $\{(x_k, \mathbf{w}_k)\}_{k=1}^K$ , entails two challenges:

1. Although we want to estimate  $h_x$  at every  $x \in \mathcal{X}$ , we only have data at  $x_1, \dots, x_K$ .
2. We need to impose to  $\hat{h}_x$  and  $\tilde{h}_k$  the corresponding moment constraints.

To estimate the spectral surface,  $h_x$ , we propose the estimator

$$\hat{h}_x(w) = \frac{\sum_{k=1}^K \mathbb{K}_b(x - x_k) \tilde{h}_k(w)}{\sum_{k=1}^K \mathbb{K}_b(x - x_k)}, \quad (2.7)$$

for  $w \in (0, 1)$ , where  $\mathbb{K}_b$  is a kernel density estimator and  $b > 0$  is a bandwidth parameter controlling smoothing in the  $x$ -direction. The estimator in (2.7) is similar to the well-known Nadaraya–Watson estimator (Nadaraya, 1964; Watson, 1964), but here—contrary to the usual nonparametric regression setting—the responses are spectral densities, and hence infinite-dimensional objects; further details on kernel regression can be found in Wand and Jones (1994, Chapter 5). If the spectral density estimates at every  $x_k$  are such that

$$\int_0^1 \tilde{h}_k(w) \, dw = 1, \quad \int_0^1 w \tilde{h}_k(w) \, dw = 1/2,$$

for  $k = 1, \dots, K$ , then

$$\begin{aligned} \int_0^1 \hat{h}_x(w) \, dw &= \frac{\sum_{k=1}^K \mathbb{K}_b(x - x_k) \int_0^1 \tilde{h}_k(w) \, dw}{\sum_{k=1}^K \mathbb{K}_b(x - x_k)} = 1, \\ \int_0^1 w \hat{h}_x(w) \, dw &= \frac{\sum_{k=1}^K \mathbb{K}_b(x - x_k) \int_0^1 w \tilde{h}_k(w) \, dw}{\sum_{k=1}^K \mathbb{K}_b(x - x_k)} = 1/2. \end{aligned}$$

for all  $x \in \mathcal{X}$ . Put differently, valid spectral surfaces can be obtained from our estimator in (2.7) if at every  $x_k$  we estimate a valid spectral density,  $\tilde{h}_k$ , i.e. a density on  $[0, 1]$  obeying the moment constraint. To ensure that each spectral density estimate,  $\tilde{h}_k$ , obeys the normalization and marginal moment constraints, we use the smooth Euclidean likelihood estimator (de Carvalho *et al.*, 2013), which for a sample of  $n_k$  pseudo-angles is defined as

$$\tilde{h}_k(w) = \sum_{i=1}^{n_k} \tilde{p}_{i,k} \beta(w; W_{i,k} \nu, (1 - W_{i,k}) \nu), \quad (2.8)$$

for  $w \in (0, 1)$ , where

$$\tilde{p}_{i,k} = \frac{1}{n_k} \{1 - (\bar{W}_k - 1/2) S_k^{-2} (W_{i,k} - \bar{W}_k)\}, \quad (2.9)$$

for  $i = 1, \dots, n_k$  and  $k = 1, \dots, K$ . Here,  $\bar{W}_k$  and  $S_k^2$  denote the sample mean and sample variance of  $W_{1,k}, \dots, W_{n_k,k}$ , that is,

$$\bar{W}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} W_{i,k}, \quad S_k^2 = \frac{1}{n_k} \sum_{i=1}^{n_k} (W_{i,k} - \bar{W}_k)^2.$$

The parameter  $\nu > 0$  in (2.8) is a concentration parameter, responsible for controlling the amount of smoothing, in the  $w$ -direction. A method for parameter selection using cross-validation is discussed in Section 2.2.4. The estimator in (2.9) can be understood as an empirical likelihood-based kernel density estimator (Chen, 1997); the weights in (2.9) differ from the usual  $1/n_k$  appearing in kernel density estimation, as they are obtained through an empirical likelihood-based method, in order to produce estimates which obey the moment constraint. Specifically, the  $\tilde{p}_{i,k}$  in (2.9) are Euclidean likelihood weights (Owen, 2001, pp. 63–66), i.e., are the solution to the optimization problem

$$\begin{aligned} \max_{\mathbf{p}_k \in \mathbb{R}^{n_k}} \quad & -\frac{1}{2} \sum_{i=1}^{n_k} (n_k p_{i,k} - 1)^2 \\ \text{s.t.} \quad & \sum_{i=1}^{n_k} p_{i,k} = 1 \\ & \sum_{i=1}^{n_k} W_{i,k} p_{i,k} = 1/2. \end{aligned} \quad (2.10)$$

Finally, our estimator in (2.7) can be rewritten as a double kernel estimator

$$\hat{h}_x(w) = \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} \tilde{p}_{i,k} \mathbb{K}_b(x - x_k) \beta(w; W_{i,k} \nu, (1 - W_{i,k}) \nu)}{\sum_{k=1}^K \mathbb{K}_b(x - x_k)}. \quad (2.11)$$

Next we provide details on practical implementation.

### 2.2.4 Details on implementation

We select the tuning parameters via leave-one-out cross-validation for each parameter separately. Specifically, for the concentration parameter  $\nu$  we choose

$$\nu^* = \arg \min_{\nu > 0} \sum_{k=1}^K \sum_{i=1}^{n_k} -\log\{\tilde{h}_{-i}(W_{i,k})\}, \quad (2.12)$$

where

$$\tilde{h}_{-i}(w) = \sum_{j \neq i} \tilde{p}_{j,k} \beta(w; W_{j,k} \nu, (1 - W_{j,k} \nu)),$$

whilst for the bandwidth  $b$  we select

$$b^* = \arg \min_{b > 0} \int_0^1 \sum_{k=1}^K \{\tilde{h}_k(w) - \hat{h}_{-k}(w)\}^2 dw, \quad (2.13)$$

with

$$\hat{h}_{-k}(w) = \frac{\sum_{j \neq k} \mathbb{K}_b(x_j - x_k) \tilde{h}_j(w)}{\sum_{j \neq k} \mathbb{K}_b(x_j - x_k)}.$$

In principle,  $\mathbb{K}_b$  should be a symmetric and unimodal density. While there are many kernel functions that verify these basic requirements, it is well known that the choice of the kernel has little impact on the corresponding estimators; see [Wand and Jones \(1994, Chapter 2\)](#) and references therein. In practice, we use a normal kernel.

Next, we give computational details on how to implement the double kernel estimator using `k.smooth` from the R package `stats` ([R Core Team, 2014](#)).

**Pseudocode for double kernel estimator**

1. Compute  $\nu^*$  and  $b^*$  using `optim` according to (2.12) and (2.13), respectively.
2. Construct a grid  $\{w_j\}_{j=1}^J \in (0, 1)$  and compute  $\tilde{h}_k(w_j)$  according to (2.8).
3. **for**  $j = 1, \dots, J$ , **do**:  
     Compute  $\hat{h}_x(w_j)$  using `ksmooth` with data

$$\{(x_k, \tilde{h}_k(w_j))\}_{k=1}^K.$$

## 2.3 Simulation study

### 2.3.1 Models, configurations, and preliminary experiments

We construct samples of pseudo-angles  $\{\mathbf{w}_k\}_{k=1}^K$  from the Dirichlet spectral surface, a covariate-adjusted extension of the Dirichlet model (Coles and Tawn, 1991), based on the pd spectral density

$$h_x(w) = \frac{a_x b_x \Gamma(a_x + b_x + 1) (a_x w)^{a_x - 1} \{b_x (1 - w)\}^{b_x - 1}}{2 \Gamma(a_x) \Gamma(b_x) \{a_x w + b_x (1 - w)\}^{a_x + b_x + 1}}. \quad (2.14)$$

Here  $a_x : \mathcal{X} \mapsto (0, \infty)$ ,  $b_x : \mathcal{X} \mapsto (0, \infty)$ , and  $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$ . The values of the parameters in (2.14) are chosen to produce two scenarios: a symmetric Dirichlet spectral surface with  $(a_x, b_x) = (x, x)$ , where  $x \in \mathcal{X}_{\text{sDir}} = [1.5, 4]$ ; an asymmetric Dirichlet spectral surface with  $(a_x, b_x) = (x, 100)$ , where  $x \in \mathcal{X}_{\text{aDir}} \in [0.9, 4]$ . For each of the two scenarios, we consider  $K \in \{20, 50, 100\}$ , and for every  $K$ , the values for  $\{n_k\}_{k=1}^K$  are chosen randomly according to two configurations, both based on the actual number of pseudo-angles in the data application (see Table 2.2):

- Configuration 1: corresponds to  $K$  sampled values from the number of pseudo-angles in the data application.
- Configuration 2: samples  $K$  values from the same set, but multiplied by a factor of 5.

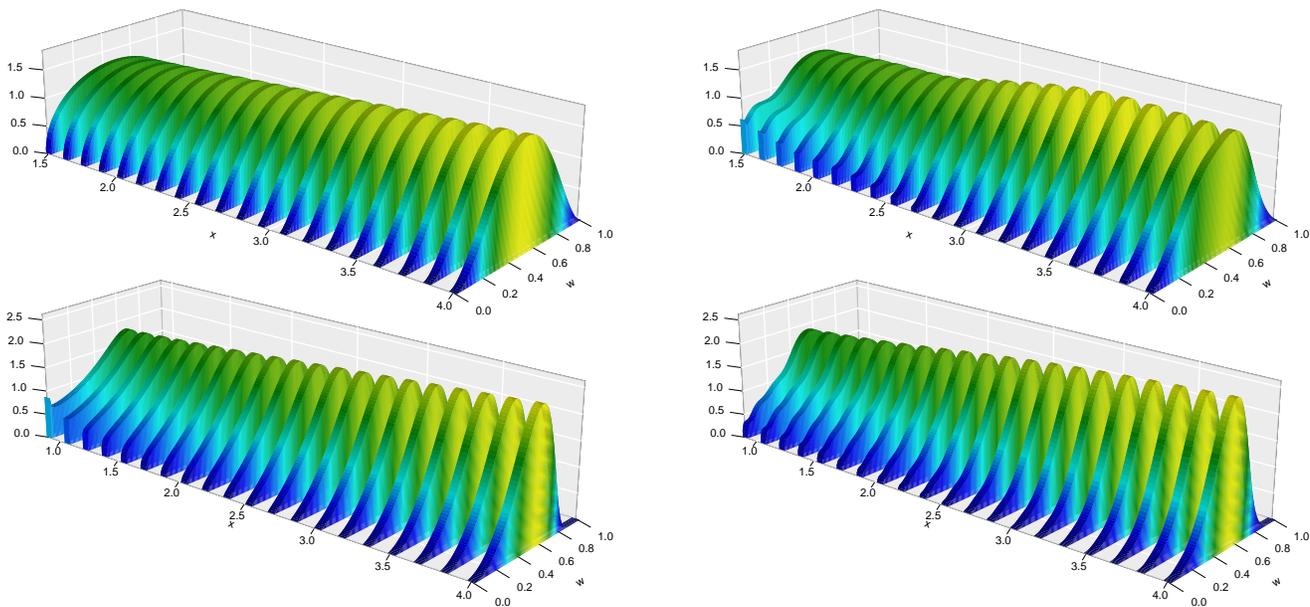


Figure 2-2. True (left) cross sections of the spectral surface and corresponding estimates (right) from the symmetric (top) and asymmetric (bottom) Dirichlet predictor-dependent models, for  $K = 20$  values of the predictor and Configuration 1.

This gives rise to six different simulation schemes for each of the two predictor-dependent models.

We start with a single-run experiment. Figure 2-2 shows true and estimated spectral densities from the symmetric (top) and asymmetric (bottom) Dirichlet models described above, for  $K = 20$  values of the predictor and Configuration 1. Spectral density estimates were computed using the smooth Euclidean estimator in (2.8). If  $\mathbb{K}_b$  is chosen as a normal kernel with standard deviation  $b$ , and if we smooth over all the predictor space using the double kernel estimator for both configurations and  $K \in \{20, 50, 100\}$ , we obtain what is shown in Figure 2-3. Figure 2-3 corresponds to the symmetric Dirichlet spectral surface, where extremal dependence increases as a function of the predictor. The analogue for the asymmetric Dirichlet spectral surface is displayed in Figure 2-4

The single-run experiment in Figures 2-3 and 2-4 allows us to illustrate strengths and limitations of the double kernel estimator. Pointwise estimation is troublesome at the edge of the predictor space, due to boundary bias of  $\mathbb{K}_b$  which is a drawback of kernel-based estimators on bounded domains (Härdle, 1990, Section 4.4, and references therein). The double kernel estimator seems to have more

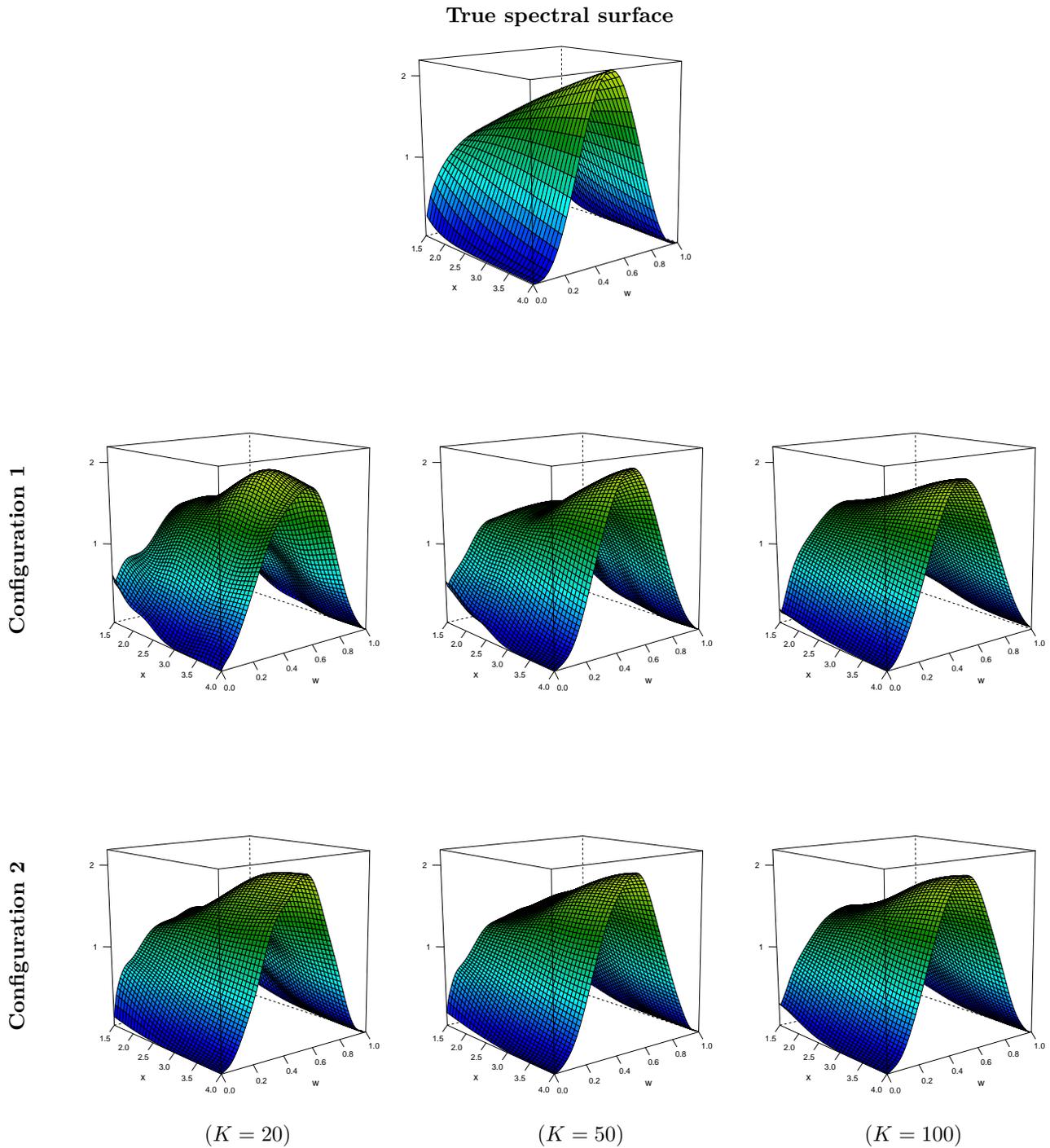


Figure 2-3. On the top: true spectral surface from the symmetric Dirichlet predictor-dependent model detailed in Section 2.3.1, followed by spectral surface estimates for Configuration 1 (above) and 2 (below).

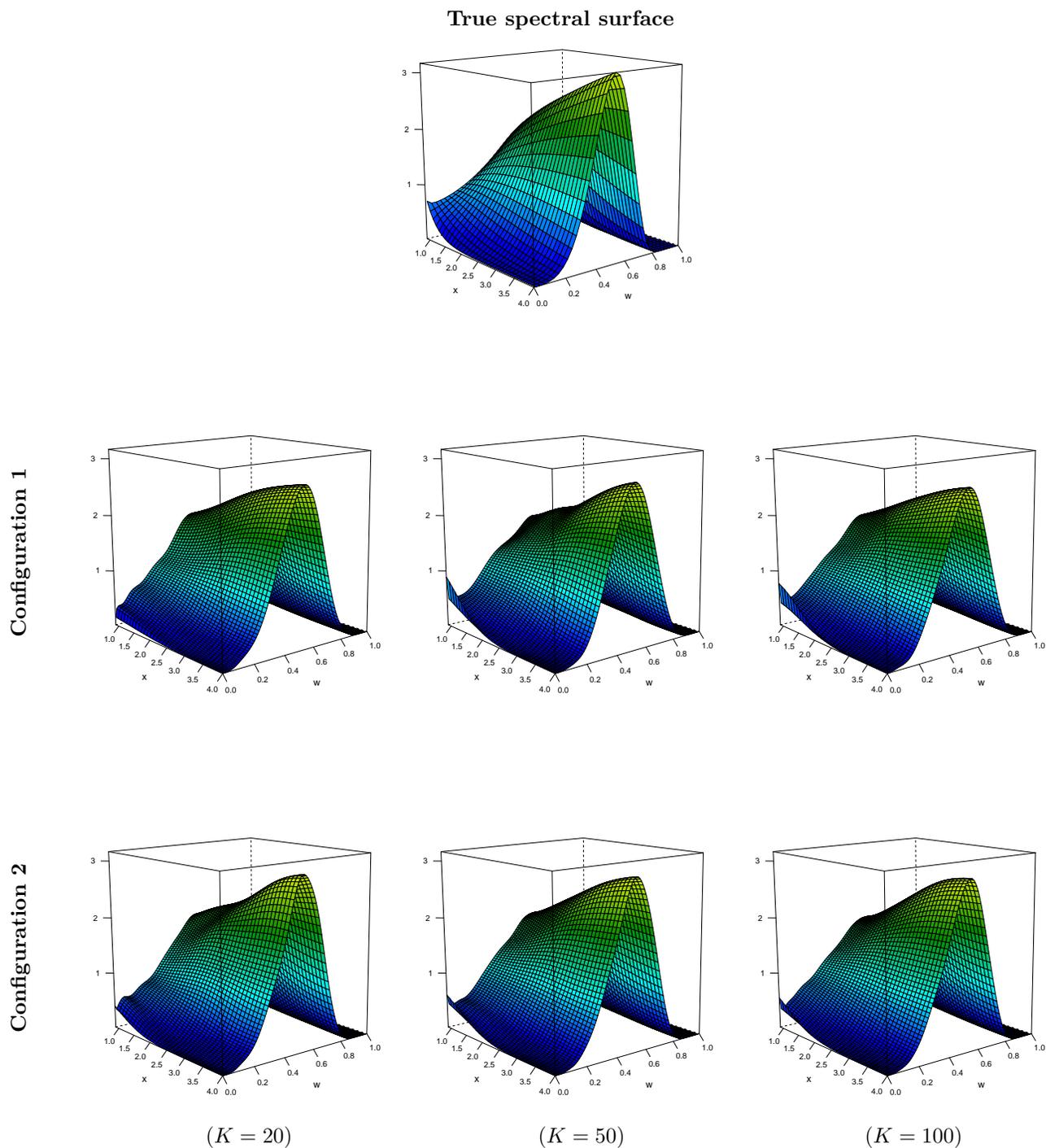


Figure 2-4. On the top: true spectral surface from the asymmetric Dirichlet predictor-dependent model detailed in Section 2.3.1, followed by estimate spectral surfaces for Configurations 1 (above) and 2 (below).

Table 2.1. Mean integrated absolute error estimates of the spectral surface computed over 1000 samples for the data-generating configurations discussed in Sect. 2.3.1.

Model	Conf.	$K = 20$	$K = 50$	$K = 100$
Symmetric	1	0.2877	0.2785	0.2698
Dirichlet	2	0.2089	0.1897	0.1717
Asymmetric	1	0.4254	0.4158	0.4146
Dirichlet	2	0.2958	0.2940	0.2953*

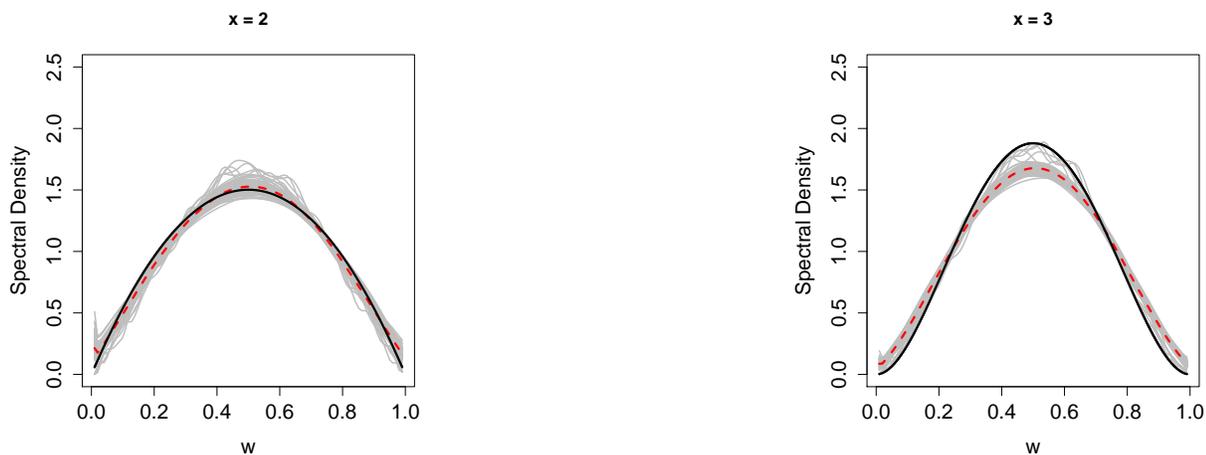
difficulties estimating the asymmetric spectral surface, probably due to the need to recover a more complicated surface. In spite of these limitations, our estimator recovers satisfactorily the shape of the true spectral surface, and thus is able to reproduce satisfactorily the evolution of extremal dependence over the predictor. Another interesting aspect is that the performance of the estimator seems to be more sensitive to changes in  $n_k$  (the number of pseudo-angles for every value of the predictor) rather than changes in  $K$  (the number of predictor values).

### 2.3.2 Simulation results

Cross sections of the spectral surface give rise to spectral densities for certain values of the predictor. To assess the precision of the estimates, Figure 2-5 displays trajectories of 100 estimates of these cross sections along with their Monte Carlo means, for  $K = 100$  and configuration 1 detailed above. These trajectories allow us to illustrate the performance of our estimator under different dependence dynamics. The top panel of Figure 2-5 displays the results for the symmetric Dirichlet spectral densities, where we can see the limitations due to boundary bias that were discussed in Section 2.3.1, mostly for  $x = 2$ . The same plot shows that extremal dependence is overestimated by some of the simulations, whereas it is slightly underestimated for  $x = 3$ . The asymmetric Dirichlet spectral densities, presented in the bottom pannel of Figure 2-5, display less dispersed estimates than their symmetric counterparts, and the asymmetry does not seem to be a major issue. In the two sets of spectral densities, the estimator shows a positive performance in recovering the different shapes of the densities, and Monte Carlo means produce reasonable estimates.

Table 3.1 shows the performance results in terms of the mean integrated absolute error, computed

## Cross sections of symmetric Dirichlet spectral surface



## Cross sections of asymmetric Dirichlet spectral surface

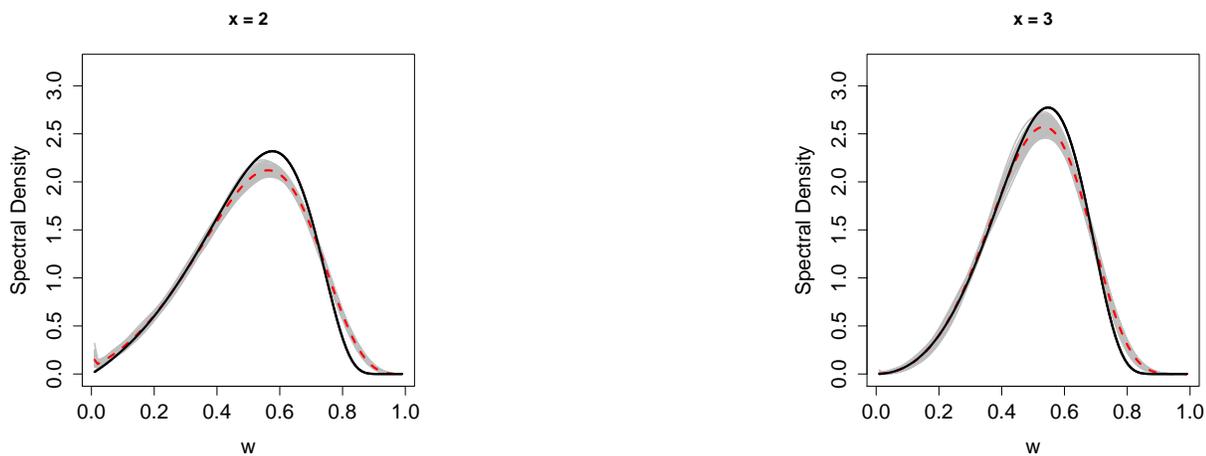


Figure 2-5. Trajectories of 100 randomly selected estimates of cross sections of the spectral surface (gray lines), using  $K = 100$  and configuration 1, as well as their corresponding true values (solid line) and Monte Carlo means (dashed line).

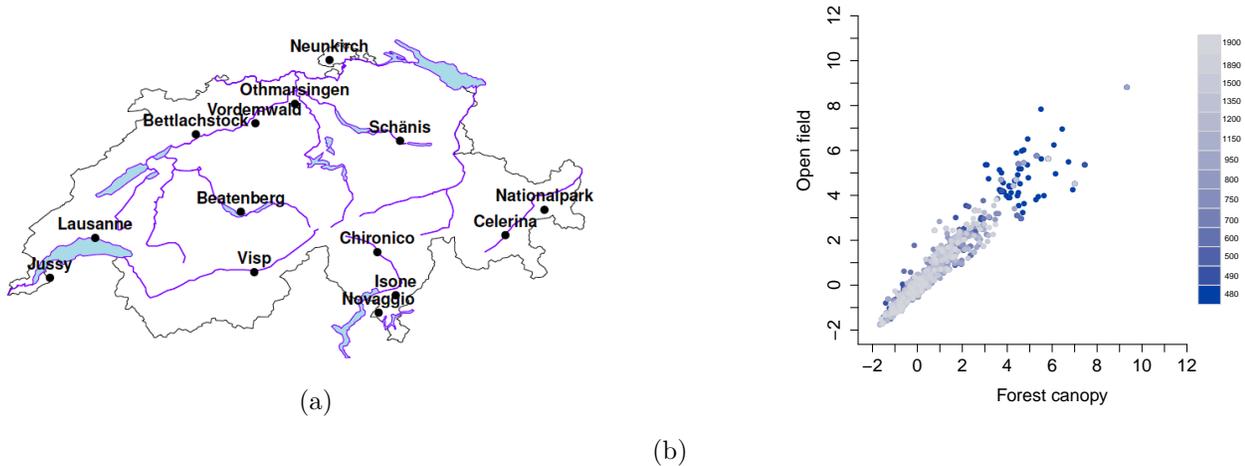


Figure 2-6. (a) Location of air temperature monitoring stations. For each site, air temperatures under the forest canopy and in a nearby open field are available. (b) Scatterplots of pairs of temperatures using an altitude-varying color palette measured in meters, for maxima in forest canopy versus open field.

over 1000 samples. Results confirm what we already observed for the single-run experiment in Figure 2-3; under the same configuration for  $K$ , changes in the number of pseudo-angles cause great improvements in the performance of the estimator. On the opposite, if we fix the number of pseudo-angles and increase  $K$ , gains are not as significant. Overall, simulations confirm that the double kernel estimator produces reasonable estimates of the spectral surface. Monte Carlo mean spectral surfaces are reported in the Appendix.

## 2.4 Extreme forest temperature illustration

### 2.4.1 Data description and preprocessing

The data were gathered from the Long-term Forest Ecosystem Research database maintained by LWF (Langfristige Waldokosystem Forschung), and consist of air temperatures under the forest canopy and in a nearby open field at 14 monitoring stations in Switzerland. (Figure 2-6(a)). The raw data consist of a pair of covariates plus two series of air temperatures per site, measured in circular metal shelters two meters above ground every 10 minutes since 1997.

We use the same preprocessing steps as in [Ferrez \*et al.\* \(2011\)](#); in particular, we take daily maxima

## 2.4. EXTREME FOREST TEMPERATURE ILLUSTRATION

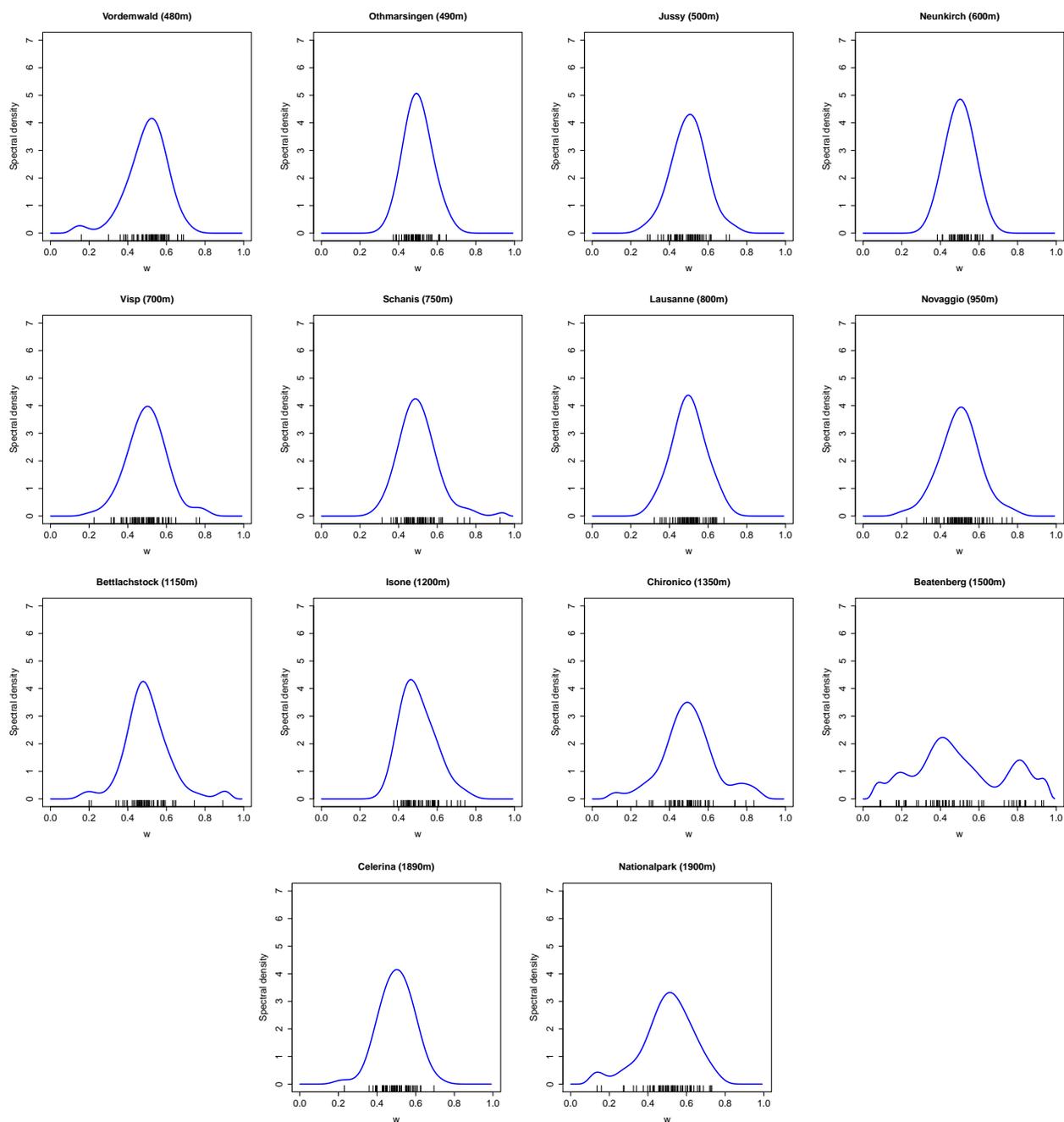


Figure 2-7. Spectral density estimates using the smoothed Euclidean likelihood estimator in (2.8) in increasing order of altitude. The associated pseudo-angles are plotted using a rug at the bottom of every plot.

of the residual series resulting from removal of the annual cycle in both location and scale, i.e., we subtract a periodic mean and divide by a periodic standard deviation. After thresholding the residuals of each sample at the 98% empirical quantile, we reduce the initial 38.923 observations to 785 pairs of residuals, one of the pairs being under the forest canopy and the other being in the open field. The number of pseudo-angles in each site vary from 42 to 65, whereas altitude varies from 480 to 1900 m; see Table 2.2 in Appendix.

### 2.4.2 Altitude-adjusted extremal dependence

Our aim here is to assess the dynamics governing the relation between extremely high temperatures in the open and under the canopy over different altitudes. Our covariate of interest is thus altitude (above sea level) of monitoring stations, and this is motivated by earlier experiments conducted by Ferrez *et al.* (2011, p. 999) who suggested that extremal dependence between temperatures under the canopy and in the open field could be linked to altitude.

We first estimate the extremal dependence of temperatures in every site. Figure 2-7 shows the estimated spectral densities using the smooth Euclidean likelihood estimator in (2.8). We can see clear changes in the dynamics of the extremal dependence. Altitudes between 1400 m and 1650 m present more dispersion—as we can see for example in the spectral density estimate corresponding to Beatenberg—but, in general, strong dependence of temperatures is noted in all sites.

Figure 2-8 shows the extremal dependence when we smooth through all altitudes between 480 m and 1900 m using the double kernel estimator. All in all, we see substantial changes in the extremal behavior, although it is not possible to identify a unique pattern that indicates a monotone evolution of the dependence. In general, we see strong dependence for almost all altitudes, with the exception of values between and 1400 m and 1650 m, where we can see more dispersion and consequently a decrease in the extremal dependence.

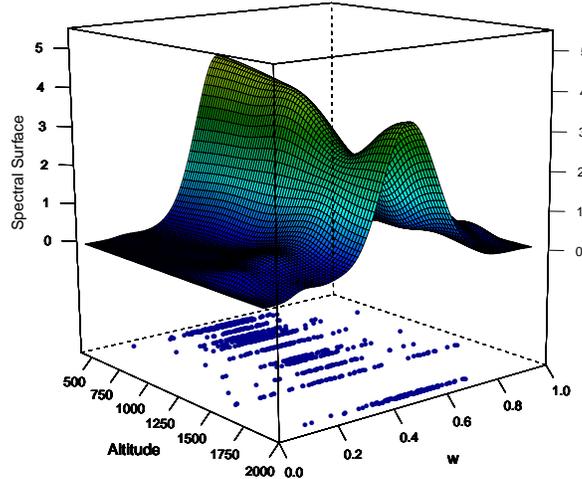


Figure 2-8. Spectral surface estimate using the double kernel estimator in (2.11); pseudo-angles are overlaid on the bottom of the box.

## 2.5 Final remarks

We propose a density regression model that allows us to assess the changes in the extremal dependence structure over the values of a discrete predictor. This is a first step to tackle the gap between the developments in non-stationary marginal distributions and bivariate distributions. We perform inference by introducing a non-parametric double kernel estimator that smooths in two steps: first in the extreme data direction using the Euclidean likelihood estimator of [de Carvalho \*et al.\* \(2013\)](#); and then in the covariate direction through an approach similar to the Nadaraya–Watson estimator, but where responses are spectral density estimates.

Our model is related to the spectral density ratio model of [de Carvalho and Davison \(2014\)](#) in the sense that covariates can be incorporated, but rather than just linking extremal distributions, our model assesses directly the evolution of extremal dependence. Furthermore, implementation of our estimator is straightforward, and inference is computationally convenient.

We test our methods in a temperature data application where altitude is considered a variable of interest. Results suggest an impact of the altitude on the extremal dependence of maximum temperatures under the forest and on a open field, showing the need to consider nonstationarity in the extreme value dependence structure.

## 2.6 Appendix

### 2.6.1 Appendix A: Monte Carlo mean spectral surfaces

Figures 2-9 and 2-10 show Monte Carlo mean spectral surfaces for the two simulation scenarios described in Section 2.3.

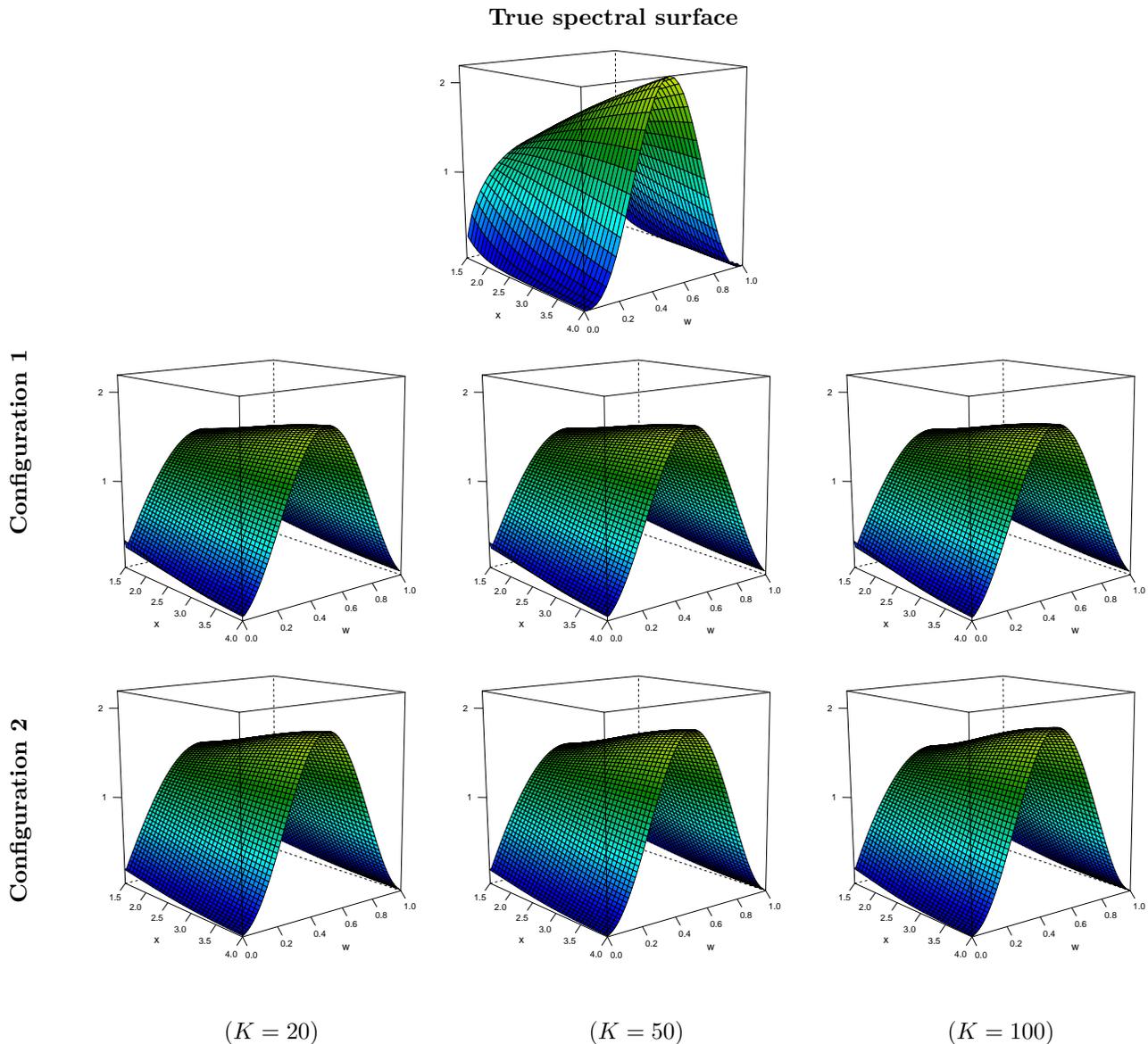


Figure 2-9. On the top: true spectral surface from the symmetric Dirichlet predictor-dependent model detailed in Section 2.3.1, followed by Monte Carlo mean spectral surfaces for configurations 1 (left) and 2 (right) with  $K = 20$  (second row),  $K = 50$  (third row) and  $K = 100$  (fourth row).

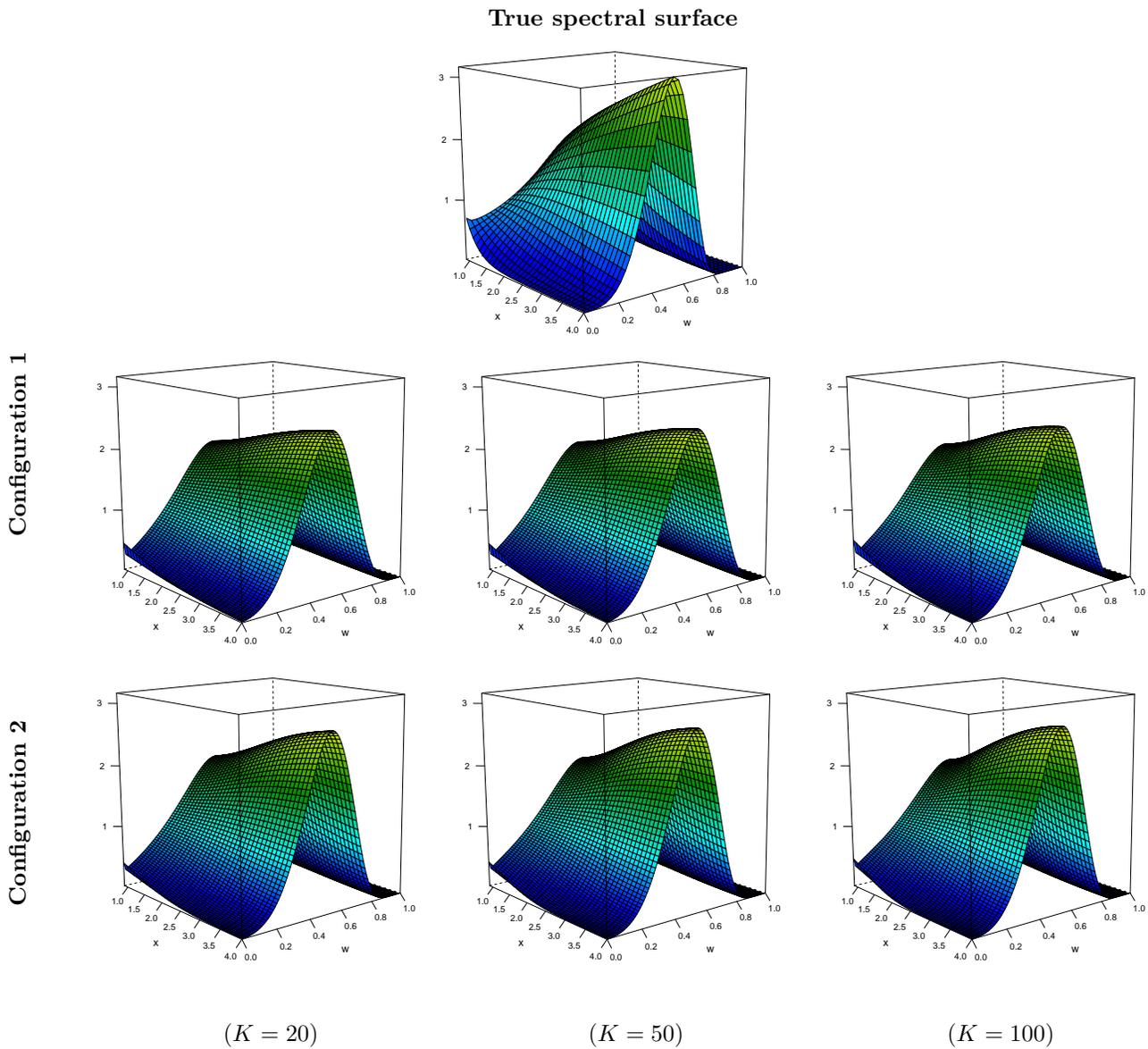


Figure 2-10. On the top: true spectral surface from the asymmetric Dirichlet predictor-dependent model detailed in Section 2.3.1, followed by Monte Carlo mean spectral surfaces for configurations 1 (left) and 2 (right) with  $K = 20$  (second row),  $K = 50$  (third row) and  $K = 100$  (fourth row).

## 2.6.2 Appendix A: Additional data application report

Locations of the monitoring stations, number of pseudo angles in each site and altitude can be found in Table 2.2.

Table 2.2. Locations of monitoring stations, along with the corresponding number of pseudo-angles and altitude.

---

Location	$n_k$	Altitude (m)
Beatenberg	57	1500
Bettlachstock	54	1150
Celerina	53	1890
Chironico	45	1350
Isonne	45	1200
Jussy	62	500
Lausanne	63	800
Nationalpark	59	1900
Neunkirch	41	600
Novaggio	65	950
Othmarsingen	57	490
Schänis	58	750
Visp	64	700
Vordemwald	61	480

---

# 3

## Time-Varying Extreme Value Dependence with Application to Leading European Stock Markets

---

### 3.1 Introduction

In recent years, international stock markets have been registering unprecedented levels of turbulence. Episodes such as the subprime crisis and the Greek debt crisis may have boosted this turbulence a little further, and led many to fear a financial doomsday. The situation has been extraordinarily delicate in Europe, where evidence of increasing extremal dependence was found by [Poon \*et al.\* \(2003, 2004\)](#) before the most recent financial crisis. We look to update suitable parts of their analysis and in particular analyze the time-varying extremal dependence in a more complete manner than has been done before. To achieve this goal, we propose an approach for modeling temporal nonstationarity in the extreme value dependence structure.

Statistical modeling of univariate extreme values has been in development since the 1970s ([Natural Environment Research Council, 1975](#)). Fundamental to practical applications to complex problems have been the development of methodology to account for nonstationarity in the distributions of interest, which was first strongly advocated by [Davison and Smith \(1990\)](#). Typical approaches to this problem are based around the generalized linear modeling idea of allowing the parameters of a marginal distribution to depend on covariates; more flexible approaches involving generalized additive modeling were introduced by [Chavez-Demoulin and Davison \(2005\)](#). [Eastoe and Tawn \(2009\)](#) present related ideas where data are pre-processed according to their dependence on covariates.

Statistical methods for modeling multivariate extreme values were introduced by [Tawn \(1988\)](#), and developed in [Tawn \(1990\)](#) and [Coles and Tawn \(1991\)](#). Since this time, much work has been done

on developing dependence modeling frameworks for extremes, yet surprisingly little has focused on how to incorporate nonstationarity into the dependence structure. Exceptions include [Eastoe \(2009\)](#), who introduces a conditionally independent hierarchical model, [Jonathan \*et al.\* \(2014\)](#), who develop methodology for including covariates in the model of [Heffernan and Tawn \(2004\)](#), and [de Carvalho and Davison \(2014\)](#), who develop a semiparametric model for settings where several multivariate extremal distributions are linked through the action of a covariate on an unspecified baseline distribution. In addition, [Huser and Genton \(2014\)](#) have also recently developed nonstationary models for spatial extremes. In this work we aim to augment the literature on modeling nonstationarity in the dependence structure by proposing flexible methodology for a simple set-up. Working within a tail dependence framework known as *asymptotic dependence*, we suppose that the relevant bivariate extreme value distribution evolves over a certain covariate of interest. The approach that we take is fully nonparametric, which is advantageous since neither the form of the bivariate distribution at a given covariate, nor the form of dependence on the covariate can be parametrically specified.

Our methodology is particularly suited to assessing temporal changes in extremal dependence, which is the situation that we would like to investigate in our motivating example. [Poon \*et al.\* \(2003, 2004\)](#) studied the dependence between stock market returns in the US, UK, France, Germany, and Japan. The main focus of their works was to highlight that not all markets exhibit a sufficient strength of tail dependence to be asymptotically dependent, and to propose alternative dependence summaries. However, considering only the European markets, they noted that there was evidence for relatively strong left tail dependence, and we also find evidence for asymptotic dependence in the left tails of these major European markets. As noted by [Poon \*et al.\* \(2003\)](#), the dependence is not stationary in time, and a main focus of this work is to explore this nonstationarity using a full model for the time-varying dependence structure, rather than simply summary statistics.

In the next section we provide a background on dependence modeling for extreme values, and introduce our proposed framework for incorporating nonstationarity. In [Section 3.3](#) we introduce our estimation and inference methods; numerical illustrations follow in [Section 3.4](#). The focus of [Section 3.5](#) is on applying the proposed methods to returns from three major European stock markets—using CAC, DAX, and FTSE—to assess the evolution of their extremal dependence structure over time. We conclude in [Section 3.6](#).

## 3.2 Predictor-dependent modeling for bivariate extremes

### 3.2.1 Bivariate statistics of extremes

Let  $\{(Y_{i,1}, Y_{i,2})\}_{i=1}^N$  be a collection of independent and identically distributed (iid) random vectors with continuous marginal distributions  $F_{Y_1}$  and  $F_{Y_2}$ . We are concerned with assessments of the extremal dependence between the components of the vectors, and thus without loss of generality we shall suppose that they have standard Pareto margins, i.e.,  $P(Y_j > y) = y^{-1}$ , for  $y > 1$  and  $j = 1, 2$ . Let

$$(M_{1,N}, M_{2,N}) := \frac{1}{N} \left( \max_{1 \leq i \leq N} \{Y_{i,1}\}, \max_{1 \leq i \leq N} \{Y_{i,2}\} \right)$$

be the standardized vector of componentwise maxima. Then if

$$P(M_{1,N} \leq y_1, M_{2,N} \leq y_2) \rightarrow G(y_1, y_2), \quad \text{as } N \rightarrow \infty, \quad (3.1)$$

where  $G$  is a non-degenerate distribution function, then the limit  $G$  has the form

$$G(y_1, y_2) = \exp \left\{ -2 \int_{[0,1]} \max \left( \frac{w}{y_1}, \frac{1-w}{y_2} \right) H(dw) \right\}, \quad (3.2)$$

where,  $y_1, y_2 > 0$ . Here,  $G(y_1, y_2)$  is the so-called *bivariate extreme value distribution* and  $H$  is a probability measure—known as the *spectral measure*. A consequence of Pickands' (1981) representation theorem is that the spectral measure needs to obey the following marginal moment constraint

$$\int_{[0,1]} w H(dw) = 1/2. \quad (3.3)$$

See Coles (2001, Theorem 8.1). Let  $R = Y_1 + Y_2$  and  $W = Y_1/(Y_1 + Y_2)$ . de Haan and Resnick (1977) have shown that the convergence in (3.2) is equivalent to

$$P(W \in \cdot \mid R > \tau) \rightarrow H(\cdot), \quad \tau \rightarrow \infty. \quad (3.4)$$

In practice, convergence (3.4) is more often useful than (3.2) and tells us that when the ‘radius’  $R$  is large, the ‘pseudo-angles’  $W$  are approximately distributed according to  $H$ , and approximately independent of  $R$ . The distribution of mass of  $H$  on  $[0, 1]$  describes the extremal dependence structure of the random vector  $(Y_1, Y_2)$ . The extreme cases of this distribution are given by *asymptotic independence*, whereby all mass is placed at the vertices of  $[0, 1]$ , giving  $G(y_1, y_2) = \exp\{-(y_1^{-1} + y_2^{-1})\}$ , and by complete dependence, whereby all mass is placed at the center of the interval, yielding  $G(y_1, y_2) = \exp\{-\max(y_1^{-1}, y_2^{-1})\}$ . We refer to situations where  $H$  has mass away from the vertices as *asymptotic dependence* and this will be the framework of our modeling. Nevertheless, asymptotic independence is a relatively common situation in practice, and can be detected when  $R$  and  $W$  are not found to be independent for any values of  $R$ , with the mass of  $W$  moving closer to 0 and 1 as events become more extreme. In this situation, no continuous models for  $H$  will provide useful information on the extremal dependence structure. Finally, a standard assumption for statistical modeling is that  $H$  is absolutely continuous with *spectral density*  $h(w) = dH(w)/dw$ , and this will be our framework.

### 3.2.2 Predictor-dependent modeling framework

Formally,  $\mathcal{F} = \{F_x : x \in \mathcal{X}\}$  is a set of pd (predictor-dependent) probability measures if the  $F_x$  are probability measures, indexed by a predictor  $x \in \mathcal{X} \subseteq \mathbb{R}$ . Analogously, we define  $\mathcal{H} = \{H_x : x \in \mathcal{X}\}$  as a set of pd spectral measures on  $[0, 1]$  if the  $H_x$  are probability measures satisfying

$$\int_{[0,1]} w H_x(dw) = 1/2, \quad x \in \mathcal{X}. \quad (3.5)$$

If  $H_x(w) := H_x[0, w]$  is absolutely continuous, its pd spectral density is  $h_x(w) = dH_x(w)/dw$ . Further aspects of pd spectral measures are discussed in [de Carvalho \(2015\)](#).

Our main modeling object of interest will be the set of pd spectral densities  $\{h_x(w) : w \in [0, 1], x \in \mathcal{X}\}$ , which we will refer to as the *spectral surface*. A simple spectral surface can be obtained with the pd spectral density  $h_x(w) = \beta(w; \mu_x, \mu_x)$ , where  $\mu : \mathcal{X} \mapsto (0, \infty)$ , and  $\beta(\cdot; p, q)$  denotes the beta density with shape parameters  $p, q > 0$ . In [Figure 3-1 \(a\)](#), we represent a spectral surface based on this model, with  $\mu_x = x$ , for  $x \in \mathcal{X} = [0.5, 50]$ . As can be seen, larger values of the predictor  $x$  lead to larger levels of extremal dependence. Other spectral surfaces can be readily constructed from parametric models

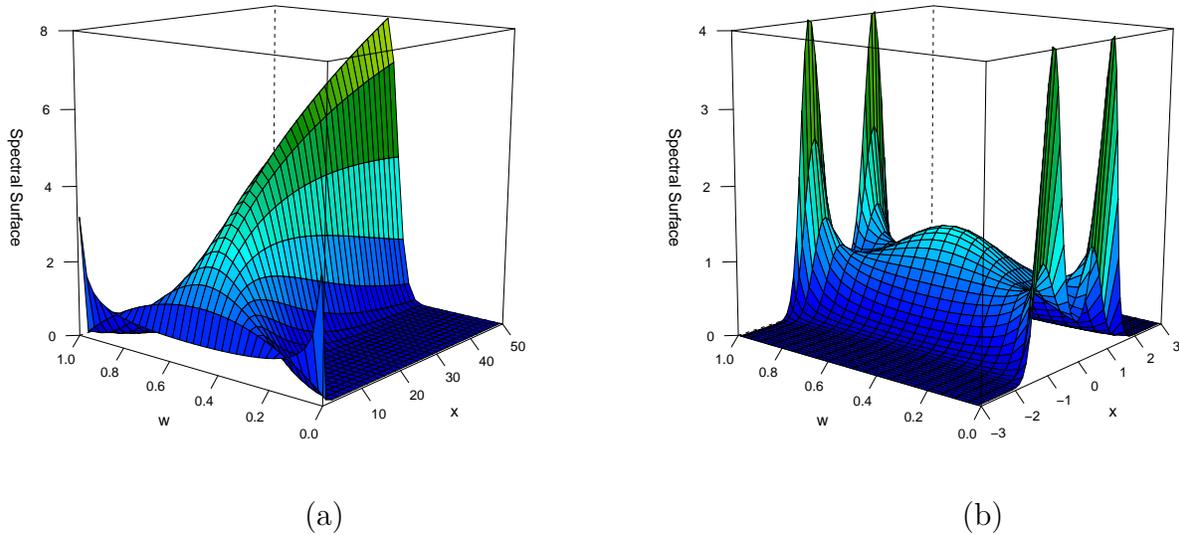


Figure 3-1. (a) Spectral surface from a predictor-dependent beta family, with  $\mu_x = x$ , for  $x \in \mathcal{X} = [0.5, 50]$ . (b) Spectral surface from a predictor-dependent logistic family, with  $\alpha_x = \Phi(x^2)$ , for  $x \in \mathcal{X} = [-3, 3]$ .

for the spectral density.

**Example 6** (pd logistic model). The logistic spectral surface is a covariate-adjusted extension of the logistic model (Coles, 2001, p. 146), and it is based on the pd spectral density

$$h_x(w) = \frac{1}{2} \left( \frac{1}{\alpha_x} - 1 \right) \{w(1-w)\}^{-1-1/\alpha_x} \{w^{-1/\alpha_x} + (1-w)^{-1/\alpha_x}\}^{\alpha_x-2}, \quad w \in (0, 1), \quad (3.6)$$

where  $\alpha : \mathcal{X} \mapsto (0, 1]$ . The closer  $\alpha_x$  is to 0, the higher the level of extremal dependence, while the closer  $\alpha_x$  is to 1, the closer we get to independence. Spectral surfaces with simple ‘shapes’ can be obtained by modeling  $\alpha_x$  with either a distribution function,  $F(x)$ , or a survivor function,  $1 - F(x)$ . If  $\alpha_x = F(x)$ , the level of extremal dependence decreases as the predictor increases, whereas if  $\alpha_x = 1 - F(x)$ , we get closer to independence as the value of the predictor increases; more sophisticated shapes can be obtained with  $\alpha_x = (F \circ G)(x)$ , for a certain continuous function  $G : \mathcal{X} \mapsto \mathbb{R}$ . In Figure 3-1 (b) we represent the logistic spectral surface in (3.6) with  $\alpha_x = \Phi(x^2)$ , for  $x \in \mathcal{X} = [-3, 3]$ , where  $\Phi$  denotes the standard normal distribution function.

**Example 7** (pd Dirichlet model). The Dirichlet spectral surface is a covariate-adjusted extension of the Dirichlet model (Coles and Tawn, 1991), and it is based on the pd spectral density

$$h_x(w) = \frac{a_x b_x \Gamma(a_x + b_x + 1) (a_x w)^{a_x - 1} \{b_x(1 - w)\}^{b_x - 1}}{2\Gamma(a_x)\Gamma(b_x)\{a_x w + b_x(1 - w)\}^{a_x + b_x + 1}}, \quad w \in (0, 1), \quad (3.7)$$

where  $a : \mathcal{X} \mapsto (0, \infty)$  and  $b : \mathcal{X} \mapsto (0, \infty)$ . Spectral surfaces with simple shapes can be obtained with  $a_x = b_x = \exp(x)$ , while if more complex dynamics are desirable, it can be based on  $a_x = \exp\{\mathcal{A}(x)\}$ ,  $b_x = \exp\{\mathcal{B}(x)\}$ , where  $\mathcal{A} : \mathcal{X} \mapsto \mathbb{R}$  and  $\mathcal{B} : \mathcal{X} \mapsto \mathbb{R}$  are continuous functions.

The basic idea of a pd spectral measure is not especially complicated, and inference for such a model would be simple if: (i) we knew our data conformed to a particular parametric family, and (ii) we knew precisely how that family depended on  $x$ . However, since we do not have knowledge of either of these aspects, the natural approach to take is a nonparametric one. We describe our estimation strategy in Section 3.3.

### 3.2.3 Related predictor-dependent objects of interest

Our estimation target  $\{h_x(w) : w \in [0, 1], x \in \mathcal{X}\}$  can be used for constructing other objects of interest when modeling bivariate extremes. For example, a pd version of Pickands (1981) dependence function can be defined as

$$A_x(w) = 1 - w + 2 \int_0^w H_x(u) du, \quad x \in \mathcal{X}, \quad w \in [0, 1],$$

leading to the pd extremal coefficient  $C_x = 2A_x(1/2)$ . In addition, we can define the pd bivariate extreme value (BEV) distribution as

$$G_x(y_1, y_2) = \exp \left\{ -2 \int_{[0,1]} \max \left( \frac{w}{y_1}, \frac{1-w}{y_2} \right) H_x(dw) \right\}, \quad (3.8)$$

for  $x \in \mathcal{X}$ , and  $y_1, y_2 > 0$ .

**Example 8.** Using the pd spectral density from Example 6, we obtain  $A_x(w) = \{(1-w)^{1/\alpha_x} + w^{1/\alpha_x}\}^{\alpha_x}$

and  $C_x = 2^{\alpha_x}$ , while the logistic BEV spectral surface is based on the pd BEV,

$$G_x(y_1, y_2) = \exp\{-(y_1^{-1/\alpha_x} + y_2^{-1/\alpha_x})^{\alpha_x}\}, \quad x \in \mathcal{X}, \quad y_1, y_2 > 0.$$

### 3.3 Estimation and inference

#### 3.3.1 Derivation of pseudo-angles

Consider Equation (3.4). We are now supposing nonstationarity in the dependence structure such that

$$\mathbb{P}(W_x \in \cdot \mid R_x > \tau) \rightarrow H_x(\cdot), \quad \tau \rightarrow \infty, \quad (3.9)$$

where we subscript the  $W$  and  $R$  by  $x$  to denote the fact that the dependence structure depends on  $x$ . Note that we still assume Pareto margins, and this still yields a limiting Pareto distribution for the radius  $R_x$  for each  $x$ . Typically, when stationarity is assumed, one searches for a high threshold in  $R$ , such that  $W$  and  $R$  are approximately independent above the threshold, and uses all  $W$  associated to threshold exceedances of  $R$  for inference. Supposing that  $x$  does not impact upon the rate of convergence in the limit (3.9), we can do the same thing. However, for prudence, we assess the dependence of  $R_x$  on  $x$ , and if any relationship is detected, then we threshold  $R_x$  using parametric quantile regression (Koenker, 2005), and take the corresponding  $W_x$  for inference. Below we use  $n$  to denote the number of pseudo-angles that resulted from thresholding  $R_{x_i} = Y_{x_i,1} + Y_{x_i,2}$ , for  $i = 1, \dots, N$ . Further details on the derivation of pseudo-angles for our data application can be found in Section 3.5.3.

We note that we are not allowing the margins to change over the predictor. This is however a sensible modeling assumption for our data application, because returns are known to be approximately stationary. Indeed, as posed by Resnick (2007, p. 7) “Returns have more attractive statistical properties than prices such as stationarity.”

### 3.3.2 Predictor-dependent spectral density estimation

Here we outline our estimator for the family of densities  $\{h_x(w) : w \in [0, 1], x \in \mathcal{X}\}$ . Assume observations  $\{W_{x_i}\}_{i=1}^n$ , where the covariates  $x_i$  are either continuous or discrete; for ease of notation, below we write  $W_i$  to denote  $W_{x_i}$ . Let  $K_b(x) = (1/b)K(x/b)$  be a kernel with bandwidth  $b > 0$ . For any  $x \in \mathcal{X}$ , we define the estimator

$$\hat{h}_x(w) = \sum_{i=1}^n \pi_{b,i}(x) \beta(w; \nu W_i \theta_b(x) + \tau, \nu \{1 - W_i \theta_b(x)\} + \tau), \quad w \in (0, 1), \quad (3.10)$$

where

$$\theta_b(x) = \frac{1/2}{\sum_{i=1}^n \pi_{b,i}(x) W_i}, \quad \pi_{b,i}(x) = \frac{K_b(x - x_i)}{\sum_{j=1}^n K_b(x - x_j)}, \quad i = 1, \dots, n.$$

The moment constraint (3.5) is satisfied, since

$$\int_0^1 w \hat{h}_x(w) dw = \frac{\sum_{i=1}^n K_b(x - x_i) \{\nu W_i \theta_b(x) + \tau\}}{(\nu + 2\tau) \sum_{i=1}^n K_b(x - x_i)} = \frac{\nu/2 + \tau}{\nu + 2\tau} = 1/2,$$

for all valid  $\tau \geq 0$ , upon substitution of  $\theta_b(x)$ .

The two kernels ( $K_b$  and  $\beta$ ) and the three parameters involved in our estimator can be interpreted as follows. The bandwidth  $b$  is the scale parameter of the kernel  $K_b$  and controls the amount of smoothing in the  $x$ -direction. The choice of the kernel  $K_b$  is subject to the typical considerations. In principle,  $K_b$  should be symmetric and unimodal, since there is a sense in which density estimators based on kernels that do not satisfy these requirements are inadmissible (Cline, 1988). While there are many kernel functions that do satisfy these basic requirements, it is well known that the choice of the kernel has little impact on the corresponding estimators; see Wand and Jones (1994, Ch. 2) and references therein. The parameter  $\nu$  is asymptotically inversely proportional to the variance of the kernel  $\beta$  and has the main role of controlling the amount of smoothing in the  $w$ -direction. The additional parameter  $\tau$  has the role of adjusting slightly the center of the kernel, allowing more flexible estimation, whilst not affecting the moment constraint. Note that  $\tau = 0$  yields a kernel with mean equal to  $W_i$ , whilst  $\tau = 1$  yields a kernel with mode  $W_i$ . In addition,  $\theta_b(x)$  assesses by how much the pseudo-angles

deviate from  $1/2$ . To see this, note that  $\theta_b(x) = (1/2)/\widehat{E}(W_i)$ , where  $\widehat{E}(W_i) = \sum_{i=1}^n \pi_{b,i}(x)W_i$  is the Nadaraya–Watson estimator (Nadaraya, 1964; Watson, 1964) of  $E(W_x) = \int_{[0,1]} wH_x(dw) = 1/2$ , for all  $x \in \mathcal{X}$ .

Plug-in estimators for the related pd objects of interest discussed in Section 3.2.3 can be readily obtained; particularly

$$\widehat{H}_x(w) = \sum_{i=1}^n \pi_{b,i}(x)B(w; \nu W_i \theta_b(x) + \tau, \nu \{1 - W_i \theta_b(x)\} + \tau), \quad w \in (0, 1),$$

where  $B(w; p, q)$  is the regularized incomplete beta function, with  $p, q > 0$ ; in addition, the plug-in estimators for the pd Pickands dependence function, extremal coefficient, and bivariate extreme value distribution can be written as

$$\begin{aligned} \widehat{A}_x(w) &= 1 - w + 2 \sum_{i=1}^n \pi_{b,i}(x) \int_0^w B(u; \nu W_i \theta_b(x) + \tau, \nu \{1 - W_i \theta_b(x)\} + \tau) du, \\ \widehat{C}_x &= 2\widehat{A}_x(1/2) = 1 + 4 \sum_{i=1}^n \pi_{b,i}(x) \int_0^{1/2} B(u; \nu W_i \theta_b(x) + \tau, \nu \{1 - W_i \theta_b(x)\} + \tau) du, \\ \widehat{G}_x(y_1, y_2) &= \exp \left\{ -2 \int_0^1 \max \left( \frac{u}{y_1}, \frac{1-u}{y_2} \right) \right. \\ &\quad \left. \times \sum_{i=1}^n \pi_{b,i}(x) \beta(u; \nu W_i \theta_b(x) + \tau, \nu \{1 - W_i \theta_b(x)\} + \tau) du \right\}, \end{aligned} \tag{3.11}$$

for  $x \in \mathcal{X}$ , and  $y_1, y_2 > 0$ .

### 3.3.3 Connections to smoothing on the unit interval

Kernel density estimation on the unit interval is a challenging problem; see Chen (1999), Jones and Henderson (2007), de Carvalho *et al.* (2013), Geenens (2014), and the references therein. In this section we comment on the connections of our estimator in (3.10), with the smooth Euclidean likelihood spectral density of de Carvalho *et al.* (2013), which can be regarded as a moment constrained kernel density estimator on the unit interval, in the sense that it obeys (3.3).

If all covariates  $x$  take the same value, so that the estimation problem reduces to estimating the

spectral density for an identically-distributed set of pseudo-angles  $\{W_i\}_{i=1}^n$ , then (3.10) becomes

$$\hat{h}(w) = \frac{1}{n} \sum_{i=1}^n \beta \left( w; \nu \frac{W_i}{2\bar{W}} + \tau, \nu \left\{ 1 - \frac{W_i}{2\bar{W}} \right\} + \tau \right), \quad w \in (0, 1). \quad (3.12)$$

Hence, our estimator in (3.12) has connections with the smooth Euclidean spectral density estimator in de Carvalho *et al.* (2013, p. 1190), which is given by

$$\begin{aligned} \tilde{h}(w) &= \frac{1}{n} \sum_{i=1}^n \{1 - (\bar{W} - 1/2)S^{-2}(W_i - \bar{W})\} \beta\{w; W_i\nu, (1 - W_i)\nu\}, \\ &= \frac{1}{n} \sum_{i=1}^n \beta\{w; W_i\nu, (1 - W_i)\nu\} - \frac{1}{n} \sum_{i=1}^n (\bar{W} - 1/2)S^{-2}(W_i - \bar{W})\} \beta\{w; W_i\nu, (1 - W_i)\nu\}, \end{aligned} \quad (3.13)$$

for  $w \in (0, 1)$ ; here  $\bar{W}$  and  $S^2$  are the sample mean and sample variance of  $W_1, \dots, W_n$ , that is,

$$\bar{W} = \frac{1}{n} \sum_{i=1}^n W_i, \quad S^2 = \frac{1}{n} \sum_{i=1}^n (W_i - \bar{W})^2.$$

A heuristic argument can be used to see the connection between (3.12) and (3.13), where for simplicity we focus on the case  $\tau = 0$ . The right-hand term in (3.13) is asymptotically negligible because  $\bar{W} = 1/2 + o_p(1)$ , so that for large  $n$ , we have  $\tilde{h}(w) \approx n^{-1} \sum_{i=1}^n \beta\{w; W_i\nu, (1 - W_i)\nu\}$ ; on the other hand, we also have that for large  $n$ ,  $\hat{h}(w) \approx n^{-1} \sum_{i=1}^n \beta\{w; W_i\nu, (1 - W_i)\nu\}$ , since  $\bar{W} = 1/2 + o_p(1)$ , as  $n \rightarrow \infty$ . While both estimators obey the moment constraint (3.5), they impose it through different approaches: our estimator enforces (3.3) by rescaling the pseudo-angles with a factor of  $(2\bar{W})^{-1}$ ; the smooth Euclidean spectral density enforces (3.3) additively, through the right-hand term in (3.13).

### 3.3.4 Tuning parameter selection and bootstrap

We select the tuning parameters by  $K$ -fold cross-validation (Hastie *et al.*, 2001, Section 7.10.1) using the Kullback–Leibler criterion (Bowman, 1984). Specifically, let  $\{\mathbf{W}_1, \dots, \mathbf{W}_K\}$  be the full sample of pseudo-angles split into  $K$  blocks. In the analyses in Sections 3.4 and 3.5, we split the blocks according the values of the accompanying covariate  $x$ , so that each  $\mathbf{W}_k = (W_{k,1}, \dots, W_{k,n_k})$  is in a similar part of the covariate space. Letting  $\hat{h}_{x(-k)}$  denote the estimator leaving out the  $k$ th sample,  $\mathbf{W}_k$ , of length

$n_k$ , we select

$$(\hat{b}, \hat{\nu}, \hat{\tau}) = \arg \min_{(b, \nu, \tau) \in \mathcal{R}_{\mathcal{X}, n}} \sum_{k=1}^K \sum_{j=1}^{n_k} -\log \hat{h}_{x_{k,j}}^{(-k)}(W_{k,j}),$$

with

$$\begin{aligned} \mathcal{R}_{\mathcal{X}, n} &= \{(b, \nu, \tau) \in (0, \infty)^3 : \nu W_i \theta_b(x) + \tau > 0, \nu \{1 - W_i \theta_b(x)\} + \tau > 0, \text{ for } i = 1, \dots, n; x \in \mathcal{X}\} \\ &= \{(b, \nu, \tau) \in (0, \infty)^3 : \nu \{1 - W_i \theta_b(x)\} + \tau > 0, \text{ for } i = 1, \dots, n; x \in \mathcal{X}\}. \end{aligned} \quad (3.14)$$

The constrained optimization yields well-defined estimates, since it guarantees the positivity of the beta parameters in our estimator. The latter equality in (3.14) follows from noting that  $\nu W_i \theta_b(x) + \tau > 0$ , for all  $x \in \mathcal{X}$ ; further details on practical implementation of tuning parameter selection are given in Section 3.4.2.

Cross sections of the spectral surface give rise to spectral densities for certain values of the covariate. To measure the precision of the estimates, approximate confidence bands can be constructed through a resampling cases bootstrap algorithm (Davison and Hinkley, 1997, Section 6.2.4), and we will adopt this approach for constructing approximate confidence bands in Section 3.5. Specifically our resampling cases bootstrap algorithm is as follows:

### Resampling cases bootstrap

for  $r = 1, \dots, B$  do:

1. Sample  $j_1^*, \dots, j_n^*$  with replacement from a discrete uniform distribution over  $\{1, \dots, n\}$ .
2. for  $i = 1, \dots, n$  do: set  $(x_i^*, W_i^*) = (x_{j_i^*}, W_{j_i^*})$ .
3. Estimate tuning parameters  $(\hat{b}^*, \hat{\nu}^*, \hat{\tau}^*)$ .

4. Estimate pd spectral density through (3.10), i.e., compute

$$\hat{h}_x^r(w) = \sum_{i=1}^n \pi_{\hat{b}^*,i}^*(x) \beta(w; \hat{\nu}^* W_i^* \theta_{\hat{b}^*}^*(x) + \hat{\tau}^*, \hat{\nu}^* \{1 - W_i^* \theta_{\hat{b}^*}^*(x)\} + \hat{\tau}^*), \quad w \in [0, 1],$$

where

$$\theta_{\hat{b}^*}^*(x) = \frac{1/2}{\sum_{i=1}^n \pi_{\hat{b}^*,i}^*(x) W_i^*}, \quad \pi_{\hat{b}^*,i}^*(x) = \frac{K_{\hat{b}^*}(x - x_i^*)}{\sum_{j=1}^n K_{\hat{b}^*}(x - x_j^*)}, \quad i = 1, \dots, n.$$

Such confidence bands are indeed approximate as they do not reflect the uncertainty stemming from transforming the data to the same margins.

## 3.4 Simulation study

### 3.4.1 Data-generating configurations and preliminary experiments

We study the performance of our methods under the logistic and Dirichlet pd models introduced in Examples 6 and 7. Regarding the logistic spectral surface, we take  $\alpha_x = \Phi(x)$  and consider  $x \in \mathcal{X}_{\text{logistic}} = [\Phi^{-1}(0.2), \Phi^{-1}(0.4)]$ . For the Dirichlet spectral surface we consider two scenarios: a symmetric Dirichlet spectral surface with  $(a_x, b_x) = (x, x)$ , for  $x \in \mathcal{X}_{\text{sDir}} = [0.8, 4]$  and an asymmetric Dirichlet spectral surface with  $(a_x, b_x) = (x, 100)$ , for  $x \in \mathcal{X}_{\text{aDir}} = [0.5, 2]$ . In Figure 3-2 we plot the true and estimated spectral surfaces for the three cases described above on a single experiment with  $n = 500$ . The top panel of Figure 3-2 corresponds to the logistic spectral surface, where extremal dependence decreases as a function of the predictor. The center panel shows the symmetric Dirichlet spectral surface, where we observe weaker dependence for lower values of the covariate, whereas stronger dependence prevails for higher values. Finally, an increasing asymmetric dependence dynamic is displayed in the bottom panel, where we have plotted the asymmetric Dirichlet spectral surface.

The single run experiment in Figure 3-2 allows us to illustrate strengths and limitations of the methods. Even though there is a good fit—which is discussed in further detail in Section 3.4.2—we can anticipate from this figure that our estimator suffers from limitations inherent to kernel-based

estimators. For example, pointwise estimation for the pd logistic model does not perform so well when the spectral surface peaks at  $x = \Phi^{-1}(0.2)$ , but this is mostly due to the boundary bias of  $K_b$  which is a drawback of kernel-based estimators on bounded domains; see [Hardle \(1990, Section 4.4\)](#) and references therein. For the pd Dirichlet models, the estimates are slightly less smooth due to the need to recover more complicated surfaces. In particular, when the true spectral density is low, and thus there are few data points, estimation will be naturally more challenging. In spite of these limitations, our estimator successfully recovers the shape of the true spectral surface, and thus is able to reproduce accurately the evolution of extremal dependence over the covariate.

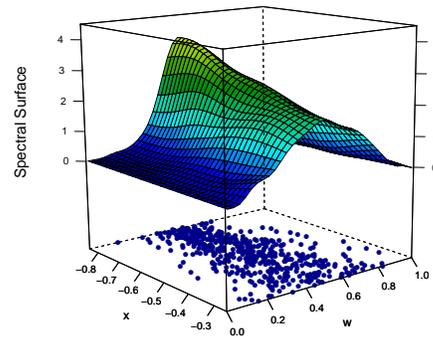
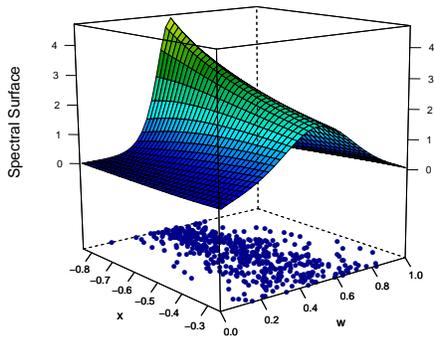
### 3.4.2 Simulation results

For the simulation studies, we took 1000 samples of sizes 350 and 500 for the three pd models presented in Section 3.4.1. For the samples of size 500, Figure 3-3 displays trajectories of 100 randomly selected estimates of cross sections of the spectral surface (i.e., pd spectral density estimates at fixed values of  $x$ ) along with their Monte Carlo means. These trajectories allow us to illustrate the performance of our estimator under different dependence dynamics. For instance, the top left and center right panels in Figure 3-3 show the limitations due to boundary bias that was discussed in Section 3.4.1. Indeed, our numerical experience suggests that the method's performance tends to be better in the center of the covariate space (center column of Figure 3-3), and that U-shaped spectral surfaces are in general more difficult to estimate. The center panel corresponds to the symmetric Dirichlet spectral model, which displays a good mean performance in the first two cases, but with some bias in the third case. Finally, the asymmetric Dirichlet spectral model, presented in the bottom panel, displays more dispersed estimates than for the other two models, although the Monte Carlo mean produces suitable approximations. The asymmetry does not seem to be a major problem. Overall, for the three models, the estimator displays reasonable performance in recovering the different shapes of the densities and Monte Carlo means produce reliable estimates.

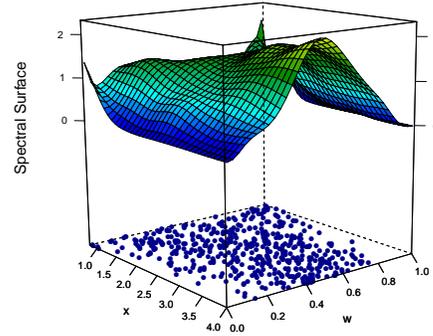
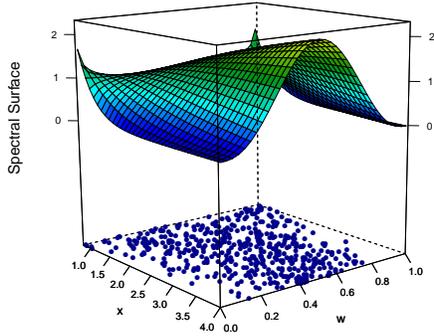
We assess the performance of our estimator using the mean integrated absolute error (MIAE),

$$\text{MIAE} = \mathbb{E} \left( \int_x \int_0^1 |\hat{h}_x(w) - h_x(w)| dw dx \right). \quad (3.15)$$

Logistic spectral surface



Symmetric Dirichlet spectral surface



Asymmetric Dirichlet spectral surface

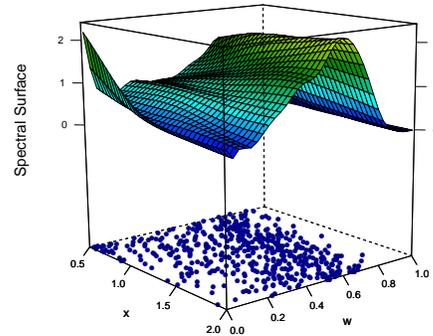
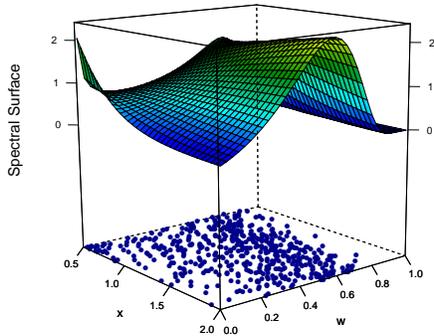


Figure 3-2. True spectral surfaces (left) and corresponding estimates (right). Top panel: pd logistic model with  $\alpha_x = \Phi(x)$ , for  $x \in \mathcal{X}_{\text{logistic}} = [\Phi^{-1}(0.2), \Phi^{-1}(0.4)]$ . Center panel: pd Symmetric Dirichlet model with  $(a_x, b_x) = (x, x)$ , for  $x \in \mathcal{X}_{\text{sDir}} = [0.8, 4]$ . Bottom panel: pd Asymmetric Dirichlet model with  $(a_x, b_x) = (x, 100)$ , for  $x \in \mathcal{X}_{\text{aDir}} \in [0.5, 2]$ . The simulated pseudo-angles based on which the estimates on the right are produced are overlaid on the bottom of the boxes.

Table 3.1. Mean integrated absolute error estimates computed over 1000 samples for the data-generating configurations discussed in Section 3.4.1.

$n$	pd Model	Specification	MIAE
350	Logistic	$\alpha_x = \Phi(x)$	0.0832
	Symmetric Dirichlet	$(a_x, b_x) = (x, x)$	0.4042
	Asymmetric Dirichlet	$(a_x, b_x) = (x, 100)$	0.2736
500	Logistic	$\alpha_x = \Phi(x)$	0.0750
	Symmetric Dirichlet	$(a_x, b_x) = (x, x)$	0.3664
	Asymmetric Dirichlet	$(a_x, b_x) = (x, 100)$	0.1879

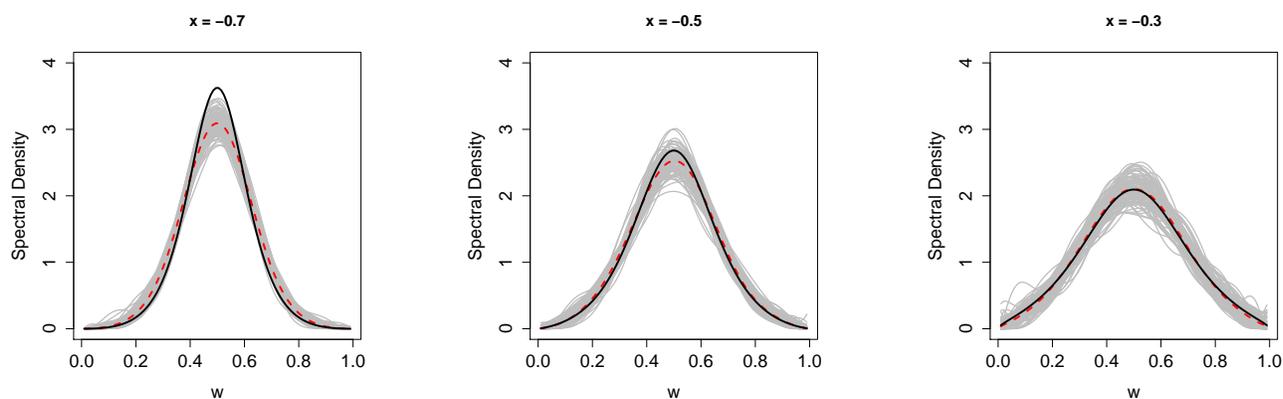
and report the results in Table 3.1. Although we cannot compare MIAE across models due to the different domains of the covariate, the figures demonstrate that performance increases with sample size as one would anticipate. Overall, simulations confirm that our methods produce acceptably accurate estimates of the spectral surface. Monte Carlo mean spectral surfaces are reported in the Appendix.

To conclude this section, we provide some practical implementation details on the tuning parameter selection. Since optimization over  $\mathcal{R}_{x,n}$  (defined in Eq. (3.14)) is computationally expensive, we found that  $\mathcal{R}_{x,n}$  can be well approximated by the set

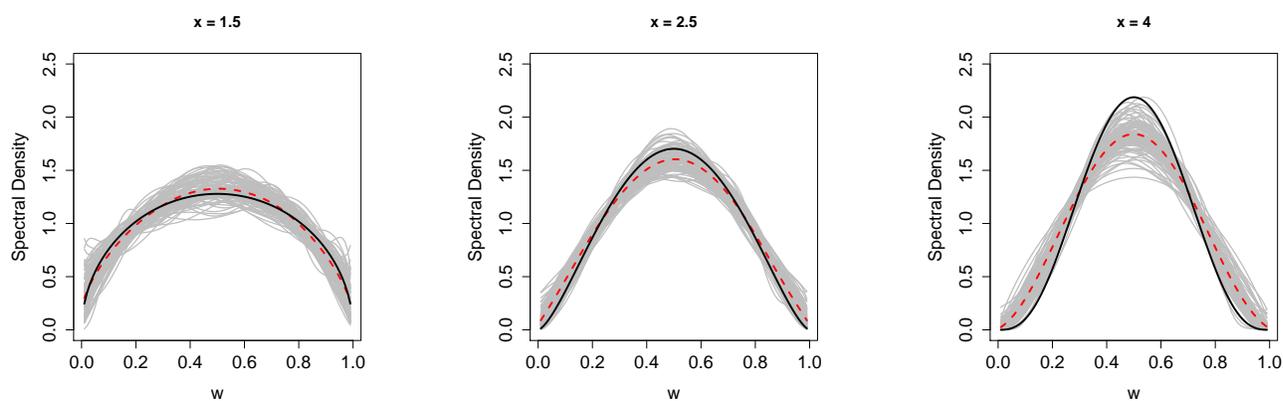
$$\mathcal{R}_n = \{(b, \nu, \tau) \in (0, \infty)^3 : \nu\{1 - W_i\theta_b(x_j)\} + \tau > 0, \text{ for } i, j = 1, \dots, n\},$$

i.e., a version determined only at the observed covariate values. Furthermore, for large  $n$ , unconstrained optimization over  $(0, \infty)^3$  typically also performs well. We thus recommend the user to initially try unconstrained optimization for large  $n$ , or optimization over  $\mathcal{R}_n$  for moderate  $n$ . Only if the resulting parameter values do not yield a valid estimator over the study region of interest, then one needs to implement the constrained optimization over  $\mathcal{R}_{x,n}$ .

## Cross sections of logistic spectral surface



## Cross sections of symmetric Dirichlet spectral surface



## Cross sections of asymmetric Dirichlet spectral surface

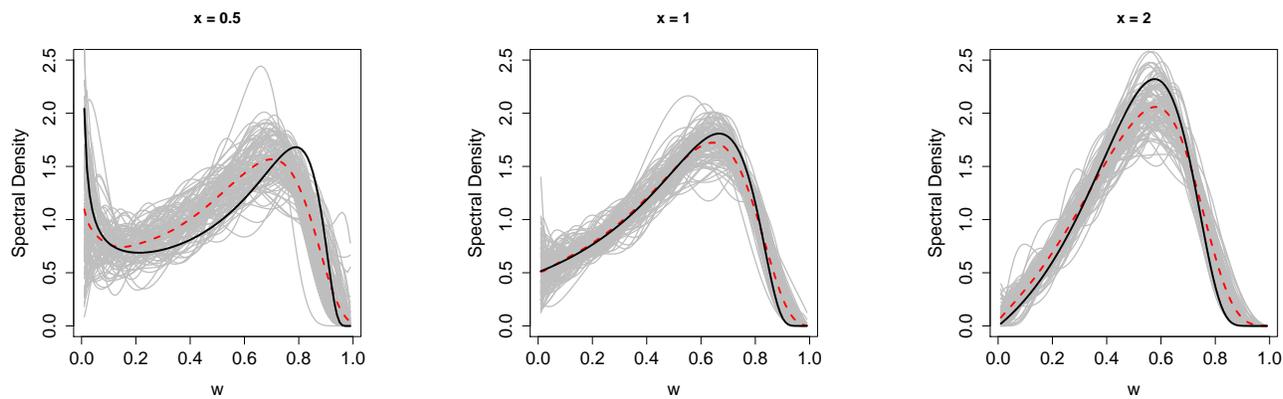


Figure 3-3. Trajectories of 100 randomly selected estimates of cross sections of the spectral surface (gray lines), as well as their corresponding true values (solid line) and Monte Carlo means (dashed line).

## 3.5 Dynamics of joint extremal losses in leading European stock markets

### 3.5.1 Background and motivation for empirical analysis

In 1999, eleven European Union (EU) countries formed the Economic and Monetary Union (EMU), which led them to adopt a common currency and monetary policy as well as the conduction of coordinated economic policies.

The process of creation of EMU was the outcome of three stages of development, on which further details can be found on the European Central Bank website:

<https://www.ecb.europa.eu/ecb/history;>

see also [James \(2012\)](#). To join the Eurozone (countries who adopted the Euro as their common currency) member states had to qualify by meeting the criteria of the Maastricht Treaty in terms of budget deficits, inflation, interest rates, and other monetary requirements. At the moment the Euro is the single currency shared by 19 of the 28 EU members. The remaining 9 countries, including the UK, are endowed with ‘opt-out’ clauses which exempts them from using the Euro as their currency. In recent years there have been several studies providing evidence for an increased integration of European stock markets, and the EMU has been frequently put forward as the causal driver for this increase, along with some other determinants ([Fratzscher, 2002](#); [Kim \*et al.\*, 2005](#); [Hardouvelis \*et al.\*, 2006](#); [Büttner and Hayo, 2011](#), and the references therein). [Hardouvelis \*et al.\* \(2006\)](#) found however that the UK, who has chosen not entering the eurozone, showed no increase in stock market integration by that time.

Although there is a wealth of studies analyzing stock market integration over time, few attempts have been made to ascertain the dynamics governing extreme value dependence of stock market returns over time. An exception in this respect is the seminal paper of [Poon \*et al.\* \(2003\)](#), which provides evidence of increasing levels of extremal dependence for three major stock markets within Europe [CAC (France), DAX (Germany), and FTSE (UK)]. The subperiod analysis of [Poon \*et al.\* \(2003, Section 3.3.2\)](#) is however exploratory, in the sense that they arbitrarily partitioned the sample period into three periods, and thus estimation of extremal dependence on each period only takes data from

that period into account.

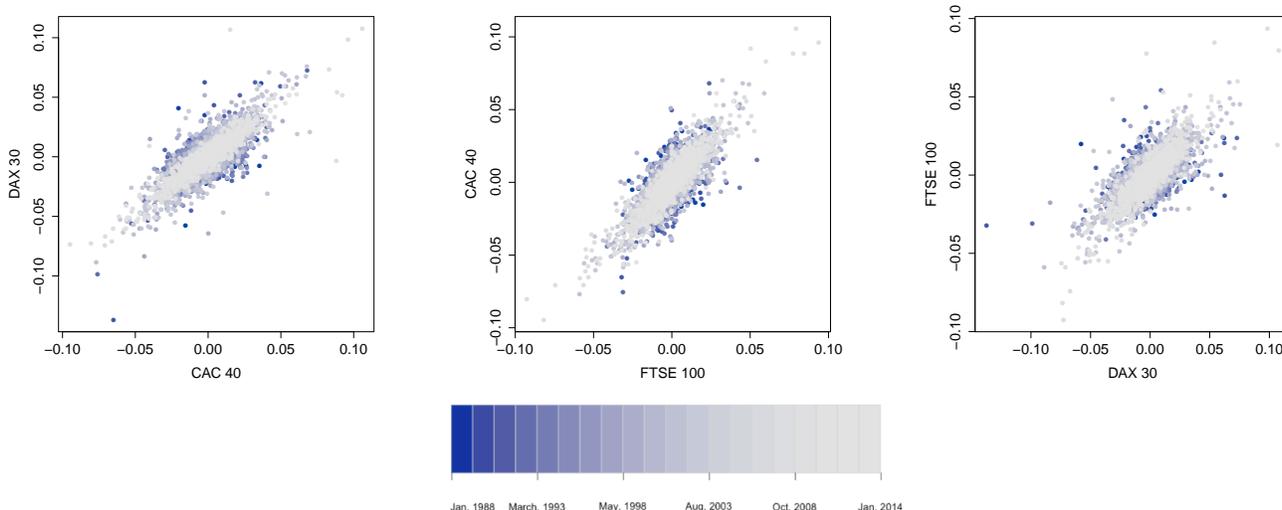


Figure 3-4. Scatterplots using a time-varying color palette for daily returns for CAC 40 (FR), DAX 30 (GR) and FTSE 100 (UK) spanning the period from January 1, 1988 to January 1, 2014.

Below we apply our methods to address a similar question to that of [Poon \*et al.\* \(2003, 2004\)](#). Specifically, one of our main interests is on disentangling the dynamics governing the dependence of extreme losses on three leading European stock markets—using CAC, DAX, and FTSE—in recent years. The motivation for choosing these markets is twofold: these are the stock markets of the European economies member of G5; these are also the same European stock markets considered by [Poon \*et al.\* \(2003, 2004\)](#). Moreover they display a stronger type of extremal dependence than some of the other markets studied by [Poon \*et al.\* \(2003, 2004\)](#), i.e., asymptotic dependence as defined in [Section 3.2.1](#).

### 3.5.2 Data description and exploratory considerations

Our data consist of daily closing stock index levels of three leading European stock markets: CAC 40, DAX 30, and FTSE 100 (henceforth CAC, DAX, and FTSE). Index values were gathered from Datastream in terms of local currency. Our sample period spans from January 1, 1988 to January 1, 2014 ( $N = 6784$  observations), and hence it includes the Great Moderation (from the mid-1980s to 2007) and Great Recession (December 2007 to June 2009) which are by all standards challenging modeling

issues. Scatterplots of possible combinations of pairs of return series are displayed in Figure 3-4. This figure is depicted using a time-varying color palette which allows us to uncover a higher frequency of joint extremes over recent periods, thus suggesting that extremal dependence between returns may have been increasing over time. This is in line with the findings of Poon *et al.* (2003, 2004). As mentioned in Section 3.5.1 we choose to focus on extreme losses, and thus as a unit of analysis we use daily negative returns, which can be used as proxies for losses in these markets.

To verify that our methods are sensible for modeling, we need to assess whether negative returns are asymptotically dependent. As mentioned in Section 3.2.1, in the modeling of extreme events two different classes of extreme value dependence can arise: asymptotic dependence and asymptotic independence. Dependence between moderately large values can arise in both cases, but the very largest values from each variable can occur together only under asymptotic dependence. To make ideas concrete, let  $Y_1$  and  $Y_2$  be any two negative returns of interest, transformed to have unit Pareto margins. Under an exploratory setting, two measures of tail dependence can be obtained to summarize the strength of extremal dependence:

$$\chi = \lim_{\tau \rightarrow \infty} \mathbf{P}(Y_1 > \tau \mid Y_2 > \tau), \quad \bar{\chi} = \lim_{\tau \rightarrow \infty} \frac{2 \log \mathbf{P}(Y_1 > \tau)}{\log \mathbf{P}(Y_1 > \tau, Y_2 > \tau)} - 1.$$

Here,  $\chi \in [0, 1]$  measures the strength of dependence within the class of asymptotically dependent variables, whereas  $\bar{\chi} \in [-1, 1]$  is often used to measure the strength of dependence within the class of asymptotically independent variables. Taken together, the pair  $(\chi, \bar{\chi})$  provides a summary of extremal dependence for the vector  $(Y_1, Y_2)$ . Asymptotic dependence implies  $\chi = 0$  and  $\bar{\chi} = 1$ . Roughly speaking, if  $\bar{\chi} > 0$  then we often speak about ‘positive extremal dependence,’ whereas if  $\bar{\chi} < 0$  we use the expression ‘negative extremal dependence.’ Indeed, for the bivariate normal dependence structure  $\bar{\chi}$  corresponds to Pearson correlation; see Heffernan (2000) for further examples.

In Figure 3-5 we present rolling window estimates of  $\chi$  and  $\bar{\chi}$  with approximate 95% confidence intervals, which is tantamount to the subperiod analysis of Poon *et al.* (2003, Section 3.3.2). Given the large uncertainty entailed in the estimation of  $\chi$  and  $\bar{\chi}$ , interpretation of these plots is far from being straightforward. Nevertheless, pointwise estimation for  $\chi$  seems reasonably different from 0 for the three pairs under study. Values for  $\bar{\chi}$  are closer to 1 as time passes, although for FTSE versus

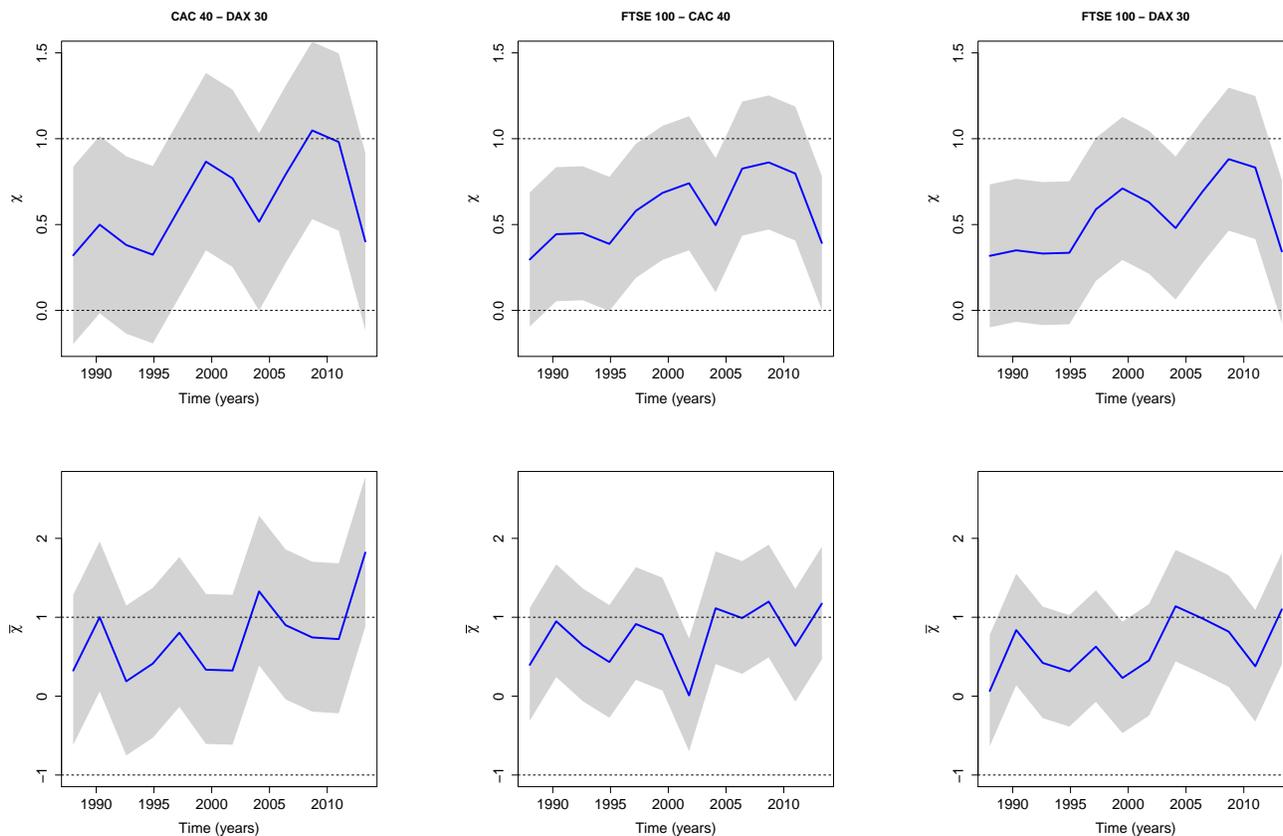


Figure 3-5. Rolling window estimates for  $\chi$  (top) and  $\bar{\chi}$  (bottom), using a window of size 600 for CAC 40 (FR), DAX 30 (GR), and FTSE 100 (UK).

CAC there are some evidences of  $\bar{\chi}$  being close to 0 at about 2003. However, the values of  $\bar{\chi}$  start to increase right after, quickly approaching 1.

### 3.5.3 Modeling time-varying extremal dependence

The time-varying color palette scatterplots in Figure 3-4 and the rolling window estimates in Figure 3-5 provide evidence of an increase in extremal dependence, but they are only exploratory. In this section we complete the analysis from Section 3.5.2, by applying our predictor-modeling approach in order to assess how the dependence structure of bivariate extreme losses in the CAC, DAX, and FTSE has been evolving over recent years. Before we proceed any further, some comments regarding implementation are necessary. As mentioned in Section 3.2, the data were transformed to standard Pareto margins. This was done as follows. Given a sample of pairs of returns  $(r_{1,1}, r_{1,2}), \dots, (r_{N,1}, r_{N,2})$ , we constructed

proxies for the unobservable pseudo-angles  $W_i$  by setting

$$W_i = \widehat{Y}_{i,1}/(\widehat{Y}_{i,1} + \widehat{Y}_{i,2}), \quad R_i = \widehat{Y}_{i,1} + \widehat{Y}_{i,2},$$

where  $\widehat{Y}_{i,1} = 1/\{1 - \widehat{F}_{r_1}(r_{i,1})\}$  and  $\widehat{Y}_{i,2} = 1/\{1 - \widehat{F}_{r_2}(r_{i,2})\}$  and where  $\widehat{F}_{r_1}$  and  $\widehat{F}_{r_2}$  are estimators of the marginal distribution functions  $F_{r_1}$  and  $F_{r_2}$ . A robust choice for  $\widehat{F}_{r_1}$  and  $\widehat{F}_{r_2}$  is the pair of univariate empirical distribution functions, normalized by  $N+1$  rather than by  $N$  to avoid division by zero. Finally, we proceed as described in Section 3.3.1. Evidence of dependence of the pseudo-radius on time was found, and so we proceed under a nonstationary assumption. Specifically, we model the 95% quantile of the pseudo-radius through parametric quantile regression, threshold the pseudo-radius according to the fit, and use the associated pseudo-angles for inference. After thresholding, the number of pseudo-angles is equal to 338, 340 and 339 for CAC–DAX, FTSE–CAC and FTSE–DAX, respectively. The pseudo-angles corresponding to these observations are plotted in the two-dimensional bottom plane in Figure 3-7. The tuning parameters  $(b, \nu, \tau)$  were computed as discussed in Sections 3.3.4 and 3.4.2 and they are estimated as:  $(0.12, 11.49, 0.62)$  for CAC–DAX;  $(0.21, 6.42, 0.57)$  for FTSE–CAC;  $(0.24, 11.03, 0.57)$  for FTSE–DAX.

In Figure 3-6 we plot cross sections of the spectral surface estimate at three important periods on the EU agenda: I) Beginning of stage one of EMU (1 July, 1990); II) beginning of stage three of EMU (1 January, 1999); III) activation of the assistance package for Greece (2 May, 2010), the first country to be shut out of the bond market, which fostered the European sovereign debt crisis (Lane, 2012). The choice of landmarks I–III) is arbitrary, but recall that our main interest is in describing how extremal dependence may change, by comparing periods sufficiently apart in time. As can be observed from the first column in Figure 3-6, at around 1990 the dependence between extreme losses of CAC–DAX and FTSE–DAX was similar, exhibiting some evidence of extremal independence. The second column in Figure 3-6 reveals that about a decade later this dynamic changed, and that extreme losses of CAC–DAX and FTSE–DAX started to show some mild signs of extremal dependence. The evidence for an increase in extremal dependence of losses in these pairs of markets is even more clear at the beginning of the European sovereign debt crisis, as can be confirmed from the third column in Figure 3-6. The dynamics of the dependence of bivariate extreme losses for the pair FTSE–CAC

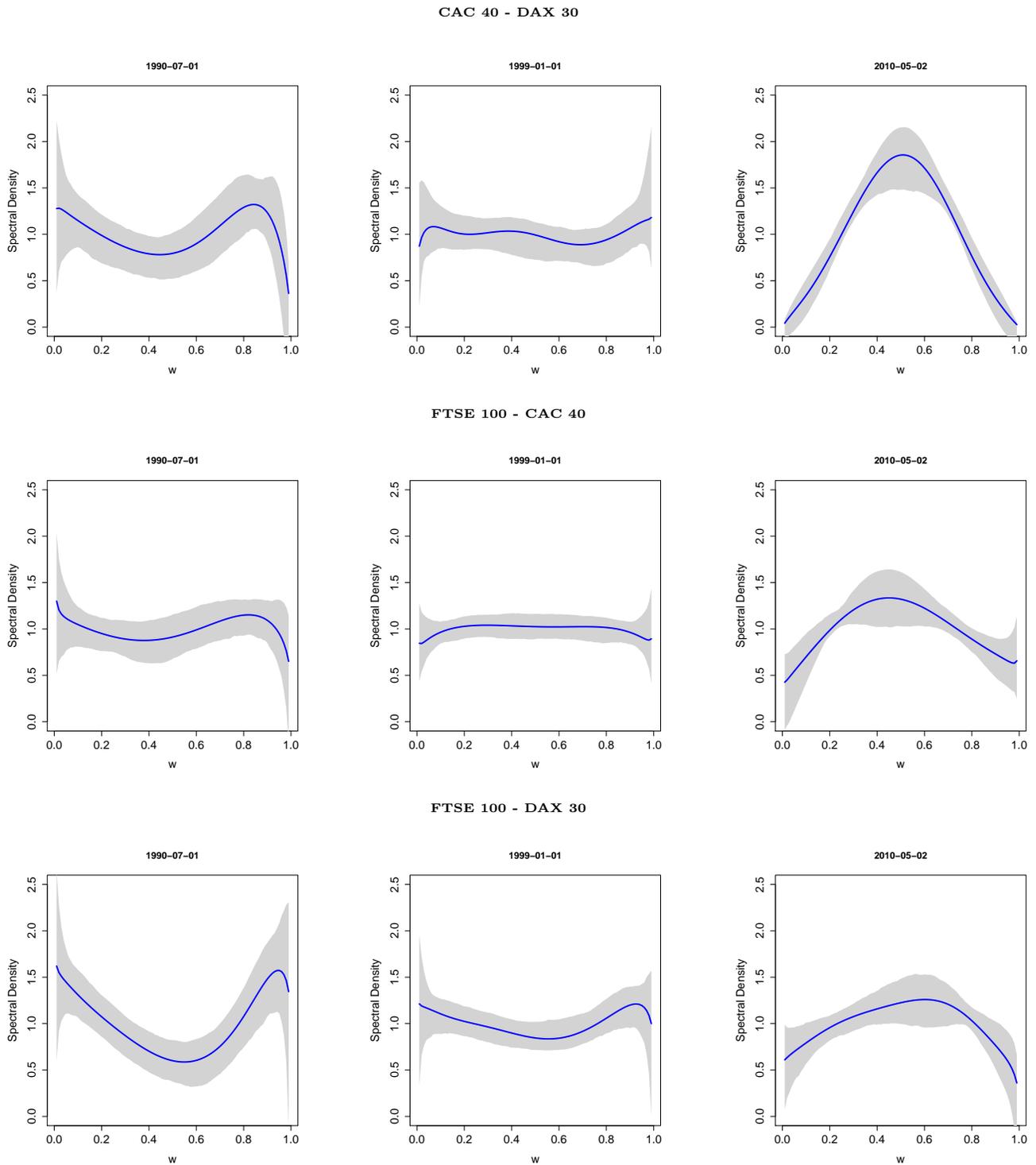


Figure 3-6. Cross sections of spectral surface estimates for CAC–DAX [top], FTSE–CAC [center], and FTSE–DAX [bottom]. The first column corresponds to the beginning of stage one of EMU (1 July, 1990), the second column corresponds to the beginning of stage three of EMU (1 January, 1999), and the third column corresponds to the time of activation of the assistance package for Greece (2 May, 2010). Confidence bands were obtained through a resampling cases bootstrap algorithm, with  $B = 1000$ .

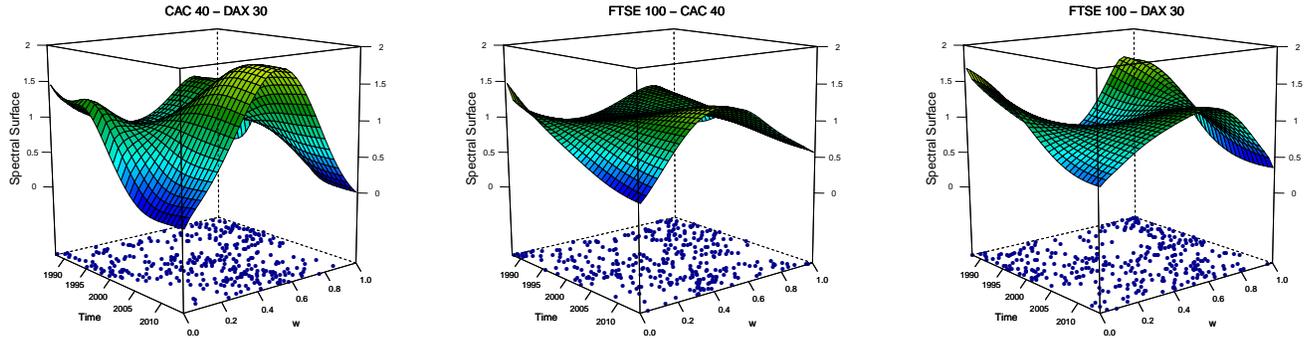


Figure 3-7. Spectral surfaces estimates for CAC–DAX, FTSE–CAC and FTSE–DAX, with pseudo-angles overlaid on the bottom of the box.

are slightly more complex, but also show evidence of increasing levels at the onset of the European sovereign debt crisis. Our findings may seem to contradict [Hardouvelis \*et al.\* \(2006\)](#)—who claimed that the UK showed no increase in stock market integration—however we note that [Hardouvelis \*et al.\* \(2006\)](#) did not assess extremal dependence. The confidence bands in Figure 3-6 were obtained through the resampling cases bootstrap algorithm, with  $B = 1000$  resamples, detailed in Section 3.3.4.

Figure 3-6 provides only a few snapshots corresponding to landmarks I–III). A more complete portrait of the temporal changes in extremal dependence is provided by the spectral surface estimate in Figure 3-7, from which the cross-sections in Figure 3-6 are derived.

All in all, we can clearly see the change from weaker dependence around 1990 to strong dependence starting from 2005, thus suggesting that in recent decades there has been an increase in the extremal dependence in the losses for these leading European stock markets. The pair CAC–DAX is the one where extremal dependence peaks the most, thus suggesting a high level of synchronization and comovement of extreme losses on those markets over recent years.

Similar conclusions can be drawn from Figure 3-8, where we plot the pd extremal coefficient as a function of time, as defined in Section 3.2.3. The extremal coefficient is equal to  $2 - \chi$ , and as such is equal to 2 under asymptotic independence, and takes values in  $[1, 2)$  under asymptotic dependence. Figure 3-8 permits the comparison with the results of [Poon \*et al.\* \(2004\)](#), who calculated  $\chi$  over subperiods. The red lines in Figure 3-8 represent the values from the analysis of Poon *et al.* for the subperiod November 1990–November 2001 (cf [Poon \*et al.\*, 2004](#), Table 3). Specifically [Poon \*et al.\*](#)

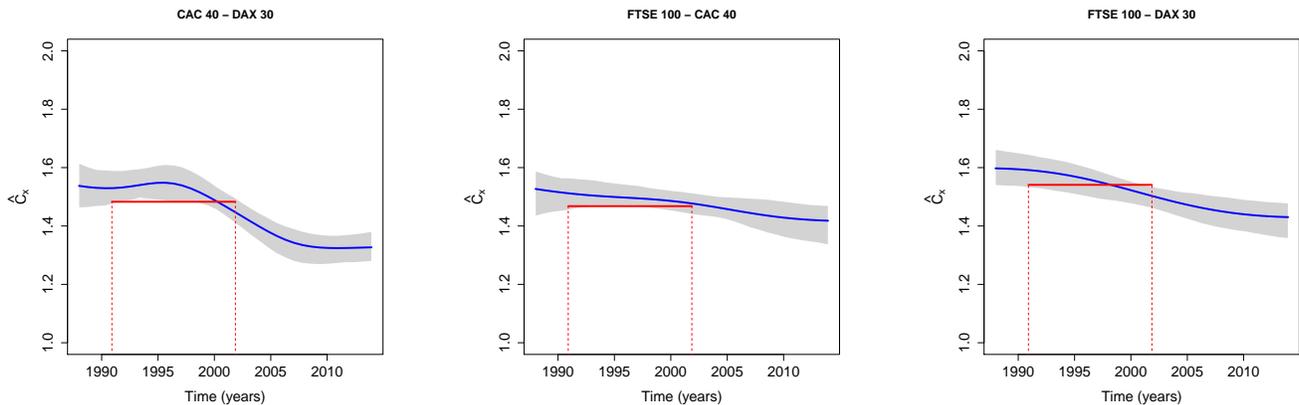


Figure 3-8. Predictor-dependent extremal coefficients. Confidence bands were obtained through a resampling cases bootstrap algorithm, with  $B = 1000$ . The red lines represent the values from the analysis of Poon et al. for the subperiod November 1990–November 2001 (cf Poon *et al.*, 2004, Table 3).

(2004) report the following values of  $\chi$  for: CAC–DAX, 0.517 (0.037); FTSE–CAC, 0.532 (0.035) and FTSE–DAX, 0.459 (0.039), with standard errors on parentheses. As can be seen from Figure 3-8, the magnitudes of the extremal coefficients estimated by Poon et al. are slightly smaller than the estimates we obtain, but are in reasonable agreement with uncertainty taken into account.

### 3.6 Final remarks

In this chapter, we develop methods for modeling nonstationary extremal dependence structures, motivated by the need to assess the comovement of extreme losses on some leading European stock markets over recent years. Although there are many studies analyzing stock market integration over time, few attempts have been made to assess the dynamics of extreme value dependence of stock market returns over time.

In a recent paper, de Carvalho and Davison (2014) develop the so-called spectral density ratio model, which can be regarded as a related approach to model nonstationary extremal dependence structures. While flexible, their approach is computationally demanding, and only applies to the setting where there are multiple observations for each value of the predictor—and thus it is inappropriate for the present setting of interest. Our methods are more resilient in the sense that they do not require a sample of pseudo-angles per each observed covariate, but apply more generally to a regression setting where

each covariate value may correspond to a single pseudo-angle. In common with any approach based on multivariate extreme value distributions, a limitation with our methods is that it will overestimate risk if data are asymptotically independent. The need for developing pd models able to cope with both asymptotic dependence and asymptotic independence is of utmost importance.

## 3.7 Appendix

### 3.7.1 Appendix A: tuning parameter selection

**Proposition 1.** Define

$$\begin{aligned}\mathcal{R}_{x,n} &= \{(b, \nu, \tau) \in (0, \infty)^3 : \nu\{1 - W_i\theta_b(x)\} + \tau > 0, \text{ for } i = 1, \dots, n, x \in \mathcal{X}\}, \\ \mathcal{R}_n &= \{(b, \nu, \tau) \in (0, \infty)^3 : \nu\{1 - W_i\theta_b(x_j)\} + \tau > 0, \text{ for } i, j = 1, \dots, n\}\end{aligned}$$

and suppose that  $b = b(n) \rightarrow 0$ ,  $\nu = \nu(n) \rightarrow \infty$  and  $\tau = \tau(n) \rightarrow 0$  when  $n \rightarrow \infty$ . Then, as  $n \rightarrow \infty$

$$\lambda(\mathcal{R}_{x,n} \setminus \mathcal{R}_n) = o_p(1), \quad \text{and} \quad \lambda((0, \infty)^3 \setminus \mathcal{R}_n) = o_p(1), \quad (3.16)$$

where  $\lambda(\cdot)$  denotes the Lebesgue measure in  $\mathbb{R}^3$ .

*Proof.* The proofs for both convergences in (3.16) are similar. Here we outline the proof for the first convergence.

$$\begin{aligned}\lambda(\mathcal{R}_{x,n} \setminus \mathcal{R}_n) &= \lambda(\{(b, \nu, \tau) \in (0, \infty)^3 : \nu\{1 - W_i\theta_b(x)\} + \tau > 0, \text{ for } i = 1, \dots, n, x \in \mathcal{X} \text{ and} \\ &\quad \nu\{1 - W_i\theta_b(x_j)\} + \tau \leq 0, \text{ for some } i, j = 1, \dots, n\})\end{aligned}$$

but according to lemma 2, when  $n \rightarrow \infty$ ,  $\theta_b(x) - 1 = o_p(1)$  and since  $\nu \rightarrow \infty$  and  $\tau \rightarrow 0$ , we have that

$$\lambda(\{(b, \nu, \tau) \in (0, \infty)^3 : \nu\{1 - W_i\theta_b(x_j)\} + \tau \leq 0, \text{ for some } i, j = 1, \dots, n\}) = o_p(1),$$

consequently,  $\lambda(\mathcal{R}_{x,n} \setminus \mathcal{R}_n) = o_p(1)$ . □

### 3.7.2 Appendix B: Monte Carlo mean spectral surfaces

Figure 3-9 shows Monte Carlo mean spectral surfaces for the three simulation scenarios described in Section 3.4.

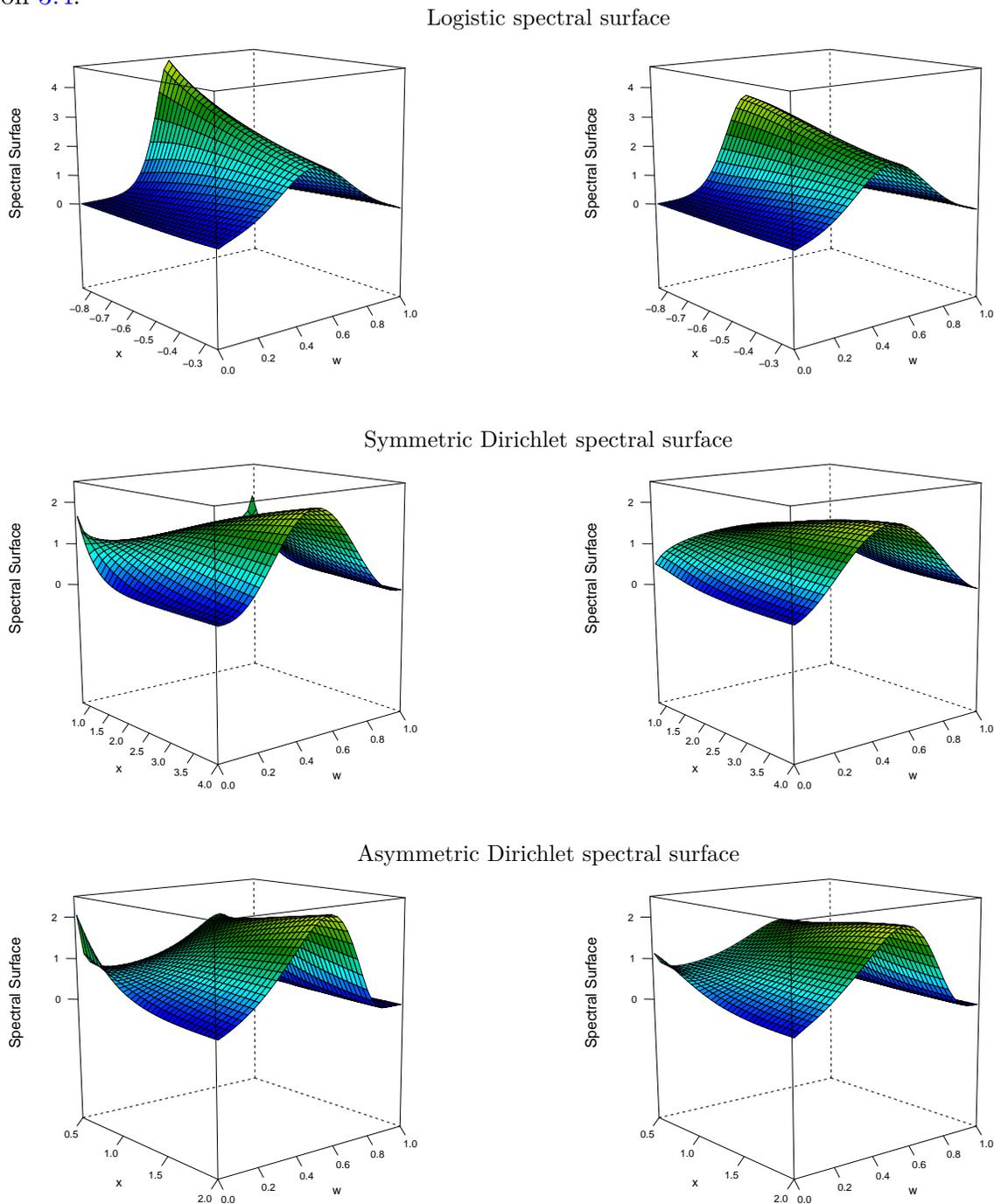


Figure 3-9. True spectral surfaces (left) and corresponding Monte Carlo means (right) resulting from the simulation study in Section 3.4.

### 3.7.3 Appendix D: supplementary data analysis reports

Recall that our data consist of daily closing stock index levels of three leading European stock markets: CAC 40, DAX 30, and FTSE 100. Index values were gathered from Datastream in terms of local currency. Our sample period spans from January 1, 1988 to January 1, 2014 (6784 observations).

Figure 3-10 shows the series of daily returns, whereas Table 3.2 presents summary statistics for CAC 40, DAX 30 and FTSE 100.

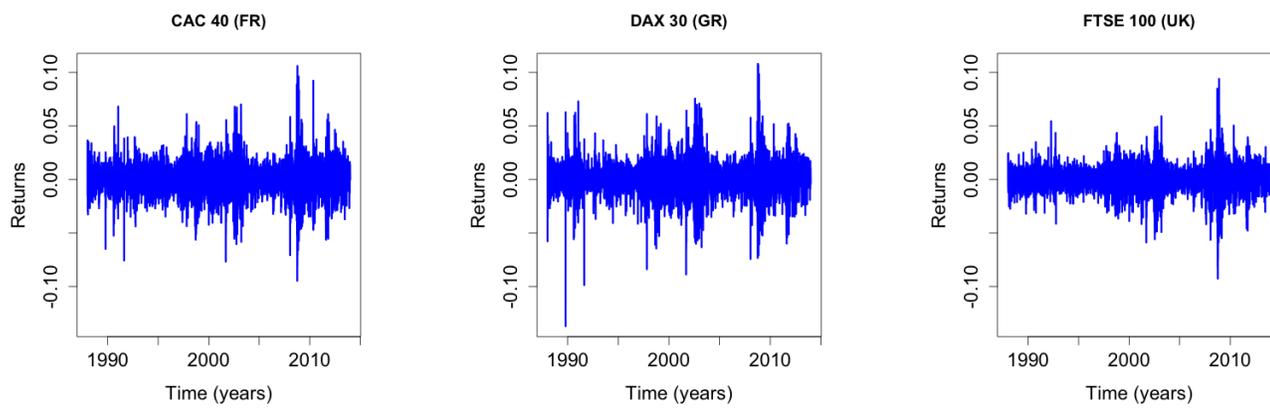


Figure 3-10. Daily returns for CAC 40 (FR), DAX 30 (GR) and FTSE 100(UK) spanning the period from January 1, 1988 to January 1, 2014.

Table 3.2. Summary statistics of the return series for the three stock markets indices: FTSE 100 (UK), CAC 40 (FR) and DAX 30 (GR). The data span the period from January 1, 1988 to January 1, 2014.

Statistic	Stock market index		
	UK	FR	GR
Mean	0	0	0
Std Deviation	0.011	0.014	0.014
Minimum	-0.093	-0.095	-0.137
Maximum	0.094	0.106	0.108
Kurtosis	9.415	7.833	9.272
Skweness	0.135	-0.047	-0.255



# 4

## Conclusions and Further Modeling

---

In this thesis, we have addressed two different frameworks in the context of nonstationary bivariate extremal dependence. In Chapter 2 we developed methodologies for a setting where different dependence structures are related with different values of a predictor, whereas in Chapter 3 we worked in a setting where the dependence is changing over the predictor. In both cases, we proposed methodologies to estimate the predictor-dependent spectral density that produce mean-constraint estimators. This chapter summarizes the main conclusions of this dissertation and gives directions of future work.

### 4.1 Conclusions

Bivariate statistics of extremes provides a natural framework for inference on the joint tail of a distribution outside the sample region. The characterization of the asymptotic distribution of the joint tail relies on the spectral measure defined in  $[0,1]$ , whose distribution or density describe the dependence among extremes. In this thesis, we have developed methodologies for modeling the spectral density of a bivariate extreme value distribution which evolves over a predictor. We have done this, working within a tail dependence framework known as asymptotic dependence, and assuming two different frameworks, both related with the notion of predictor-dependent spectral measure. The first framework, detailed in Chapter 2, considers the case where sets of extreme data are observed for every value of the predictor. In this setting, spectral densities estimates can be constructed for every predictor value through the smoothed Euclidean likelihood estimator (de Carvalho *et al.*, 2013). Using smoothing techniques, we are able to smooth over all the predictor space, producing estimates of the spectral densities for

unobserved predictor values. This two-step inference technique gives rise to a double kernel estimator, which is similar to the Nadaraya–Watson estimator, but where responses are spectral density estimates. Our model is related to the spectral density ratio model of [de Carvalho and Davison \(2014\)](#) in the sense that covariates can be incorporated, but rather than just linking extremal distributions, our model assesses directly the evolution of extremal dependence. Furthermore, implementation of our estimator is straightforward, and inference is computationally convenient. We test our methods in a temperature data application where altitude is considered a variable of interest. Results suggest an impact of the altitude on the extremal dependence of maximum temperatures under the forest and on a open field, showing the need to consider nonstationarity in the extreme value dependence structure.

The second framework of this dissertation, covered in [Chapter 3](#), considered a setting where the dependence is changing over the predictor and is particularly suited if we want to study temporal changes in extremal dependence. The observed data consist of paired values of the predictors and the pseudo-angles. This data setting produces methods that are more complex, since we can no longer divide the estimation and the parameter selection processes, as we did in [Chapter 2](#). Nevertheless, these methods are more resilient in the sense that they do not require a sample of pseudo-angles per each observed covariate value, but apply more generally to a regression setting where to each covariate value may correspond to a single pseudo-angle. Our work is motivated by the results of [Poon \*et al.\* \(2003\)](#) and [Poon \*et al.\* \(2004\)](#) who studied the dependence between stock market returns in the US, UK, France, Germany, and Japan. The main focus of their works was to highlight that not all markets exhibit a sufficient strength of tail dependence to be asymptotically dependent, and to propose alternative dependence summaries. However, considering only the European markets, they noted that there was evidence for relatively strong left tail dependence, and we also find evidence for asymptotic dependence in the left tails of these major European markets. As noted by [Poon \*et al.\* \(2003\)](#), the dependence is not stationary in time. Our main focus was to explore this nonstationarity using a full model for the time-varying dependence structure, rather than simply summary statistics. As in [Chapter 2](#), we characterize the dependence structure through the spectral density. Our model generates time-dependent spectral density estimates based on a double kernel estimator, which smooths simultaneously in both the predictor and the pseudo-angle directions.

We have made a first attempt to address the gap between the developments in nonstationary

marginal distributions and bivariate distributions, by proposing two computationally compelling regression models for the spectral density of a bivariate extreme value distribution that allows us to include nonstationarity in the dependence structure of extreme values. Although both models deal with nonstationary, they are built for different scenarios. Our first methodology applies to the setting where there are several pseudo-angles corresponding to the same value of the predictor, whereas the second one is particularly suited to assessing temporal changes in extremal dependence. Both models are shown to succeed in describing extremal dependence when we work under an asymptotic dependence assumption, but, as with any approach based on multivariate extreme value distributions, both models overestimate risk if data are asymptotically independent. In the time-dependent model, we do not assume marginal nonstationary distributions, meaning that we do not allow for the margins to change over the predictor. However, this is a sensible modeling assumption for our data application, because returns are known to be approximately stationary. Finally, this work focused on componentwise maxima, but wastefulness of data can be avoided by using models for high threshold exceedances. The efficiency of this approach comes with additional drawbacks, such as the selection of a suitable threshold and the suitability of extremal models at low levels, all of this being magnified by the difficulty of working under a nonstationary setting.

## 4.2 Further Modeling

The methodologies presented in this thesis can be applied to different contexts in extreme data applications and extended in several directions. Some of the future works that may be derived from this dissertation are detailed below.

### **Spatial extremes and high threshold exceedances**

Bivariate sequences arise naturally in many real-life data contexts, such as environmental phenomena. When the sequences are accompanied by a spatial dimension, the interest is in the modeling of the spatial dependence within extreme events in continuous space based on observations recorded on a grid. Inspired by the data application of Chapter 2, a sensible approach is to consider spatial modeling, for which max-stable processes (de Haan, 1984) are natural models. More efficient approaches can also be

achieved by considering high threshold exceedances. The models of [Huser and Genton \(2014\)](#) and [Huser and Davison \(2014\)](#) could be an appropriate way to proceed in this respect, avoiding the wastefulness of data and allowing for other covariates to be introduced in the model.

### **Bayesian methodology**

Considering the often low amount of information available in extreme value analyses, it is natural to consider other sources of information, like prior information about the variable under study or related covariates. In this context, Bayesian methods supply a complete paradigm for statistical inference, providing a set of interesting additional statistical tools, such as posterior prediction. Bayesian inference in extreme value analysis can be carried out from a parametric point of view, but a more appealing framework can be based on Bayesian nonparametrics, since the extreme value dependence structures are infinite-dimensional. In the bivariate setting, the novel work of [Guillote \*et al.\* \(2011\)](#) take a step into that direction, proposing Bayesian inference using a censored likelihood approach, with a prior distribution for the spectral measure concentrated on a countable union of finite dimensional families of smooth spectral measures which is dense in the set of all spectral measures. Nevertheless, its generalization to arbitrary dimensions is not straightforward. A second possible direction is based on approximation of the spectral measure by a mixture of Dirichlet distributions. This approximation is theoretically valid in arbitrary dimension, but the moment constraint makes Bayesian inference very challenging in greater dimension. [Sabourin and Naveau \(2014\)](#) proposed a reparametrization of the Dirichlet mixture model that turns to be unconstrained, and so it can be used in a Bayesian framework to infer the extremal dependence structure. A third direction is based on recent developments for density estimation on compact sets. [Barrientos \*et al.\* \(2012\)](#) proposed models for sets of predictor-dependent probability distributions with bounded domain. Application of these models to extreme value analysis is restricted to the imposition of the moment constraint.

### **Models for asymptotic independence**

As described in Section 3.6, a major drawback of models for multivariate extremes is the classical assumption of asymptotic dependence. [Ledford and Tawn \(1997\)](#) introduced a versatile model to bridge

the gap between asymptotic dependence and independence, but the need for developing more models able to cope with both asymptotic dependence and asymptotic independence is of utmost importance. In the nonstationary context, the hidden angular measure of [Ramos and Ledford \(2009\)](#) can be extended to a predictor-dependent version. Issues related with asymptotic dependence versus asymptotic independence in spatial processes can be addressed with the approach of [Ancona-Navarrete and Tawn \(2000\)](#).

### Higher dimensional setting

Conceptually, it is not difficult to see how to generalize the approaches in Chapters 2 and 3 to arbitrary dimensions. However, from a practical point of view, there are some serious obstacles to be overcome. The spectral measure is an arbitrary probability measure on the  $(d - 1)$ -dimensional unit simplex satisfying the moment constraints in (1.35), and it may have density on each of the  $2^d - 1$  faces of the unit simplex [Coles and Tawn \(1991\)](#). Double kernel estimators in Sections 2.2.3 and 3.3.2 can be extended by replacing the beta kernel by Dirichlet kernels defined on  $\mathbb{R}^{d-1}$ . The moment constraint problem can be overcome by careful selection of the Dirichlet parameters. Efficient methodologies should be also proposed for numerical computations.



## Bibliography

---

- ANCONA-NAVARRETE, M. A. and TAWN, J. A. (2000) A comparison of methods for estimating the extremal index. *Extremes* **3**, 5–38.
- BARRIENTOS, A. F., JARA, A. and QUINTANA, F. A. (2012) Fully nonparametric regression for bounded data using bernstein polynomials. *Technical report, Department of Statistics, Pontificia Universidad Católica de Chile*.
- BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. (2004) *Statistics of extremes: theory and applications*. Wiley & Sons.
- BORTOT, P and TAWN, J. A. (1984) Models for the extremes of Markov chains. *Applied Statistics* **47**, 405–423.
- BOWMAN, A. W. (1984) An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* **71**, 353–360.
- BRUUN, T. and TAWN, J. A. (1984) Comparison of approaches for estimating the probability of coastal flooding. *Applied Statistics* **47**, 405–423.
- BÜTTNER, D. and HAYO, B. (2011) Determinants of European stock market integration. *Economic Systems* **35**, 574–585.
- CAPÉRAÀ, P., FOUGERES, A. L. and GENEST, C. (1997) A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84**, 567–577.
- CHAVEZ-DEMOULIN, V. and DAVISON, A. C. (2005) Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society, Series C* **54**, 207–222.
- CHAVEZ-DEMOULIN, V., EMBRECHTS, P. and HOFERT, M. (2015) An extreme value approach for modeling operational risk losses depending on covariates. *J Risk Insur* (DOI: 10.1111/jori.12059).
- CHEN, S. X. (1997). Empirical likelihood-based kernel density estimation. *Australian Journal of Statistics* **39**, 47–56.
- CHEN, S. X. (1999). Beta kernel estimators for density functions. *Computational Statistics & Data Analysis* **31**, 131–145.

- CHOW, Y-S., GEMAN, S. and WU, L D. (1983). Consistent cross-validated density estimation. *Annals of Statistics* **11**, 25-38.
- CLINE, D. B. H. (1988) Admissible kernel estimators of a multivariate density. *Annals of Statistics* **16**, 1421-1427.
- COLES, S. G. and TAWN, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society, Series B* **53**, 377–392.
- COLES, S. G. and TAWN, J. A. (1994). Statistical methods for multivariate extremes: An application to structural design. *Applied Statistics* **43**, 1–48.
- COLES, S. G., HEFFERNAN, J. and TAWN, J. A. (1999). Dependence measures for extreme value analyses. *Extremes* **2**, 339–365.
- COLES, S. G. (2001). *An Introduction to the Statistical Modeling of Extreme Values*. London: Springer.
- CONDON, L. E., GANGOPADHYAY, S. and PRUITT, J. A. (1994). Climate change and non-stationary flood risk for the upper Truckee River basin. *Hydrology and Earth System Sciences* **19**(1), 159–175.
- COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- DAVISON, A. C. and SMITH, R. L. (1990) Models for exceedances over high thresholds (with Discussion). *Journal of the Royal Statistical Society, Series B* **52**, 393–442.
- DAVISON, A.C. and HINKLEY, D.V. (1997) *Bootstrap Methods and Their Application*. Cambridge, UK: Cambridge University Press.
- DE CARVALHO, M. and DAVISON, A. C. (2011), Semiparametric estimation for  $K$ -sample multivariate extremes. *Proceedings 58th International Statistical Institute*, 2961–2969.
- DE CARVALHO, M., OUMOW, B., SEGERS, J. and WARCHOL, M. (2013) A Euclidean likelihood estimator for bivariate tail dependence. *Communications in Statistics-Theory and Methods* **42**, 1176–1192.
- 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. (2015) *Risk and Extremes: Challenges and Opportunities* (to appear). New York: Wiley.
- DE HAAN L. (1975) On regular variation and its applications to the weak convergence of sample extremes. *Mathematical Centre Tracts* **32**, 1–124.

- DE HAAN, L. and RESNICK, S. I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **40**, 317–377.
- DE HAAN L. (1984) A spectral representation for max-stable processes. *Annals of Probability* **12**, 1194–1204.
- DE HAAN L. and RONDE J. (1998) Sea and wind: multivariate extremes at work. *Extremes* **1**, 7–45.
- DE HAAN, L., NEVES, C. and PENG, L. (2008). Parametric tail copula estimation and model testing. *Journal of Multivariate Analysis* **99**, 1260–1275.
- DEHEUVELS, P. (1991) On the limiting behavior of the Pickands estimator for bivariate extreme value distributions. *Statistics & Probability Letters* **12**, 429–439.
- DOBLER, C., BÜRGER, G. and STÖTTER, J. (2012). Assessment of climate change impacts on flood hazard potential in the Alpine Lech watershed. *Journal of Hydrology* **460**, 29–39.
- DREES, H. (2001) Exceedance over threshold. In *Encyclopedia of Environmetrics*, Vol. 2 (EL-SHAARAWI, A.H. and PIEGORSCH, W.W., eds), 715–728. Wiley.
- DUNSON, D. B., PILLAI, N. and PARK, J H (2007) Bayesian density regression. *Journal of the Royal Statistical Society, Series B* **69**, 163–183.
- EASTOE, E. and TAWN, J. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society, Series C* **58**, 22–45.
- EASTOE, E. F. (2009). A hierarchical model for non-stationary multivariate extremes: a case study of surface-level ozone and NO<sub>x</sub> data in the UK. *Environmetrics* **20**, 428–444.
- EINMAHL, J., DE HAAN, L. and PITERBARG, V. (2001) Nonparametric estimation of the spectral measure of an extreme value distribution. *Annals of Statistics* **29**, 1401–1423.
- EINMAHL, J., KRAJINA, A. and SEGERS, J. (2008) A method of moments estimator of tail dependence. *Bernoulli* **14**, 1003–1026.
- EINMAHL, J., 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. and SEGERS J. (2009) Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Annals of Statistics* **37**, 2953–2989.
- EMBRECHTS, P., KLUPPELBERG, C. and MIKOSCH, T. (1997) *Modelling Extremal Events*. New York: Springer.

- FERREZ, J., DAVISON, A.C. and REBETEZ, M. (2011) Extreme temperature analysis under forest cover compared to an open field. *Agricultural and Forest Meteorology* **151**, 992–1001.
- FISHER, R. A. and TIPPETT, L. H. C. (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical proceedings of the Cambridge Philosophical Society* **24**, 180–190.
- FOUGERES, A. L. (2004) Multivariate extremes. In *Extreme Values in Finance, Telecommunications, and the Environment* (FINKENSTADT, B. and ROOTZEN, H., eds), 373–388. New York: Chapman & Hall/CRC.
- FRATZSCHER, M. (2002) Financial market integration in Europe: on the effects of EMU on stock markets. *International Journal of Finance & Economics* **7**, 165–193.
- GARDES, L. and GIRARD, S. (2010). Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes* **13**(2), 177–204.
- GEENENS, G. (2014). Probit transformation for kernel density estimation on the unit interval. *Journal of the American Statistical Association* **109**, 346–359.
- GEORGIEV, A. A. (1988) Asymptotic properties of the multivariate Nadaraya–Watson regression function estimate: the fixed design case. *Statistics & probability letters* **7**, 35–40.
- GNEDENKO, B.V. (1948) On a local limit theorem of the theory of probability. *Uspekhi Matematicheskikh Nauk* **3**, 187–194.
- GUDENDORF, G. and SEGERS, J. (2011) Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *Journal of multivariate analysis* **102**(1), 37–47.
- GUDENDORF, G. and SEGERS, J. (2012) Nonparametric estimation of multivariate extreme-value copulas. *Journal of Statistical Planning and Inference* **142**(12), 3073–3085.
- GUILLOTE, S., PERRON, F. and SEGERS, J. (2011) Non-parametric Bayesian inference on bivariate extremes. *Journal of the Royal Statistical Society, Series B* **59**, 475–499.
- GUMBEL, E.J. (1958) *Statistics of Extremes*. New York: Columbia University press.
- HALL, P. and PRESNELL, B. (1999) Density estimation under constraints. *Journal of Computational and Graphical Statistics* **8**(2), 259–277.
- HARDLE, W. (1990) *Applied Nonparametric Regression*. Cambridge, UK: Cambridge University Press.
- HARDOUVELIS, G. A., MALLIAROPULOS, D. and PRIESTLEY, R. (2006) EMU and European stock market integration. *The Journal of Business* **79**, 365–392.

- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2001) *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, New York: Springer.
- HEFFERNAN, J. E. (2000). Directory of coefficients of tail dependence. *Extremes* **3**, 279–290.
- 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**, 497–546.
- HUSER, R. and GENTON, M. G. (2014) Non-stationary dependence structures for spatial extremes. *arXiv preprint (arXiv:1411.3174)*.
- HUSER, R. and GENTON, M. G. (2014) Space-time modelling of extreme events. *Journal of the Royal Statistical Society, Series B* **76**(2), 439–461.
- JAMES, H. (2012). *Making the European Monetary Union*. Cambridge, MA: Harvard University Press.
- JOE H. (1994) Multivariate extreme value distributions with applications to environmental data. *Canadian Journal of Statistics* **22**, 47–64.
- JONATHAN, P., EWANS, K. and RANDELL, D. (2014) Non-stationary conditional extremes of northern north sea storm characteristics. *Environmetrics* **25**, 172–188.
- JONES, M. C. and HENDERSON, D. A. (2007). Kernel-type density estimation on the unit interval. *Biometrika* **94** 977–984.
- KIM, S. J., MOSHIRIAN, F. and WU, E. (2005) Dynamic stock market integration driven by the European monetary union: An empirical analysis. *Journal of Banking & Finance* **29**, 2475–2502.
- KIRILIOUK, A., SEGERS, J. and WARCHOL, M. (2015) Nonparametric estimation of extremal dependence. In: *Extreme Value Modelling and Risk Analysis: Methods and Applications*, Eds D Dey, J Yan. Boca Raton, FL: Chapman and Hall/CRC.
- KOENKER, R. (2005) *Quantile Regression*. Cambridge, UK: Cambridge University Press.
- KOTZ, S. and NADARAJAH, S. (2000) *Extreme Value Distributions*. Imperial College Press.
- LANE, P. R. (2012) The European sovereign debt crisis. *The Journal of Economic Perspectives* **26**, 49–67.
- LEDFORD, A.W. and TAWN, J.A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169–187.

- LEDGORD, A.W. and TAWN, J.A. (1996) Extreme value theorems of uncertain process with application to insurance risk model. *Soft Computing* **17**(4), 549–556.
- LEDGORD, A.W. and TAWN, J.A. (1997) Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society, Series B* **59**, 475–499.
- LINDSAY, B. G. (1988) Composite Likelihood Methods. *Contemporary Mathematics* **80**, 221–239.
- MAULIK, K., RESNICK, S. and ROOTZÉN, H. (2002) Asymptotic independence and a network traffic model. *Journal of Applied Probability* **39**, 671–699.
- NADARAYA, E.A. (1964) On estimating regression. *Theory of Probability & Its Applications* **9**, 141–142.
- NATURAL ENVIRONMENT RESEARCH COUNCIL (1975) *Flood Studies Report*. London: Natural Environment Research Council.
- NELDER, J. A. and WEDDERBURN, R. W. M. (1972) On estimating regression. *Theory of Probability & Its Applications* **9**, 141–142.
- NIKOULOPOULOS, A. K., JOE, H. and LI, H. (2009) Extreme Value Properties of Multivariate  $t$  Copulas. *Extremes* **12**(2), 129–148.
- OPITZ, T. (2013) Extremal  $t$  processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis* **122**(1), 409–413.
- OWEN, A. (2001), *Empirical Likelihood*. Boca Raton, FL: Chapman & Hall.
- PACIOREK, C. J. and SCHERVISH, M. (2006) Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17**(5), 483–506.
- PICKANDS, J. (1981) Multivariate extreme value distributions. *Proceedings 43rd Session International Statistical Institute* **2**, 859–878.
- POON, S.-H., ROCKINGER, M. and TAWN, J. A. (2003) Modelling extreme-value dependence in international stock markets. *Statistica Sinica* **13**, 929–953.
- POON, S.-H., ROCKINGER, M. and TAWN, J. A. (2004) Extreme-value dependence in financial markets: Diagnostics, models and financial implications. *Review of financial studies*, **17**, 581–610.
- R CORE TEAM (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- 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.
- REISS, R. D. and THOMAS, M. (2001). *Statistical Analysis of Extreme Values*. Birkhauser.
- RESNICK, S., ROOTZÉN, H. (2000) Self-similar communication models and very heavy tails. *Annals of Applied Probability* **10**, 753–778.
- RESNICK, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer.
- SABOURIN, A. and NAVEAU, P. (2014) Bayesian dirichlet mixture model for multivariate extremes: A re-parametrization. *Computational Statistics & Data Analysis* **71**, 542–567.
- SCHLATHER, M. and TAWN, J. (2003) A dependence measure for multivariate and spatial extreme values: properties and inference. *Biometrika* **90**, 139–156.
- SCHMIDT, R. and STADTMÜLLER, U. (2006) Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics* **33**, 307–335.
- SEGERS, J. (2012) Max-stable models for multivariate extremes *arXiv preprint arXiv:1204.0332*.
- SIBUYA, M. (1960) Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics* **11**, 195–210.
- SMITH, R.L. (1994) Multivariate threshold methods. In *Extreme Value Theory and Applications*, (GALAMBOS, J., LECHNER, J. and SIMIU, E., eds), 225–248. Kluwer.
- SMITH, R. L., TAWN, J. A. and COLES, S. G (1997) Markov chain models for threshold exceedances. *Biometrika* **84**, 249–268.
- TAWN, J. A. (1988) Bivariate extreme value theory: Models and estimation. *Biometrika*, **75**, 397–415.
- TAWN, J. A. (1990) Modelling multivariate extreme value distributions. *Biometrika*, **77**, 245–253.
- THIBAUD, E. and OPITZ, T. (2014) Efficient inference and simulation for elliptical Pareto processes. *arXiv:1401.0168v1*.
- TOWLER, E., RAJAGOPALAN, B., GILLELAND, E., SUMMERS, R. S., YATES, D. and KATZ, R. W. (2010). Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory. *Water Resources Research* **46**(11), W11504.
- VARIN, C., REID, N. and FIRTH, D. (2011) An Overview of Composite Likelihood Methods *Statistica Sinica* **21**, 5–42.
- WAND, M.P. and JONES M.C. (1994) *Kernel smoothing*. CRC Press.
- WATSON, G.S. (1964) Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics* **26**, 359–372.