

PhD THESIS DECLARATION

The undersigned

SURNAME Marcon

FIRST NAME Giulia

PhD Registration Number 1538537

Thesis title: Non-Parametric Inference and Simulation
in Extreme Value Theory

PhD in Statistics

Cycle 26th

Candidate's tutor Simone A. Padoan

Year of thesis defence 2015/2016

DECLARES

Under her responsibility:

- 1) that, according to Italian Republic Presidential Decree no. 445, 28th December 2000, mendacious declarations, falsifying records and the use of false records are punishable under the Italian penal code and related special laws. Should any of the above prove true, all benefits included in this declaration and those of the temporary embargo are automatically forfeited from the beginning;
- 2) that the University has the obligation, according to art. 6, par. 11, Ministerial Decree no. 224, 30th April 1999, to keep a copy of the thesis on deposit at the "Biblioteche Nazionali Centrali" (Italian National Libraries) in Rome and Florence, where consultation will be permitted, unless there is a temporary embargo protecting the rights of external bodies and the industrial/commercial exploitation of the thesis;

- 3) that the Bocconi Library will file the thesis in its “Archivio istituzionale ad accesso aperto” (institutional registry) which permits online consultation of the complete text (except in cases of a temporary embargo);
- 4) that, in order to file the thesis at the Bocconi Library, the University requires that the thesis be submitted online by the candidate in unalterable format to Società NORMADEC (acting on behalf of the University), and that NORMADEC will indicate in each footnote the following information:
 - thesis Non-Parametric Inference and Simulation in Extreme Value Theory;
 - by Marcon Giulia;
 - defended at Università Commerciale “Luigi Bocconi” – Milano in January, 2016;
 - the thesis is protected by the regulations governing copyright (Italian law no. 633, 22th April 1941 and subsequent modifications). The exception is the right of Università Commerciale “Luigi Bocconi” to reproduce the same for research and teaching purposes, quoting the source;
 - the thesis is subject to a temporary embargo for 24 months;
- 5) that the copy of the thesis submitted online to NORMADEC is identical to the copies handed in/sent to the members of the Thesis Board and to any other paper or digital copy deposited at the University offices, and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;
- 6) that the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (Italian law, no. 633, 22nd April 1941 and subsequent integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal, and shall be exempt from any requests or claims from third parties;
- 7) that the thesis meets one of the temporary embargo hypotheses included in the declaration “TEMPORARY EMBARGO REQUEST OF THE PhD THESIS” undersigned elsewhere.

Date 13/11/2015

SURNAME Marcon

FIRST NAME Giulia

Contents

1	Introduction	1
1.1	Proposal	3
1.2	Computational Details	5
2	Preliminaries and Background	7
2.1	Univariate Extremes	8
2.2	Multivariate Extremes	9
2.2.1	Dependence Measures	12
3	Extremal Dependence Modelling	17
3.1	A Review of Parametric Models	17
3.2	A Review of Nonparametric Estimators	21
4	Nonparametric Estimation of a Pickands Dependence Function	25
4.1	Madogram Estimator	26
4.2	Estimation based on Bernstein Polynomials	28
4.2.1	Bernstein Polynomials on the Simplex \mathcal{S}_d	28
4.2.2	Shape Preserving Estimator	30
4.2.3	Confidence Bands	34
4.3	Asymptotic Results	36
4.4	Numerical Results	40
4.5	Weekly Maxima of Hourly Rainfall in France	50
5	Simulation based on Bernstein polynomial representations	53
5.1	Motivation	54
5.2	Extremal Dependence Functions in Bernstein Form	55
5.3	Simulation Scheme	57
6	Bayesian Inference	65
6.1	Extremal Dependence	66
6.2	Bayesian Modelling of H and A	68

6.3 Model Formulation and Algorithms	71
--	----

List of Figures

1.1	Map of French weekly precipitation maxima from 1993-2011.	2
2.1	Example of bivariate critical regions, angular density and Pickands dependence function	16
4.1	Visual illustration of the gain in implementing the Bernstein-Bézier projection of \widehat{A}_n	30
4.2	Confidence bands of $\widehat{\beta}$	36
4.3	Asymptotic convergence of the madogram and its Bernstein projection	38
4.4	Estimates of BP-MD with three different values of the polynomial order	40
4.5	Estimates of the Pickands dependence function for the symmetric logistic model obtained with four nonparametric estimators and their projected version and different sample sizes	41
4.6	Estimates of the bivariate Pickands with bootstrap variability bands for the SL, AL, HR, EST models with strong, mild and weak dependence	47
4.7	Estimates of trivariate Pickands and bootstrap variability bands for the SL, AL, HR, EST models with strong, mild and weak dependence	48
4.8	Length of the β vector of coefficients, p_k , when increasing the order of the polynomial k and the dimension d	49
4.9	Estimated pairwise extremal coefficients and Pickands dependence functions of French weekly precipitation maxima	51
5.1	Simulated bivariate vectors of heavy rainfall in France	59
5.2	Modelling of h and A , for the Symmetric Logistic model, and Simulation	60
5.3	Modelling of h and A , for the Asymmetric Logistic model, and Simulation	61
5.4	Modelling of h and A , for the Hüsler-Reiss model, and Simulation	62
5.5	Modelling of h and A , for the Tilted Dirichlet model, and Simulation	63
6.1	Median curve of the MCMC sample from the <i>pseudo</i> -posterior distribution of the Pickands dependence function, along with 95% credible bands, and Boxplots of posterior samples of each coefficient β_j	75

6.2	Pickands dependence function and angular distribution function of the Symmetric Logistic model with strong, mild and weak dependence, estimated by the MCMC posterior and MCMC <i>pseudo</i> -posterior samples. . . .	80
6.3	Pickands dependence function and angular distribution function of the Asymmetric Logistic model with strong, mild and weak dependence, estimated by the MCMC posterior and MCMC <i>pseudo</i> -posterior samples. . . .	80
6.4	Pickands dependence function and angular distribution function of the Hüsler-Reiss model with strong, mild and weak dependence, estimated by the MCMC posterior and MCMC <i>pseudo</i> -posterior samples.	81
6.5	Pickands dependence function and angular distribution function of the Tilted Dirichlet model with strong, mild and weak dependence, estimated by the MCMC posterior and MCMC <i>pseudo</i> -posterior samples.	81

List of Tables

4.1	Pointwise and simultaneous coverage probabilities	35
4.2	MISE of four estimators of the Pickands dependence function based on a symmetric logistic dependence model for different parameter values and sample sizes.	43
4.3	MISE of four estimators of the Pickands dependence function based on an asymmetric logistic, Hüsler-Reiss and extremal skew- t dependence models for different parameter values	44
4.4	Percentage improvement of the MISE gained with the projection method for the symmetric logistic model.	45
4.5	Percentage improvement of the MISE when using the projection method, for the asymmetric logistic, Hüsler-Reiss and extremal skew- t models.	46
4.6	Expectation of $\frac{1}{Q} \sum_{q=1}^Q \left(\mathbf{b}_k(\mathbf{w}_q) \hat{\beta}_k - \hat{A}_n(\mathbf{w}_q) \right)^2$ for increasing values of Q for bivariate and trivariate maxima vectors.	49
4.7	Computational time (in minutes) for the estimation of the Pickands dependence function by means of the projection estimator $\tilde{A}_{n,k}$	49

Tesi di dottorato "Non-Parametric Inference and Simulation in Extreme Value Theory"
di MARCON GIULIA

discussa presso Università Commerciale Luigi Bocconi-Milano nell'anno 2016

La tesi è tutelata dalla normativa sul diritto d'autore (Legge 22 aprile 1941, n.633 e successive integrazioni e modifiche).

Sono comunque fatti salvi i diritti dell'università Commerciale Luigi Bocconi di riproduzione per scopi di ricerca e didattici, con citazione della fonte.

List of Symbols

\mathcal{A}	Family of convex functions A with upper and lower bound defined in (C2), page 13
\mathcal{A}_k	Constrained multivariate Bernstein-Bézier polynomial families in \mathcal{A}
\mathcal{B}_k	Family of Bernstein-Bézier polynomials of degree k
\mathcal{C}	Space of continuous real-valued functions
\mathcal{H}	Space of valid angular distributions
\mathcal{H}_p	Subspace of polynomial angular distributions in \mathcal{H}
\mathcal{P}_k	Space of order k polynomials
L^2	Hilbert space under the L^2 norm
L^∞	Banach space of bounded real-valued functions
ℓ^∞	Space of sequences of bounded real-valued functions
$T_{\mathcal{A}}$	Tangent cone of \mathcal{A}
\mathcal{S}_d	Unit Simplex on \mathbb{R}^d
$\xrightarrow{\text{as}}$	Almost sure convergence
$\xrightarrow{\text{d}}$	Convergence in distribution of random vectors
\rightsquigarrow	Weak convergence of functions in ℓ^∞
$\ f\ _\infty$	L^∞ norm, $\sup_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) $, for $f : \mathcal{X} \rightarrow \mathbb{R}$
$\ f\ _2$	L^2 norm, $(\int_{\mathcal{X}} f^2(\mathbf{x}) d\mathbf{x})^{1/2}$, for $f : \mathcal{X} \rightarrow \mathbb{R}$
ν	Multivariate \mathbf{w} -madogram
ν_n	Multivariate \mathbf{w} -madogram statistics
$\hat{\nu}_n$	Multivariate empirical \mathbf{w} -madogram

\widehat{A}_n	Preliminary estimator of the Pickands dependence function
$\widetilde{A}_{n,k}$	Shape preserving estimator of the Pickands dependence function
C	Extreme value copula
C_n	Extreme value empirical copula
\mathcal{C}	C -Brownian bridge
\mathcal{C}_n	Zero-mean Gaussian empirical process based on eq. (4.18)
\mathbb{C}	Gaussian process on $[0, 1]^d$ defined in eq. (4.18)
\mathbb{C}_n	Gaussian empirical process
\widehat{C}_n	Extreme value empirical copula with empirical marginal distributions
Z	Gaussian process, defined in Theorem 4.1

Acknowledgements

*I know that history is going to be dominated by an improbable event,
I just don't know what that event will be*
(N. Taleb)

The completion of my Ph.D. thesis has been as exhausting as exciting, more than I would have expected. To stay on the topic of extreme values, I would summarize my Ph.D. through the proverb “bad luck never comes alone”...I started off with a difficult problem, while trying to solve it I encountered some bad luck, and a few grey hair came along. Fortunately, I have also met, during these past five years, several people who helped me to face the difficulties. Now, when I look at my Ph.D. thesis, I see a set of tools to handle simultaneous extremes, with plenty of plots that recall smiles and hammocks.

I would like to begin by giving a special thanks to my advisor *Dr. Simone Padoan*, he has been more than a mentor. I thank him for all the help and guidance while working on my thesis. When I have lacked motivation, he has been patient, understanding and supportive. He has been available whenever I have needed guidance and, most of all, he has believed in me. He managed to instil in me his enthusiasm for extreme value theory and his energy for conducting scientific research. I hope I will as enthusiastic and energetic as he is, and that I will be able to work as well as he does.

I would like to thank *Prof. Pietro Muliere*, who has assisted me as co-advisor and has been a source of knowledge, advices and encouragement, as well as *Prof. Sonia Petrone*, Ph.D. coordinator, for the her suggestions and time dedicated to me. My gratitude is also extended to *Prof. Philippe Naveau* who invited me in Paris for a research collaboration. His work has been supported by the French National Research Agency (ANR-09-RISK-007, MOPERA, ANR McSim), the LEFE-Multirisk and the Extremoscope projects. I also acknowledge Meteo France for the precipitation time series.

I could not have succeeded without the invaluable support of several people. There is a person I need to mention especially, *Isadora Antoniano Villalobos*. She has been a friend, a colleague, a source of friendship and energy, like a sister, ever since we met. She has supported me, not only by providing assistance with the research, but also emotionally through the rough road to finish this thesis.

I also want to thank present and past Ph.D. students that have helped me immensely. I have shared with them enjoyable moments, inside and outside the university. The group has been a source of friendship as well as good advice and collaboration. To name a few, thanks to *Aleks, Amir, Andrea, Irina, Lorenzo, Maria Anna, Mario, Marta, Maxim, Polina, Sajid* and *Silvia*...

On the non-scientific side of things, I want to thank all my caring friends and relatives that encourage me in spite (and probably because) of the distance. In particular, thanks to *Annalisa, Chiara, Gaetano, Laura, Luca* and *Silvia* for providing their support, friendship and keeping me smiling.

Last, but certainly not least, a most special thanks to *mum* and *dad*. Thanks for giving me the possibility to live this experience, for encouraging me when I was tired and about to give up, for enjoying my work and supporting me during (but not only) these past five years, for your unconditional love and care.

I am really grateful. Thank You All!

*Man can believe the impossible,
but man can never believe the improbable*
(O. Wilde)

Abstract

Extreme values of real phenomena are events which have low frequency of occurrence and large uncertainty, but can have a great impact on real life. The statistical analyses of extreme events is of crucial importance for assessing the risk in many areas such as the environmental studies, financial markets, actuarial sciences, engineering, industry and health. In this wide range of applications, the extremes are high-dimensional by nature and it is of interest to model and predict such events. The estimation of extreme values concerns the assessment of the marginal components together with their dependence structure among multivariate maxima. Regardless of the marginal behaviour, there are different approaches to infer the extremal dependence, for instance estimating the stable tail or the Pickands dependence functions, or the so-called spectral or angular measure. The main issue of this thesis is to estimate the extremal dependence of multivariate maxima using a nonparametric approach. Beyond the bivariate case, nonparametric estimation of the dependence function remains a challenging task and an active research field. We explore two methodologies that allow the estimation of the extremal dependence through a polynomial in the Bernstein form where some necessary constraints are easily addressed. The first proposal provides a nonparametric estimator of the Pickands dependence function through the multivariate madogram. Additionally, its regularized version based on a projection procedure takes into account the shape constraints on the family of Pickands dependence functions by means of a representation in terms of the Bernstein-Bézier polynomials. The large-sample theory of the estimator is developed and its finite-sample performance is evaluated with a simulation study. Second, we propose an inferential Bayesian framework, in the bivariate case, which is versatile in the sense of simultaneously estimating the extremal dependence in terms of the Pickands dependence function and the spectral measure. In order to derive a valid extremal dependence structure, the angular measure must satisfy what is known as the mean condition, analogous to the shape restrictions of the Pickands dependence function. Once the joint behaviour is detected from the data, using a nonparametric model, we can simulate bivariate extremes. The approaches are illustrated by analysing clusters consisting of seven weather stations that have recorded weekly maxima of hourly rainfall in France from 1993 to 2011.

Chapter 1

Introduction

*Il est impossible que
l'improbable n'arrive jamais*
(E. Gumbel)

Statisticians often may work as consultants for economists, engineers or hydrologists. In the past, they were usually asked to answer questions like “*How likely* is this to happen?”. Basic statistical theory has developed around the normal distribution and the central limit theorem, where the normal distribution acts as the limit for the distribution of the sample mean. Nowadays, many statistical problems focus on studying the frequencies of the extremes of a real process rather than its mean, therefore the normal distribution is far from being a satisfactory model. For example, one may need to model large-impact environmental or financial events in order to answer questions like “*How high, how long, how fast* is this to happen?”. At first glance, estimating the chances of events that rarely happen or even have never happened before sounds like a “mission impossible”. The Extreme Value Theory shows under that appropriate conditions it is in fact possible to provide an answer to such a question.

The extremal dependence is a probabilistic concept linked to the tail behaviour where the events have low frequency of occurrence, and large uncertainty. The statistical analysis of extreme events is crucial when considering the risk assessment in areas such as financial markets, environment and health. For example, insurance companies and governments need to monitor and statistically quantify the frequency of natural disasters. It is a key issue, in such situations, to plan suitable risk management and take preventive actions. The scientific question is how to estimate the extremes of natural phenomena observed at several locations in order to understand their simultaneous occurrence. This means their modelling the joint behaviour of several variables at high levels.

In recent years, there has been an increase in the demand of inference methods for assessing the extremal dependence. This is especially due to growing requests for multi-

variate analysis of extreme values in the fields of environmental sciences. A simple way of modelling multivariate extremes is to apply the definition of block-maximum (Coles 2001, Ch. 3) to each of the variables considered, and therefore to focus on the so-called vector of component-wise maxima and its distribution (de Haan and Ferreira 2006, Ch. 6). In these contexts, the dimension of the multivariate vector under study is often greater than two.

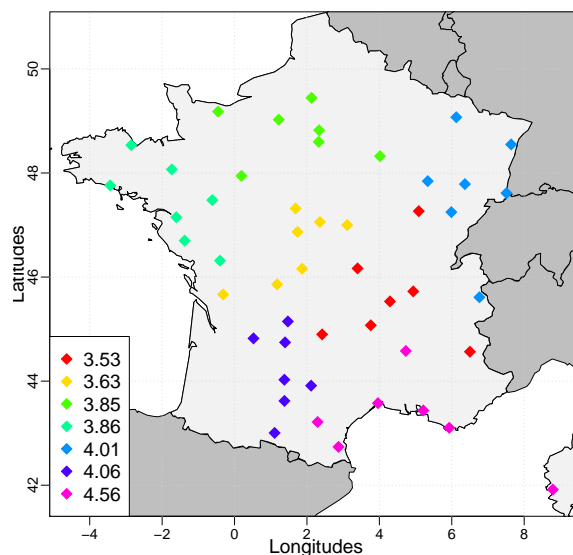


Figure 1.1: Analysis of French weekly precipitation maxima from 1993-2011. Clusters of 49 weather stations and their estimated extremal coefficients in dimension $d = 7$ obtained with the projected version of the madogram estimator, see Section 4.5 for details.

For example, Figure 1.1 displays a map of clusters composed of seven French weather stations each (see Bernard et al. (2013) for details on how the clusters are constructed). The data consist of weekly maxima of hourly rainfall recorded at each station. It would be of interest to hydrologists to infer the dependence within each of the seven-dimensional vectors of component-wise maxima and to compare the dependence structures among clusters. Furthermore, it would be desirable to reproduce such dependence in order to cover missing values or evaluate heavy precipitations in unobserved locations. Such an endeavour represents the main motivation of this thesis.

In order to provide a comprehensive analysis, both the marginal distributions and the dependence structure, need to be estimated. The aim of this thesis is to address the latter with a particular focus on multivariate and high-dimensional extremes.

The modelling of the dependence through parametric models is a widely discussed topic previously used in applications (e.g. Coles 2001, Beranger and Padoan 2015); a downside to this approach is that a specific model can be useful for a certain application but restrictive for others. As a consequence, in recent times much attention is devoted to

study nonparametric estimators or estimation methods for assessing the extremal dependence. While the advantage of a nonparametric approach is the flexibility in capturing the dependence, in high dimensions, i.e. greater than two, the estimation is a challenging task. This thesis addresses the problem of estimating the extremal dependence from observations following a multivariate extreme value distribution with known and unknown margins.

1.1 Proposal

Beyond the bivariate case, nonparametric estimation of the dependence function remains a challenging task and an active research field. The extremal dependence can be inferred alternatively by estimating the angular (or spectral) measure or the stable tail dependence function, where the latter is equivalently represented by the Pickands dependence function. We explore two methodologies that allow the estimation of the extremal dependence through a polynomial in the Bernstein form where some necessary constraints are easily addressed.

The thesis is organised as follows. Chapter 2 provides a background of extreme value theory and introduces the quantities of interest considered in the thesis. Chapter 3 is a brief review of some parametric and nonparametric models commonly used in the literature. Chapters 4, 5 and 6 contain the proposed methods and results developed in this thesis.

Specifically, two nonparametric estimators of the multivariate extremal dependence are provided in Chapter 4, in terms of the Pickands dependence function. The multivariate madogram is easy to compute and it can be interpreted as a kind of distance with respect to the case of complete dependence. Comparing to other nonparametric estimators, it outperforms the modelling of mild and weak dependences, but we show that it may not guarantee the convexity constraint required for the Pickands dependence function to be valid. For this reason, a second estimator has been proposed, which is an approximation of the madogram's projection onto the space of all functions satisfying convexity, specific lower and bounds and endpoints. This is represented in a polynomial form, specifically in a Bernstein form. From the simulation study we observe a gain in efficiency when comparing with several nonparametric estimators. Assessing for the variability of the function estimates is also possible, by computing the quantiles of the bootstrap distribution of the polynomial coefficients. The asymptotic properties of the two proposed methods are also discussed. We show the applicability of this framework in high-dimensional problems, analysing the seventh-dimensional dependence structure of heavy rainfall in France. These results have been presented at several conferences and seminars, and the paper [Marcon et al. \(2015\)](#) was awarded a prize in the poster session

of METMA VII - GRASPA14 conference.

Despite the ability of the proposed method for estimating multivariate max-stable maxima, there are some drawbacks. The first arises from the choice of the polynomial order, which makes the estimate more or less smooth. Simulation results show that an adequate choice is related to the strength of the dependence, so it would be useful to compute an index to synthesise this information. At the same time, the Bernstein approximated projection relies on a pilot estimate. Basic nonparametric estimators are usually good as first guesses but they appear unstable for small sample sizes. For these reasons, in Chapter 6 we approach the problem, for the bivariate setting, from a Bayesian perspective.

In chapter 6 we describe an inferential Bayesian framework that is versatile when estimating the extremal dependence. Through the use of polynomials in a Bernstein form, it is possible to estimate, in the bivariate case, the multiform by which the extremal dependence can be with two key peculiarities of the proposal. In the first place, it is possible to accommodate different representations of the dependence structure such as the Pickands dependence function and the spectral or angular measure. We recall that both dependence forms need to satisfy some constraints so that the extreme value distribution is a proper distribution. In second place, model fitting, inference and model assessment can be achieved via posterior simulation techniques. The main benefits that this entails are that the selection of the polynomial degree is performed by estimating it as part of the inferential procedure and there is no need for a preliminary estimate in order to perform a constraint estimation. Additionally, the estimation of marginal distributions is possible through a unique inferential procedure, even though it has not been discussed in this thesis. Although the analytical expression of the posterior distribution of the model parameters is unknown, inference is however possible via a trans-dimensional MCMC scheme. We show the ability of the proposed method to capture the extremal dependence of some popular parametric models, by means of a simulation study. This work has been presented to some conferences (ISBA 2014, EVA2015 and BNP10, in the last of which it was awarded a prize in the poster session). An extended abstract has been pre-accepted and the paper will be submitted soon.

The flexibility of Bernstein polynomials is exploited also in Chapter 5. The possibility of rewriting the Bernstein polynomial basis functions as Beta densities allows us to model the angular distribution as a mixture of densities and provide an easy simulation scheme for the angular component, once the dependence structure is modelled. This constitutes a working paper together with S. Padoan, P. Naveau and P. Muliere.

Other aspects of the proposed method can be examined in depth. For example, in Section 2.2.1 we see that, knowing the angular distribution which describes the dependence structure of data, allows the calculation of an approximated estimation of the probability

of exceedance,

$$P(Y_1 > y_1, Y_2 > y_2) \approx 2 \int_{\mathcal{S}_d} \min \left\{ \frac{w}{y_1}, \frac{1-w}{y_2} \right\} dH(w).$$

Also referred to as “return period”, this illustrates how likely and severe extreme events (such as heavy rainfall, flooding, and earthquakes) can be. It provides the probability that an event of any given magnitude will occur in any given period. This probability is often used to assess the risk of these events, especially for human populations. Estimating “return values” is also of interest because it allows to compute quantile regions with very small probability.

A further and challenging extension may be the inclusion of information, from the spatial domain and covariates, for the prediction of extreme events. The estimation of future extreme episodes of a real process, such as heavy-rainfall, heat-waves, simultaneous losses in the financial market, etc., is of a crucial importance for risk management. For an accurate assessment of multiple extreme values, their dependence structure needs to be appropriately modelled and inferred.

1.2 Computational Details

Simulations and data analysis were performed using the **R** package `ExtremalDep` (https://r-forge.r-project.org/R/?group_id=1998), which contains all functions implemented to get the following results. It provides parametric and nonparametric statistical methods for modelling and estimating the dependence of extreme values (Beranger et al. 2015). Specifically, the nonparametric modelling corresponds to the developments in Chapter 4. The functions `madogram`, `beed` and `beed.confband` provide the nonparametric estimation of the Pickands dependence function and its simultaneous confidence bands. The functions used to obtain the results described in Chapters 5 and 6 will be available soon.

Tesi di dottorato "Non-Parametric Inference and Simulation in Extreme Value Theory"
di MARCON GIULIA

discussa presso Università Commerciale Luigi Bocconi-Milano nell'anno 2016

La tesi è tutelata dalla normativa sul diritto d'autore (Legge 22 aprile 1941, n.633 e successive integrazioni e modifiche).

Sono comunque fatti salvi i diritti dell'università Commerciale Luigi Bocconi di riproduzione per scopi di ricerca e didattici, con citazione della fonte.

Chapter 2

Preliminaries and Background

*Say you were standing with one foot in the oven and one foot in an ice bucket.
According to the percentage people, you should be perfectly comfortable.*

(B. Bragan)

The joke cited in this Chapter has been chosen to remark that when analysing real phenomena, sometimes, any average masks the extreme values in the set of data, which can be of interest. An average is a useful way of summarizing data, nevertheless, it can sometimes be misleading. The main concern in Extreme Value Theory (EVT) is estimating the tail of the distribution, in which case the central observations are irrelevant for statistical inference. EVT focuses on the study and development of stochastic models that can be used for inference on applied problems related to the frequency of the unusual rather than the usual. There is now an extensive range of applications of EVT in many scientific disciplines, most of them concerning the long range prediction of extreme quantities of interest. Opposite to the central limit theorem, EVT starts from the limit distribution of the sample maximum. This idea is different from a usual parametric approach in the sense that only the tail is parametrized without imposing any assumption on the central behaviour. Two main approaches are used in this field, the block-maximum and the peak over a threshold. The extreme value basic theory is based on the generalized extreme value (GEV) distribution and the generalized Pareto distribution (GPD). In this chapter we provide a brief review of some theoretical results on the extreme values of univariate and multivariate vectors of maxima. Several dependence structures can be described and here we focus on the block-maximum approach. In the next chapter we discuss the main features of some families of parametric and nonparametric extremal dependence models.

2.1 Univariate Extremes

Let M_n denotes the maximum of n independent and identical distributed (i.i.d.) variables X_1, \dots, X_n with common distribution function F , i.e.,

$$M_n = \max(X_1, \dots, X_n).$$

Fisher and Tippett in the seminal work of (1928) conjectured that if there exist sequences of constants $a_n > 0$ and b_n such that

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \xrightarrow{d} G(x)$$

where G is a non-degenerate distribution function, then G belongs to the generalized extreme value (GEV) family of distributions:

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \quad (2.1)$$

for some constants ξ, μ and $\sigma > 0$ and $(x)_+ = \max(0, x)$. For $\xi > 0$ and $\xi < 0$ the generalized extreme value distribution reduces to the Fréchet and the negative Weibull distributions, respectively. In the limit case when $\xi \rightarrow 0$ the GEV distribution becomes the Gumbel distribution.

When extending to random vectors, we face the concept of “tail dependence” which characterizes the relation among a few random variables for being simultaneously extreme. Practically, it successfully explains why “bad luck never comes alone”. If in the univariate case defining extreme values is an easy task, one of the problems in multivariate extremes is that there is no univocal definition of extremes, and defining the associated distribution function becomes more complicated. For this reason, the block-maxima approach has been introduced (Gumbel 1958). It consists in splitting the observation period into non-overlapping periods of equal size and consider the maximum observation in each period. The new observations approximately follow, under domain of attraction conditions (see next Section), an extreme value distribution.

The modern-day analysis of risk requires the estimation of the dependence among univariate and multiple extreme episodes. An accurate estimation of extremal dependence is crucial in order to not underestimate the return period of extremes. In literature there are several theoretical results, statistical models and applications. For more background on univariate and multivariate extreme-value theory, see Kotz and Nadarajah (2000), Coles (2001), Beirlant et al. (2004), de Haan and Ferreira (2006), Resnick (2007) and Falk et al. (2010).

2.2 Multivariate Extremes

The block-maxima approach is extended to the multivariate case. Define the vector $\mathbf{M}_n = (M_{n,1}, \dots, M_{n,d})$, where $M_{n,j} = \max(X_{1,j}, \dots, X_{n,j})$ is the componentwise maxima vector of an \mathbb{R}^d -valued random vector $\mathbf{X} = (X_1, \dots, X_d)$ with $\mathbf{X}_1, \dots, \mathbf{X}_n$ being n independent and identically distributed (i.i.d.) copies of \mathbf{X} . Specifically, let $I = \{1, \dots, d\}$ be an index set, with $d \in \mathbb{N}$, and F be the joint probability distribution function of \mathbf{X} , and marginal distribution functions $F_j = F(\infty, \dots, x_j, \dots, \infty)$, $j \in I$. The distribution F is typically unknown in applications, for this reason we study the asymptotic behaviour of M_n as the sample size tends to infinity. For large, but finite, sample sizes, this results in an appropriate approximate distribution for multivariate extremes. This definition of multivariate extremes, also describes how likely is that all components of \mathbf{X} are simultaneously large.

With a similar approach to the univariate case, the asymptotic distribution of the standardization of \mathbf{M}_n is given by,

$$P\left(\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \leq \mathbf{x}\right) = F^n(\mathbf{a}_n \mathbf{x} + \mathbf{b}_n) \xrightarrow{d} G(\mathbf{x}), \quad n \rightarrow \infty, \quad (2.2)$$

for all the continuity points, \mathbf{x} , of a non-degenerate distribution G , assuming that there are sequences of normalizing constants $\mathbf{a}_n = (a_{n1}, \dots, a_{nd}) > \mathbf{0}$, with $\mathbf{0} = (0, \dots, 0)$, and $\mathbf{b}_n = (b_{n1}, \dots, b_{nd}) \in \mathbb{R}^d$ which provide the convergence. The limiting distributions (2.2) identify a class of **multivariate extreme value distributions** (MEVDs) (Resnick 2007). Any distribution function F that satisfies the convergence result (2.2) is said to be in the **(maximum) domain of attraction** of G , $F \in D(G)$ (de Haan and Ferreira 2006). Moreover, a distribution G on \mathbb{R}^d is **max-stable** if, for every $n \in \mathbb{N}$, there exists sequence $\mathbf{a}_n > \mathbf{0}$ and $\mathbf{b}_n \in \mathbb{R}^d$ such that

$$G(\mathbf{a}_n \mathbf{x} + \mathbf{b}_n) = G^{1/n}(\mathbf{x}). \quad (2.3)$$

This property also implies the **max-infinite divisibility** where G is such that G^a is a distribution for every $a > 0$. Specifically, G can be defined through the distribution of the sample maxima of n i.i.d. random vectors, and it is max-infinite divisible if, for any $n \in \mathbb{N}$, there exists a distribution F_n such that $G = F_n^n$ (Resnick 2007).

The characterization of the class of MEVDs requires the specification of the marginal behaviour and the form of the dependence structure.

Firstly, by setting all x_j but one equal to infinity in (2.2), we see that $F_j \in D(G_j)$, i.e.

$$F_j^n(a_{jn}x_j + b_{jn}) \xrightarrow{d} G_j(x_j) \quad n \rightarrow \infty.$$

Thus, choosing a_{jn} and b_{jn} for all $j \in I$ as in [de Haan and Ferreira \(2006\)](#), implies that each marginal distribution of G is a generalized extreme value distribution of the form given by (2.1). The continuity of the marginal distributions implies that G is also continuous.

A suitable choice of the elements a_{nj} and b_{nj} of the sequences \mathbf{a}_n and \mathbf{b}_n ensures that, for all $y_j > 0$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} F^n \{F_1^{\leftarrow}(1 - 1/ny_1), \dots, F_d^{\leftarrow}(1 - 1/ny_d)\} \\ &= G \left(\frac{\sigma_1(y_1^{\xi_1} - 1)}{\xi_1} + \mu_1, \dots, \frac{\sigma_d(y_d^{\xi_d} - 1)}{\xi_d} + \mu_d \right) \equiv G_0(\mathbf{y}), \end{aligned} \quad (2.4)$$

for all continuity points, $\mathbf{y} > \mathbf{0}$, of G_0 , where $F_j^{\leftarrow}(1 - 1/a)$, with $a > 1$, is the left-continuous inverse of F_j , for all $j \in I$. Then, G is transformed so that its marginal distributions are equal, and the resulting G_0 is a MEVD with identical unit Fréchet marginal distributions, $F_j(y_j) = \exp\{-1/y_j\}$, $j \in I$. Now, if we take the logarithm on $F^n \{F_1^{\leftarrow}(1 - 1/ny_1), \dots, F_d^{\leftarrow}(1 - 1/ny_d)\}$ and use its first order Taylor expansion, it leads to the following measure,

$$\phi_n(\mathcal{R}_y) = n[1 - F\{F_1^{\leftarrow}(1 - 1/ny_1), \dots, F_d^{\leftarrow}(1 - 1/ny_d)\}],$$

defined on the critical region $\mathcal{R}_y = \{(\mathbf{v} \in \mathbb{R}_+^d : v_1 > y \text{ or } \dots \text{ or } v_d > y_d)\}$.

Let us define a random collection of points in $\mathbb{R}^+ \times \mathbb{R}$, $\mathbf{P}_i = \{(i/n, (\mathbf{X}_i - \mathbf{b}_n)/(\mathbf{a}_n))\}_{i=1}^\infty$, with \mathbf{X}_i i.i.d random vectors with distribution F . Let $N_n(\cdot)$ be a PPP, with intensity measure $\lambda \times \phi_n$, where λ is the Lebesgue measure, defined by

$$N_n(E) = \sum_{i=1}^{\infty} \mathbb{I}_{\{\mathbf{P}_i \in E\}}$$

for each Borel set $E \subset \mathbb{R}^+ \times \mathbb{R}$. If the limit in (2.4) holds, then, as $n \rightarrow \infty$, N_n converges weakly to N , i.e. a PPP with intensity measure $\lambda \times \phi$ where $\phi(\mathcal{R}_y) \equiv V(\mathbf{y})$, for $\mathbf{y} > \mathbf{0}$, named **exponent measure** (see [de Haan and Ferreira 2006](#)) and such that it satisfies $\phi(\infty) = 0$ (see [Falk et al. 2010](#)). Now, we are able to count the rescaled observations that fall in a critical region, having at least one large coordinate, that is

$$\begin{aligned} G_0(\mathbf{y}) &= P(N\{(0, 1] \times \mathcal{R}_y\} = 0) \\ &= \exp\{-[\lambda\{(0, 1]\} \times \phi(\mathcal{R}_y)]\} \\ &= \exp\{-V(\mathbf{y})\} \quad \mathbf{y} > \mathbf{0}, \end{aligned} \quad (2.5)$$

The same relation is provided by the limit in (2.4). We obtain, as $n \rightarrow \infty$, for all $\mathbf{y} > \mathbf{0}$ such that $0 < G_0(\mathbf{y}) < 1$,

$$\lim_{n \rightarrow \infty} n[1 - F\{F_1^{\leftarrow}(1 - 1/ny_1), \dots, F_d^{\leftarrow}(1 - 1/ny_d)\}] = -\log G_0(\mathbf{y}) \equiv V(\mathbf{y}). \quad (2.6)$$

The **exponent dependence function** $V : \mathbb{R}_+^d \rightarrow [0, \infty)$ represents the dependence structure of multiple extremes. Specifically, it is continuous, convex, homogeneous of order -1 , (i.e. $V(a\mathbf{y}) = a^{-1}V(\mathbf{y})$ for any $a > 0$) bounded by $\max(1/y_1, \dots, 1/y_d) \leq V(\mathbf{y}) \leq 1/y_1 + \dots + 1/y_d$, and for all positive y it must satisfy $V(y, 0, \dots, 0) = \dots = V(0, \dots, 0, y) = y$ (for more details see [Falk et al., 2010](#), and [de Haan and Ferreira, 2006](#)). Notice that the interpretation of the dependence form is not straightforward. From (2.6) we see that the derivation of V depends on the functional form of F which, in most real applications, is unknown.

An illustration is provided in Figure 2.1 where the red shaded sets in the panels of the first row coincide with the critical region \mathcal{R}_y . We see that in the case of weak dependence (top-left panels) only one coordinate of the extremes is large, while in the case of strong dependence (top-middle panels) all the coordinates of the extremes are high. Also, asymmetric dependences can exist (top-right panels). However, the exponent function V is not easy to interpret. For this reason, a transformation to pseudo-polar coordinates has been proposed ([Resnick 2007](#)) which simplifies the concept. This lead to a result known as the *spectral decomposition* ([de Haan and Resnick 1977](#)). Let

$$\mathcal{S}_d := \left\{ (v_1, \dots, v_d) \in [0, 1]^d : \sum_{i=1}^d v_i = 1 \right\}, \quad (2.7)$$

be the unit simplex on \mathbb{R}_+^d where one variable is fixed, e.g. $v_d = 1 - (v_1 + \dots + v_{d-1})$. For any $\mathbf{v} \in \mathbb{R}_+^d$, the sub-norm, $\|\mathbf{v}\| = |v_1| + \dots + |v_d|$, measures the distance of \mathbf{v} from $\mathbf{0}$. Other norms can also be considered (e.g. [Resnick 2007](#)). Here, we consider the transformation $T : (r, \mathbf{w}) \mapsto (\|\mathbf{v}\|, \|\mathbf{v}\|^{-1}\mathbf{v})$, with $\mathbf{v} \in \mathbb{R}_+^d / \{\mathbf{0}\}$, and induce a new measure $\psi := \phi * T$. By the homogeneity property of ϕ provided by the (2.3), we obtain that the measure ψ is given by,

$$\begin{aligned} \psi(\mathcal{W}_r) &= \phi\{(\mathbf{v} \in \mathbb{R}_+^d / \{\mathbf{0}\} : \|\mathbf{v}\| > r, \mathbf{v}/\|\mathbf{v}\| \in \mathcal{W})\} \\ &= \phi\{(r\mathbf{u} \in \mathbb{R}_+^d / \{\mathbf{0}\} : \|\mathbf{u}\| > 1, \mathbf{u}/\|\mathbf{u}\| \in \mathcal{W})\} \\ &= r^{-1}H^*(\mathcal{W}), \end{aligned}$$

for all sets $\mathcal{W}_r = r \times \mathcal{W}$ with $r > 0$ and $\mathcal{W} \subset \mathcal{S}_d$, where $H^*(\mathcal{W}) := \phi\{(\mathbf{u} \in \mathbb{R}_+^d / \{\mathbf{0}\} : \|\mathbf{u}\| > 1, \mathbf{u}/\|\mathbf{u}\| \in \mathcal{W})\}$. Hence the measure ψ becomes the product of two independent measures. The **radial (or intensity) measure** ($1/r$), given by the distance of the points from the origin, thus indicating the intensity of the extreme value; and the **angular (or spectral) measure** (H^*) which measures the angular spread of the points (e.g. see [Falk](#)

et al. 2010). Finally,

$$\begin{aligned}
\phi(\mathcal{R}) &= \psi\{T(\mathbf{v} \in \mathbb{R}_+^d / \{\mathbf{0}\} : v_1 > y_1 \text{ or } \dots \text{ or } v_d > y_d)\} \\
&= \psi\{(r, \mathbf{w}) \in (0, \infty) \times \mathcal{S}_d : r > \min(y_j/w_j, j \in I)\} \\
&= \int_{\mathcal{S}_d} \int_{\min(y_j/w_j, j \in I)}^{\infty} r^{-2} dr dH^*(\mathbf{w}) \\
&= \int_{\mathcal{S}_d} \max_{j \in I} (w_j/y_j) dH^*(\mathbf{w}),
\end{aligned} \tag{2.8}$$

where $d\psi(r, \mathbf{w}) = r^{-2} dr \times dH^*(\mathbf{w})$ for all $r > 0$ and $\mathbf{w} \in \mathcal{S}_d$. In this case, those values with radial component higher than a high threshold, are considered extremes and are illustrated by the red points in the second-row panels of Figure 2.1. Dependence increases between all \mathbf{Y}_j as the angular components move to the centre of the simplex (illustrated by the case of strong dependence in the second-row left panel). Components become less dependent as the mass increases near the vertices (the second-row middle panel illustrates the case of weak dependence).

The spectral measure H^* can be any finite measure on \mathcal{S}_d . It must satisfy a first moment conditions

$$\int_{\mathcal{S}_d} w_j dH^*(\mathbf{w}) = 1, \quad \forall j \in I,$$

which is a critical point from the estimation perspective because it is difficult to handle in the multivariate framework. The mean constraint guarantees that the marginal distributions of G_0 are unit Fréchet, so the total mass is equal to

$$H^*(\mathcal{S}_d) = \int_{\mathcal{S}_d} (w_1 + \dots + w_d) dH^*(\mathbf{w}) = \sum_{j \in I} \int_{\mathcal{S}_d} w_j dH^*(\mathbf{w}) = d.$$

A probability measure can be derived by setting $H := H^*/H^*(\mathcal{S}_d)$, thus satisfying

$$\int_{\mathcal{S}_d} w_j dH(\mathbf{w}) = 1/d, \quad \forall j \in I. \tag{2.9}$$

Concluding,

$$G_0(\mathbf{y}) = \exp \left\{ -d \int_{\mathcal{S}_d} \max_{j \in I} (w_j/y_j) dH(\mathbf{w}) \right\}, \tag{2.10}$$

by the combination of (2.4), (2.6), (2.5) and (2.8), and G_0 is a MEVD with unit Fréchet marginal distributions.

2.2.1 Dependence Measures

Different expressions of the extremal dependence have been illustrated above. In general several measures of dependence can be found for multivariate extreme value distributions in the literature. Here we mention some of them.

Extremal dependence

From (2.5) it follows that the extremal dependence can be expressed through the exponent function satisfying continuity, homogeneity of order -1 , convexity and lower and upper limits representing the complete dependence and complete independence cases respectively (see de Haan and Ferreira 2006). Let \mathbf{Y} be a random vector with distribution (2.10). Then, the case of complete dependence, i.e. $Y_1 = Y_2 = \dots = Y_d$, provides $G_0(\mathbf{y}) = \exp\{\max(1/y_1, \dots, 1/y_d)\}$ and occurs when H places the total mass 1 on the centre of the simplex $(1/d, \dots, 1/d)$. When H places mass $1/d$ on the vertices of the simplex, then $G_0(\mathbf{y}) = \exp(1/y_1 + \dots + 1/y_d)$ and Y_1, \dots, Y_d are independent.

The transformation in pseudo-polar coordinates, also provides an easier visualization of the exponent function. Considering its restriction in the simplex, we obtain

$$V(\mathbf{z}) = (z_1 + \dots + z_d) A(t_1, \dots, t_d), \quad \mathbf{z} \in \mathbb{R}_+^d.$$

where $A : \mathcal{S}_d \rightarrow [1/d, 1]$ is the **Pickands dependence function** (Pickands 1981), defined by

$$A(\mathbf{t}) := d \int_{\mathcal{S}_d} \max_{j \in I} (w_j t_j) dH(\mathbf{w}),$$

$z_j = 1/y_j$, $j \in I$, $t_j = z_j/(z_1 + \dots + z_d)$ for $j = 1, \dots, d-1$ and $t_d = 1 - (t_1 + \dots + t_{d-1})$ (see last two rows in Figure 2.1). Observe that A inherits the above properties from V with the obvious modifications. Let \mathcal{A} be the family of functions, $A : \mathcal{S}_d \rightarrow [1/d, 1]$, satisfying the conditions:

(C1) $A(\mathbf{t})$ is convex, that is, $A(a \mathbf{t}_1 + (1-a) \mathbf{t}_2) \leq a A(\mathbf{t}_1) + (1-a) A(\mathbf{t}_2)$,
for $a \in [0, 1]$ and $\mathbf{t}_1, \mathbf{t}_2 \in \mathcal{S}_d$;

(C2) $A(\mathbf{t})$ has lower and upper bounds

$$1/d \leq \bigvee (t_1, \dots, t_{d-1}, t_d) \leq A(\mathbf{t}) \leq 1,$$

for any $\mathbf{t} = (t_1, \dots, t_d) \in \mathcal{S}_d$ with $t_d = 1 - t_1 - \dots - t_{d-1}$;

Notice that condition (C2) implies that $A(\mathbf{0}) = 1$ and $A(\mathbf{e}_i) = 1$ for the boundary points of \mathcal{S}_d , $\mathbf{0} = (0, \dots, 0)$ and $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$ for $i = 1, \dots, d-1$. The lower and upper bounds represent the complete dependence and complete independence cases respectively. This function is a crucial object of study throughout the thesis, together with the angular distribution defined above.

Angular densities

The angular probability measure H , defined on the simplex \mathcal{S}_d , can place mass not only on the interior, but also on other subspaces of the simplex, such as the edges and the

vertices. For this reason, if H is differentiable, it can have several densities on such sets (**angular densities**). In particular, [Pickands \(1981\)](#) and [Kotz and Nadarajah \(2000\)](#) derived the angular densities in the bivariate case. With $d = 2$, the unit simplex $[0, 1]$ can be partitioned into $\mathcal{S}_{2,\{1\}} = \{(1, 0)\}$, $\mathcal{S}_{2,\{2\}} = \{(0, 1)\}$, $\mathcal{S}_{2,\{1,2\}} = \{(w, 1 - w), w \in (0, 1)\}$. The densities on such spaces are given by

$$h_{2,\{1\}} = H(\{0\}) = -y_1^2 \lim_{y_2 \rightarrow 0} \frac{\partial V}{\partial y_1}(y_1, y_2),$$

$$h_{2,\{2\}} = H(\{1\}) = -y_2^2 \lim_{y_1 \rightarrow 0} \frac{\partial V}{\partial y_2}(y_1, y_2),$$

and

$$h_{2,\{1,2\}}(w) = -\frac{\partial^2 V}{\partial y_1 \partial y_2}(w, 1 - w).$$

respectively, for any $y_1, y_2 > 0$. Last density refers to extremes observed in both variables, while the first two densities describe the case when extremes are only observed in one variable. A general result is described in [Coles and Tawn \(1991\)](#) for G absolutely continuous (see also [Resnick 2007](#), Example. 5.13). Specifically, let

$$\mathcal{S}_{d,\mathcal{T}} = \{\mathbf{w} \in \mathcal{S}_d : w_j = 0, \text{ if } j \notin \mathcal{T}; w_j > 0 \text{ if } j \in \mathcal{T}\},$$

be a partition of \mathcal{S}_d , with \mathcal{T} being an index set which takes value in the power set of I , \mathbb{T} . We denote by $h_{d,\mathcal{T}}$ the density on the subspace $\mathcal{S}_{d,\mathcal{T}}$, where $\mathcal{T} \in \mathbb{T}$.

Let $\mathcal{T} = \{i_1, \dots, i_{|\mathcal{T}|}\} \subset I$, and denote by $|\mathcal{T}|$ the cardinality of \mathcal{T} . If G_0 is absolutely continuous, then the angular density for any $\mathbf{y} \in \mathbb{R}_+^d$ is

$$h_{d,\mathcal{T}} \left(\frac{y_{i_1}}{\sum_{i \in \mathcal{T}} y_i}, \dots, \frac{y_{i_{|\mathcal{T}|-1}}}{\sum_{i \in \mathcal{T}} y_i} \right) = - \left(\sum_{i \in \mathcal{T}} y_i \right)^{(|\mathcal{T}|+1)} \lim_{\substack{y_j \rightarrow 0, \\ j \notin \mathcal{T}}} \frac{\partial^{|\mathcal{T}|} V}{\partial y_{i_1} \cdots \partial y_{i_{|\mathcal{T}|}}}(\mathbf{y}). \quad (2.11)$$

The third-row panels of [Figure 2.1](#) show three examples of a bivariate angular densities for the symmetric logistic model ([Gumbel 1960](#)) with a weak, strong and asymmetric dependence respectively.

When \mathcal{T} reduces to a singleton $\{i\}$, for any $i \in I$, the angular density $h_{d,\mathcal{T}}$ represents the mass of H at the vertex \mathbf{e}_i , hence (2.11) simplifies into

$$h_{d,\mathcal{T}} = H(\{\mathbf{e}_i\}) = -y_i^2 \lim_{\substack{y_j \rightarrow 0, \\ j \notin \mathcal{T}}} \frac{\partial V}{\partial y_i}(\mathbf{y}). \quad (2.12)$$

These dependence functions can be used to derive important summaries of the extremal dependence. For example,

$$\vartheta = V(1, \dots, 1) = d \int_{\mathcal{S}_d} \max_{j \in I} (w_j) dH(\mathbf{w}). \quad (2.13)$$

is the **extremal coefficient** (Smith et al. 1990) and it represents the number of independent components of the random vector \mathbf{Y} . The coefficient may take on any value between one and d , where the limiting values describe the cases of complete dependence and complete independence, respectively.

For any $\mathbf{y} > \mathbf{0}$, let us consider the region in $\mathbb{R}_+^d/\{\mathbf{0}\}$ for which all coordinates are simultaneously larger than \mathbf{y} , i.e.

$$\mathcal{F}_{\mathbf{y}} = \{\mathbf{v} \in \mathbb{R}_+^d/\{\mathbf{0}\} : v_1 > y_1 \text{ and } \dots \text{ and } v_d > y_d\}. \quad (2.14)$$

Then, we introduce the **tail dependence** function (Nikoloulopoulos et al. 2009; de Haan and Ferreira 2006) defined by

$$\ell(\mathbf{y}) := \phi\{\{\mathbf{v} \in \mathbb{R}_+^d/\{\mathbf{0}\} : v_1 > y_1 \text{ and } \dots \text{ and } v_d > y_d\}\} \equiv \phi(\mathcal{F}_{\mathbf{y}}), \quad \mathbf{y} > \mathbf{0}.$$

This identifies the number of observations that fall in the failure region and are simultaneous extreme. Using similar arguments to those in (2.8) and (2.9) it follows that

$$\ell(\mathbf{y}) = d \int_{\mathcal{S}_d} \min_{j \in I} (w_j/y_j) \, dH(\mathbf{w}) \quad \mathbf{y} > \mathbf{0}. \quad (2.15)$$

The tail and exponent dependence functions can also be used to approximate the probability that at least one extreme will occur,

$$P(Y_1 > y_1 \text{ or } \dots \text{ or } Y_d > y_d) \approx d \int_{\mathcal{S}_d} \max_{j \in I} (w_j/y_j) \, dH(\mathbf{w}), \quad (2.16)$$

when y_1, \dots, y_d are high enough thresholds and \mathbf{Y} is a random vector with unit Pareto margins.

Furthermore, from previous results, we have that $n\bar{F}(ny_1, \dots, ny_d)$ tends to $\ell(\mathbf{y})$, as $n \rightarrow \infty$, where \bar{F} is the survival function of \mathbf{Y} . From the homogeneity property of V , which is inherited by the function ℓ , we can approximate $\bar{F}(n\mathbf{y})$ by $\ell(n\mathbf{y})$, for large n . Another approximating result can be derived, from the relation (2.15), for the probability that simultaneous extremes will occur:

$$P(Y_1 > y_1 \text{ and } \dots \text{ and } Y_d > y_d) \approx d \int_{\mathcal{S}_d} \min_{j \in I} (w_j/y_j) \, dH(\mathbf{w}), \quad (2.17)$$

when y_1, \dots, y_d are high enough thresholds.

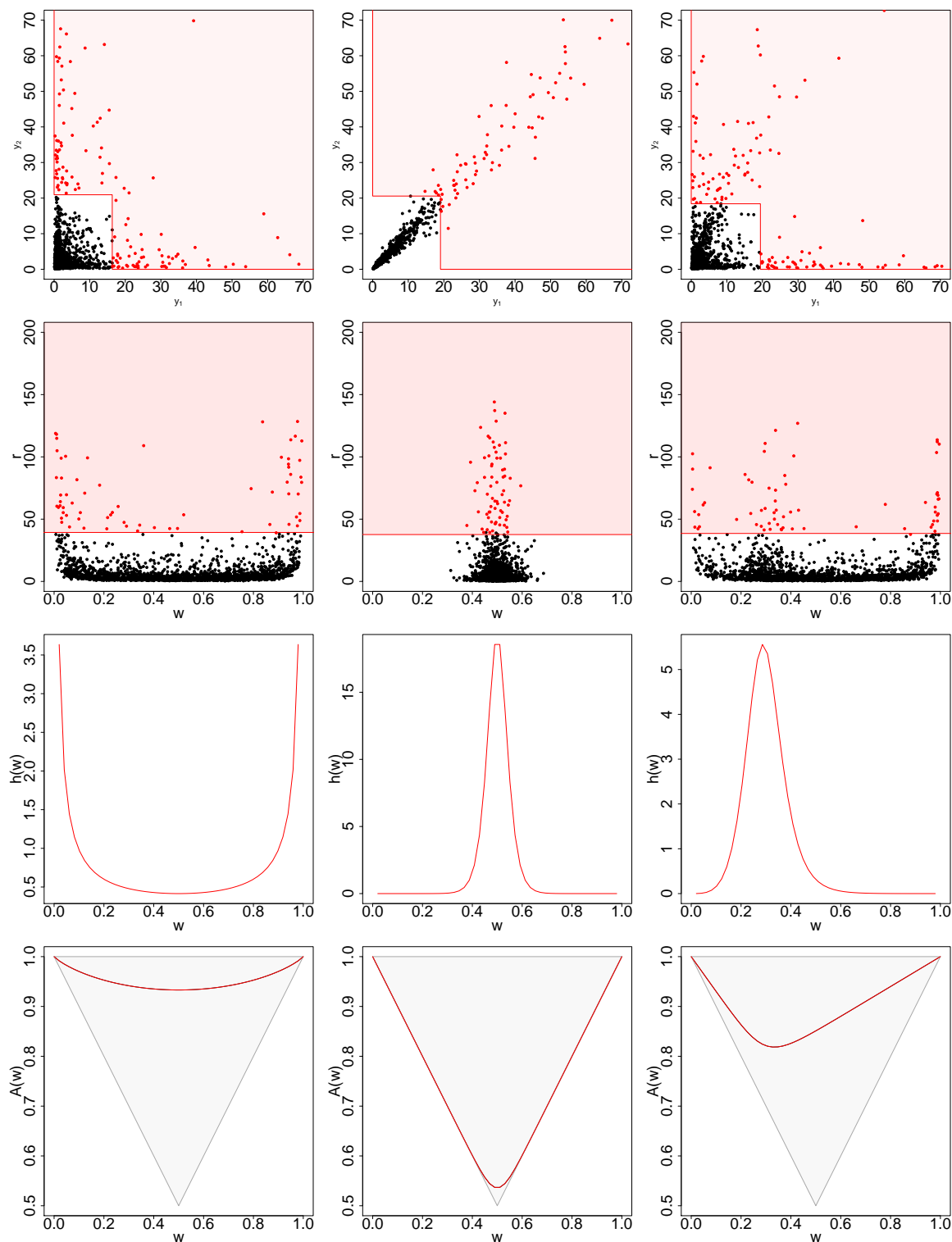


Figure 2.1: Critical regions in \mathbb{R}^2 (top-panels) and their representation in pseudo-polar coordinates (second-row-panels). Red points are the extremes with symmetric weak and strong (left-middle) and asymmetric (right) dependence. Third-row-panels display the angular densities on the simplex and bottom-panels their Pickands dependence function.

Chapter 3

Extremal Dependence Modelling

*The most important questions of life are indeed,
for the most part, really only problems of probability*

(P. Laplace)

Our main interest focuses on multivariate vectors of maxima. There is no finite parametrization that covers the whole class of dependence structures for multivariate extreme value distributions. This leads us to concentrate on nonparametric methods. The function most commonly considered to describe the extremal dependence is the angular distribution introduced in Section 2.2.1. However, the angular measure $H(\mathbf{w})$ can be difficult to handle directly due its mean constraint. For this reason, we consider instead the exponent measure function $V(\mathbf{x})$ through the Pickands dependence function $A(\mathbf{w})$. This chapter reviews some parametric models and nonparametric estimators in in Sections 3.1 and 3.2, respectively. Some of them are considered as comparative methods when studying their performance in some simulation studies.

3.1 A Review of Parametric Models

For applied purposes, it is common practise to use parametric models of multivariate extremes. In particular, the asymmetric and symmetric logistic, Hüsler-Reiss, extremal skew- t and tilted Dirichlet models are widely used. An overview can be found through Kotz and Nadarajah (2000), Coles (2001), Beirlant et al. (2004), Beranger and Padoan (2015) and Padoan (2013). From previous Chapter, it emerges that the angular measure is a crucial element for both the exponent and tail dependence functions.

Symmetric and Asymmetric logistic models

The symmetric logistic (SL) model has been introduced by Gumbel (1960) and it is widely used in the literature and application. However, its simplicity makes it inadequate when

dependences between components differs. The multivariate asymmetric logistic model (AL) extends the symmetric logistic (Tawn 1990, Coles and Tawn 1991) but it has a large number of parameters. Let $\mathcal{T} \in \mathbb{T}$ as in Section 2.2.1, then the exponent function is given by is

$$V(\mathbf{y}) = \sum_{\mathcal{T} \in \mathbb{T}} \left\{ \sum_{j \in \mathcal{T}} (\beta_{j,\mathcal{T}} y_j^{-1})^{1/\alpha_{\mathcal{T}}} \right\}^{\alpha_{\mathcal{T}}}, \quad \mathbf{y} \in \mathbb{R}_+^d,$$

where $1 \geq \alpha_{\mathcal{T}} > 0$, describes the strength of the dependence and $\beta_{\mathcal{T}}$ is such that $\beta_{j,\mathcal{T}} = 0$ if $j \notin \mathcal{T}$, for $j \in I$, $0 \leq \beta_{j,\mathcal{T}} \leq 1$ and $\sum_{\mathcal{T} \in \mathbb{T}} \beta_{j,\mathcal{T}} = 1$. The parameter $\beta_{j,\mathcal{T}}$ influences the asymmetry in the dependence structure. The number of the model parameters is $2^{d-1}(d+2) - (2d+1)$.

The angular measure is defined on all subspaces of the simplex and, from (2.11), for every $\mathcal{T} \in \mathbb{T}$ and all $\mathbf{w} \in \mathcal{S}_{d,\mathcal{T}}$, it follows that it is equal to

$$h_{d,\mathcal{T}}(\mathbf{w}) = \prod_{i=1}^{q-1} (i/\alpha_{\mathcal{T}} - 1) \prod_{j \in \mathcal{T}} \beta_{j,\mathcal{T}}^{1/\alpha_{\mathcal{T}}} w_j^{-(1/\alpha_{\mathcal{T}}+1)} \left\{ \sum_{j \in \mathcal{T}} (\beta_{j,\mathcal{T}}/w_j)^{1/\alpha_{\mathcal{T}}} \right\}^{\alpha_{\mathcal{T}}-q}.$$

The angular density on the interior of the simplex can be simplified, when $\mathcal{T} = I$, $\alpha_{\mathcal{T}} = \alpha$, $\beta_{j,\mathcal{T}} = \beta_j$, to

$$h(\mathbf{w}) = \prod_{i=1}^{d-1} (i/\alpha - 1) \prod_{j \in I} \beta_j^{1/\alpha} w_j^{-(1/\alpha+1)} \left\{ \sum_{j \in I} (\beta_j/w_j)^{1/\alpha} \right\}^{\alpha-d}, \quad \mathbf{w} \in \mathcal{S}_d.$$

The point mass at each extreme point of the simplex is provided when $\mathcal{T} = \{j\}$, for all $j \in I$. It follows from (2.12) and it is equal to $h_{d,\mathcal{T}} = \beta_{j,t}$.

The case of strong dependence arises with values of α close to zero and when α is close to 1, and the other parameters are equal, the case of weak dependence is obtained. In the first case, the mass of the angular density is mainly concentrated towards the centre of the simplex, while in the second case, the mass is concentrated on the vertices of the simplex. In case of an asymmetric dependence structure, the mass tends to be closer to the vertices of the simplex whose corresponding values of β are high.

Concerning the Pickands dependence function of the asymmetric logistic model, we have

$$A(\mathbf{w}) = \sum_{\mathcal{T} \in \mathbb{T}} \left\{ \sum_{j \in \mathcal{T}} (\beta_{j,\mathcal{T}} w_j)^{1/\alpha_{\mathcal{T}}} \right\}^{\alpha_{\mathcal{T}}}, \quad \mathbf{w} \in \mathcal{S}_d,$$

Hüsler-Reiss model

Besides the symmetric and asymmetric logistic models, another popular model is the Hüsler-Reiss (Hüsler and Reiss 1989). Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be n i.i.d. copies of a Gaussian random vector with zero-mean and unit variance, and $\rho_{i,j;n}$ be the pairwise correlation

between i -th and j -th components. Assume that, for all $i, j \in I$ and as n tends to infinity, $\log n(1 - \rho_{i,j;n}) \rightarrow \lambda_{i,j}^2 \in [0, \infty)$. Then, the exponent function is

$$V(\mathbf{y}) = \sum_{j=1}^d \frac{1}{y_j} \Phi_{d-1} \left\{ \left(\lambda_{i,j} + \frac{\log y_i/y_j}{2\lambda_{i,j}} \right)_{i \in I_j}; \bar{\Lambda}_j \right\}, \quad \mathbf{y} \in \mathbb{R}_+^d, \quad (3.1)$$

where Φ_{d-1} is $d - 1$ dimensional Gaussian distribution with partial correlation $\bar{\Lambda}_j$. The elements of the matrix $\bar{\Lambda}_j$ are $\lambda_{k,i;j} = (\lambda_{k,j}^2 + \lambda_{i,j}^2 - \lambda_{k,i}^2)/(2\lambda_{k,j}\lambda_{i,j})$, for all $j \in I$ and $k, i \in I_j$. The dependence between i -th and j -th elements of the vector of extremes is controlled by the corresponding parameter $\lambda_{i,j}$, $i, j \in I$. The case of complete dependence is achieved when $\lambda_{i,j} = 0$. Small and equal values of the parameters result in strong dependence among all the variables. The vector of extremes become independent as $\lambda_{ij} \rightarrow \infty$. Asymmetries arises when some parameter values are different and, in this case, the mass tends to concentrate around the vertices and edges associated to the smaller parameter values.

Observe that, in this case, the angular measure may concentrate mass on the interior of the simplex. Engelke et al. (2015) show that the angular density is

$$h(\mathbf{w}) = \phi_{d-1} \left\{ \left(\lambda_{i,1} + \frac{\log w_i/w_1}{2\lambda_{i,1}} \right)_{i \in I_1}; \bar{\Lambda}_1 \right\} \left\{ w_1^2 \prod_{i=2}^d (w_i 2\lambda_{i,1}) \right\}^{-1}, \quad \mathbf{w} \in \mathbb{W},$$

where ϕ_{d-1} is the $d - 1$ dimensional Gaussian density with partial correlation matrix $\bar{\Lambda}_1$. The Pickands dependence function is obtained from the exponent dependence function (3.1),

$$A(\mathbf{w}) = \sum_{j=1}^d w_j \Phi_{d-1} \left\{ \left(\lambda_{i,j} + \frac{\log w_i/w_j}{2\lambda_{i,j}} \right)_{i \in I_j}; \bar{\Lambda}_j \right\}, \quad \mathbf{w} \in \mathcal{S}_d. \quad (3.2)$$

Extremal skew- t model

The extremal skew- t model was introduced by Padoan (2011). It is the skewed extension of the extremal- t model described by Nikoloulopoulos et al. (2009) and it is more flexible than the Hüsler-Reiss. Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be n i.i.d. copies of a standard skew- t random vector with location parameter μ , scale parameter $\Omega = (\omega_{i,j})_{i,j < d}$, shape parameter $\alpha = (\alpha_1, \dots, \alpha_d)$ and $\nu > 0$ degrees of freedom (d.f.). Then, the exponent function of the limiting distribution of $\mathbf{M}_n/\mathbf{a}_n$ for $n \rightarrow \infty$, where $\mathbf{a}_n = (a_n \dots, a_n)$ is a vector of positive sequences (see eq. (4) in Padoan 2011), is

$$V(\mathbf{y}) = \sum_{j=1}^d (y_j)^{-\nu} T_{d-1} \left\{ \left[\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}} \left(\frac{\bar{y}_i}{\bar{y}_j} - \omega_{i,j} \right) \right]_{i \in I_j}; \mu, \Omega, \alpha_j, \tau_j, \nu + 1 \right\}, \quad (3.3)$$

for all $\mathbf{y} \in \mathbb{R}_+^d$, where T_{d-1} is a $d - 1$ dimensional extended Student- t with the parameter set and \bar{y} is the transformation defined in Padoan (2011, Theorem 2). Essentially, the exponent function depends on the parameters $\omega_{i,j}, \alpha_j$ and ν . The $\omega_{i,j}$ control the level of dependence between the i -th and j -th components of the random vector. The greater $\omega_{i,j}$, the stronger the dependence. The d.f., ν , is a global parameter which controls the overall level of dependence in the model. The smaller the d.f., the stronger the dependence among all components. The α_j parameters also drive the level of dependence among the components but in a more flexible way. With $\alpha_j = 0$, for all $j = 1, \dots, d$, the model reduces to the extremal- t distribution (Nikoloulopoulos et al. 2009).

The Hüsler-Reiss model is a special case of the extremal- t . At the same time, with additional conditions on the shape parameters, the limiting case of the extremal skew- t model is the skewed version of the Hüsler-Reiss model.

In this case, the Pickands dependence function is given by,

$$A(\mathbf{w}) = \sum_{j=1}^d w_j T_{d-1} \left\{ \sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}} \left[\left(\frac{\bar{w}_i}{\bar{w}_j} \right)^{1/\nu} - \omega_{i,j} \right]_{i \in I_j}; \mu, \Omega, \alpha_j, \tau_j, \nu + 1 \right\},$$

with $\mathbf{w} \in \mathcal{S}_d$.

Tilted Dirichlet model

When modelling the extremal dependence, the parametric model chosen should be flexible enough, meaning that its angular measure should place mass, not only on the interior of the simplex, but also on its vertices and edges. Comparing to the asymmetric logistic model, the tilted Dirichlet model is more parsimonious in terms of parameter, though they do not have a clear interpretation. It has been introduced by Coles and Tawn (1991). Consider a continuous function $h' : \mathcal{S}_d \rightarrow [0, \infty)$ such that $m_j = \int_{\mathbb{T}_d} v_j h'(\mathbf{v}) d\mathbf{v} < \infty$ for all $j \in I$. Then, the function

$$h(\mathbf{w}) = d^{-1} (m_1 w_1 + \dots + m_d w_d)^{-(d+1)} h' \{ \mathbf{m}\mathbf{w} / (m_1 w_1 + \dots + m_d w_d) \}, \quad \mathbf{w} \in \mathcal{S}_d$$

is a valid angular density as it satisfies the first moment condition (2.9), its mass is centred at $(1/d, \dots, 1/d)$ and it integrates to one. An example is provided, if h' is the density of the Dirichlet distribution, where the resulting angular density is equal to the following,

$$h(\mathbf{w}) = \frac{\Gamma(\sum_{j \in I} \alpha_j + 1)}{d (\sum_{j \in I} \alpha_j w_j)^{d+1}} \prod_{j=1}^d \frac{\alpha_j}{\Gamma(\alpha_j)} \left(\frac{\alpha_j w_j}{\sum_{j \in I} \alpha_j w_j} \right)^{\alpha_j - 1}, \quad \mathbf{w} \in \mathcal{S}_d, \quad (3.4)$$

with parameter set $\{\alpha_j > 0\}_{j \in I}$. The particular case of symmetry is obtained when $\alpha_1 = \dots = \alpha_d$. The case of complete independence is obtained with $\alpha_j \rightarrow 0$, for all $j \in I$. When all values of $\boldsymbol{\alpha}$ are smaller or equal to 1, the model yields to the case

of weak dependence, as the mass concentrates on the vertices of the simplex. When all values of the parameters are greater or equal to 1, the mass concentrates in the centre of the simplex leading to strong dependence. Then, $\alpha_j \rightarrow \infty$ defines the case of complete dependence.

In this case, the exponent function and its Pickands dependence function, can not be analytically computed, nonetheless they can still be evaluated numerically.

3.2 A Review of Nonparametric Estimators

Although computationally and mathematically elegant, parametric models may force the dependence structure in spite of their flexibility. The general dependence of multivariate max-stable distributions can not be parametrised (e.g. [de Haan and Ferreira, 2006](#), Ch. 6). Consequently, several inferential methods have been proposed for modelling the extremal dependence. For example, if \mathcal{A} is the family of functions A satisfying conditions (C1)-(C2), there is no parametric representation of A flexible enough to represent the entire class \mathcal{A} . For this reason, several nonparametric estimators of the Pickands have been proposed. Some examples are found in [Pickands \(1981\)](#), [Capéraà et al. \(1997\)](#), [Hall and Tajvidi \(2000\)](#), [Bücher et al. \(2011\)](#), [Zhang et al. \(2008\)](#), [Gudendorf and Segers \(2011, 2012\)](#) and [Berghaus et al. \(2013\)](#) to name a few. Note that a generic function has to be a member of the family \mathcal{A} in order to be a valid Pickands dependence function ([Falk, Hüsler, and Reiss, 2010](#), Ch. 4). The converse is not true, however, as conditions (C1)-(C2) are not sufficient and the resulting estimate of $G_0(\mathbf{y})$ in (2.5) may not be a proper distribution. For instance, an *ad hoc* example of a function that satisfies (C1)-(C2) and is not a valid Pickands dependence function is discussed [Beirlant et al. \(2004, p.257\)](#). A characterization of the class of stable tail dependence functions has been given in [Ressel \(2013\)](#) and the necessary and sufficient conditions that a multivariate function must meet in order to be a tail dependence function have been shown.

A possibility may be to rephrase those conditions for the Pickands dependence function and implement them through a statistical procedure. However, so far it has not been possible to show that the nonparametric estimators proposed in the literature, especially in their multivariate version, fulfil, without further adjustments, the conditions (C1)-(C2). So, for practical applications it is already a considerable result to provide an estimator that meets such conditions. In the next Chapter we propose a method to non-parametrically estimate the multivariate Pickands dependence function. In the bivariate case, a fast-to-compute and easy-to-interpret estimator based on a type of madogram was introduced by [Naveau et al. \(2009\)](#) but it has two drawbacks. It is only defined for the bivariate case and it does not necessarily satisfy the conditions (C1)-(C2). Our first task is to propose a new type of madogram for a multivariate setting with $d \geq 2$ (see also the

work of [Fonseca et al. \(2013\)](#)). Still, conditions (C1)-(C2) are not necessarily checked for this extension and another aim of Chapter 4 is to “regularise” it by projecting it onto \mathcal{A} through a Bernstein polynomial based method. Hence, by construction, the new regularized estimator stemming from a multivariate madogram estimate will belong to \mathcal{A} . In the bivariate case, the regularization strategy has been initially investigated by [Pickands \(1981\)](#) who proposed, in order to satisfy the convexity condition, to consider the greatest convex minorant (see also [Capéraà et al. 1997](#), [Hall and Tajvidi 2000](#)). [Smith et al. \(1990\)](#) proposed to modify a first guess estimator \widehat{A}_n using kernel methods and [Hall and Tajvidi \(2000\)](#) to approximate \widehat{A}_n using constrained smoothing splines. However, as discussed by [Fils-Villetard et al. \(2008\)](#), the impact of these adjustments on the asymptotic properties of the estimator changes from one case to another, while a general result is unknown. The projection estimator approach introduced by [Fils-Villetard et al. \(2008\)](#) provides a general theoretical framework that is suitable for estimating the Pickands dependence function and it has recently been extended in the multivariate case by [Gudendorf and Segers \(2012\)](#). In any case, only approximate projection estimators belonging to finite-dimensional subsets $\mathcal{A}_k \subseteq \mathcal{A}$, which are increasingly accurate with increasing k , can be obtained. Typically, approximate projection estimators of the Pickands dependence function have been constructed through piecewise linear functions ([Fils-Villetard et al. 2008](#); [Gudendorf and Segers 2012](#)). Although the Pickands is a smooth function, these estimators may provide non-smooth estimates, furthermore they can be time consuming for large sample sizes and challenging to implement in high dimensions (e.g. $d > 3$), since multivariate integrals are involved in the estimation procedure ([Gudendorf and Segers 2012](#)). To bypass these computational hurdles, our strategy is to project any Pickands dependence function estimators via a sequence of restricted Bernstein polynomials ([Lorentz, 1986](#); [Sauer, 1991](#)). The benefit is that the Pickands is accurately approximated while preserving its shape, indeed the estimator is guaranteed to fulfil the desired shape restrictions everywhere in the domain. In virtue of their optimal shape restriction properties ([Carnicer and Peña, 1993](#)), Bernstein polynomials are suitable for nonparametric curve estimation (e.g. [Petrone 1999](#); [Chang et al. 2005](#)) and shape-preserving regressions ([Wang and Ghosh 2012](#)).

In Chapter 4 we show their usefulness when estimating the Pickands dependence function. Our inference approach displays nice asymptotic properties, it is computationally efficient and it is feasible to implement in moderately high dimensions with an acceptable computational cost, and the uncertainty of the estimates can be easily assessed through bootstrap confidence bands. We stress that with our approach, in practice, we are able to nonparametrically estimate the Pickands dependence function, for instance, up to the dimension seven. This, it seems, has not been done before in a real application.

Let \mathbf{X} be a max-stable distributed random vector and let \mathbf{X}_m , $m = 1, \dots, n$, be independent and identically distributed (i.i.d.) replicates of it. For comparison purposes, we

briefly discuss some well known estimators of the Pickands dependence function (see [Gudendorf and Segers 2011](#) for a review). Let $Y_{m,i} = -\log F_i(X_{m,i})$ be standard exponential random variables for $m = 1, \dots, n$ and $i = 1, \dots, d$. Define for each $m = 1, \dots, n$,

$$\tilde{Y}_m(\mathbf{w}) = \bigwedge_{i=1, \dots, d} \left(\frac{Y_{m,i}}{w_i} \right), \quad \mathbf{w} \in \mathcal{S}_d, \quad (3.5)$$

Then, the multivariate Pickands statistic ([Pickands 1981](#)), referred to as P for brevity, is

$$A_n^P(\mathbf{w}) = n / \sum_{m=1}^n \tilde{Y}_m(\mathbf{w}). \quad (3.6)$$

Closely related to (3.6) we have that the multivariate version of the statistic described in [Hall and Tajvidi \(2000\)](#), referred to as HT for brevity, is

$$A_n^{HT}(\mathbf{w}) = A_n^P(\mathbf{w}) / A_n^P(\mathbf{e}_i), \quad (3.7)$$

where $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$ for $i = 1, \dots, d - 1$. In the bivariate case, the function A_n^{HT} satisfies condition C2, but not necessarily C1. In the multivariate case it has not been proved that the estimator satisfies condition C2. Lastly, the multivariate version of the statistic proposed by [Capéraà et al. \(1997\)](#), referred to as CFG for brevity, is

$$A_n^{CFG}(\mathbf{w}) = \exp \left(-\frac{1}{n} \sum_{m=1}^n \log \tilde{Y}_m(\mathbf{w}) - \gamma \right), \quad (3.8)$$

where $\gamma = 0.5772\dots$ is the Euler–Mascheroni constant. Once again, the function A_n^{CFG} does not necessarily satisfy the conditions (C1)–(C2). Asymptotic properties of these statistics have been derived by [Gudendorf and Segers \(2011\)](#). By replacing in (3.5) the marginal distributions F_1, \dots, F_d with the empirical distribution functions

$$F_{n,i}(x) = \frac{1}{n} \sum_{m=1}^n \mathbb{I}(X_{m,i} \leq x), \quad i = 1, \dots, d, \quad (3.9)$$

where $\mathbb{I}(E)$ is the indicator function of the event E , the resulting expression (3.6)–(3.8) provide the definitions of the P, HT and CFG estimators which we denote by \hat{A}_n^P , \hat{A}_n^{HT} and \hat{A}_n^{CFG} respectively. For these estimators, weak convergence result has been shown by [Gudendorf and Segers \(2012\)](#).

Also likelihood and Bayesian inferential methods have been widely investigated. Examples of likelihood based methods are the approximate likelihood (e.g. [Coles and Tawn 1994](#); [Cooley et al. 2010](#)) and the composite likelihood (e.g. [Padoan et al. 2010](#); [Davison and Gholamrezaee 2012](#)). Examples of Bayesian techniques are found in [Apputhurai and Stephenson \(2011\)](#), [Sabourin et al. \(2013\)](#), [Sabourin and Naveau \(2014\)](#).

Tesi di dottorato "Non-Parametric Inference and Simulation in Extreme Value Theory"
di MARCON GIULIA

discussa presso Università Commerciale Luigi Bocconi-Milano nell'anno 2016

La tesi è tutelata dalla normativa sul diritto d'autore (Legge 22 aprile 1941, n.633 e successive integrazioni e modifiche).

Sono comunque fatti salvi i diritti dell'università Commerciale Luigi Bocconi di riproduzione per scopi di ricerca e didattici, con citazione della fonte.

Chapter 4

Nonparametric Estimation of a Pickands Dependence Function

*It seems that the rivers know the theory.
It only remains to convince the engineers
of the validity of this analysis*

(E. Gumbel)

This Chapter addresses the problem of estimating the multivariate Pickands dependence function from observations following a multivariate extreme value distribution with known and unknown margins. Such a problem has been widely considered and different approaches exist already, particularly in the bivariate setting. The aim of this Chapter is two-fold: in a first step, a new nonparametric estimator is introduced in Section 4.1, which is a natural multivariate generalization of the bivariate madogram estimator introduced by Naveau et al. (2009). The associated multivariate madogram estimator, denoted by \hat{A}_n does not belong to the space \mathcal{A} of functions satisfying conditions (C1)–(C2). Therefore we “regularize” it, in the sense that \hat{A}_n is projected onto \mathcal{A} through a Bernstein polynomials based method. Accordingly the projection estimator \tilde{A}_n belongs to \mathcal{A} as discussed in Section 4.2. Thus, we obtain a nonparametric estimator that satisfies the shape constraints shared by dependence functions (convexity, within some lower and upper bounds, and with specific values at the boundary points of the simplex). This construction exploits the Bernstein-Bézier polynomial representation of the dependence function. More precisely, it is obtained as the projection of an original estimator (whatever it is) on constrained multivariate Bernstein-Bézier polynomial families, combined with an approximation based on sieves. One of the advantage of using the polynomials in Bernstein form is that, accounting for the variability of the estimator, it is also possible to compute simultaneous credible bands (see Section 4.2). The large-sample theory for both estimators is developed in Section 4.3 and their finite-sample performance is evaluated

with a simulation study (see Section 4.4).

4.1 Madogram Estimator

The multivariate statistic introduced in this Section is an extension of the bivariate madogram statistic discussed by Naveau et al. (2009) (see also Fonseca et al. (2013)). Let \mathbf{X} be a random vector with continuous marginal distribution functions F_1, \dots, F_d and whose copula C is an extreme-value copula with stable tail dependence function ℓ and Pickands dependence function A ; see Chapter 2 for details and properties.

Definition 4.1. For $\mathbf{w} \in \mathcal{S}_d$, the multivariate \mathbf{w} -madogram (MD), denoted by $\nu(\mathbf{w})$, is defined as the expected distance between the componentwise maximum and the componentwise mean of the variables $\{F_1^{1/w_1}(X_1), \dots, F_d^{1/w_d}(X_d)\}$, that is

$$\nu(\mathbf{w}) = \mathbb{E} \left(\bigvee_{i=1, \dots, d} \{F_i^{1/w_i}(X_i)\} - \frac{1}{d} \sum_{i=1, \dots, d} F_i^{1/w_i}(X_i) \right), \quad (4.1)$$

For $w_i = 0$ and $0 < u < 1$, we set $u^{1/w_i} = 0$, by convention.

Proposition 4.1. Let $\mathbf{U} \in [0, \infty)^d$ be a random vector with dependent components. Assume that $\mathbb{E}\{U_i\} < \infty$ for all $i \in I$. Set

$$\bar{U}_d = \frac{1}{d} \sum_{i \in I} U_i, \quad M_d = \max_{i \in I} \{U_i\}.$$

Define the mapping $\rho : [0, \infty)^d \rightarrow [0, \infty)$ by

$$\rho(\mathbf{U}) = \mathbb{E}[M_d - \bar{U}_d], \quad (4.2)$$

Then, the mapping ρ has the following properties:

1. $\rho(\mathbf{U}) \geq 0$ and $\rho(\mathbf{U}) = 0$ iff $U_i = U_j$, for all $i, j \in I$, almost surely;
2. $\rho(U_1, \dots, U_d) = \rho(U_{i_1}, \dots, U_{i_d})$ for any permutation i_1, \dots, i_d on the set I ;
3. for every variable Y with finite mean, we have

$$\rho(\mathbf{U}) \leq \sum_{i \in I} \rho(Y, U_i) + \left(\frac{d}{2} - 1\right) \left(\mathbb{E}[Y] + \frac{1}{d} \sum_{i \in I} \mathbb{E}[U_i]\right) \quad (4.3)$$

and

$$\rho(\mathbf{U}) \leq \rho(\mathbf{U} + Y) + \sum_{i \in I} \mathbb{E}[\max\{Y, U_i\}] - \mathbb{E}[\max\{\mathbf{U} + Y\}]. \quad (4.4)$$

Proposition 4.2. *Let \mathbf{X} be a max-stable random vector with exponent V and Pickands dependence function A , then*

$$\nu(\mathbf{w}) = \frac{V(1/w_1, \dots, 1/w_d)}{1 + V(1/w_1, \dots, 1/w_d)} - c(\mathbf{w}),$$

where $c(\mathbf{w}) = d^{-1} \sum_{i=1}^d w_i / (1 + w_i)$. It follows that

$$V(1/w_1, \dots, 1/w_d) = \frac{\nu(\mathbf{w}) + c(\mathbf{w})}{1 - \nu(\mathbf{w}) - c(\mathbf{w})},$$

and

$$A(\mathbf{w}) = \frac{\nu(\mathbf{w}) + c(\mathbf{w})}{1 - \nu(\mathbf{w}) - c(\mathbf{w})}. \quad (4.5)$$

Remark 4.1. When $d = 2$, expression (4.3) is the triangular inequality and if (4.2) is applied to $(U_1, U_2) := \{F^\lambda(X_1), F^{1-\lambda}(X_2)\}$, for any $\lambda \in (0, 1)$, then the λ -madogram is obtained (Naveau et al. 2009).

Remark 4.2. One advantage of the madogram as defined in (4.1) is that it can be interpreted as an L1-distance. Specifically, let $\rho(u, v) = \mathbb{E}|u - v|$, where

$$u = \bigvee_{i=1, \dots, d} \left\{ F_i^{1/w_i}(X_i) \right\}, \quad v = \frac{1}{d} \sum_{i=1, \dots, d} F_i^{1/w_i}(X_i).$$

Then, $\rho(u, v) \geq 0$ measures the distance between u and v , that is between the maximum and the mean of the elements $\{F_1^{1/w_1}(X_1), \dots, F_d^{1/w_d}(X_d)\}$. If $w_1 = \dots = w_d = 1/d$, then the L_1 distance is zero and this happens if and only if all components $F_i(X_i)$ are equal with probability one, that is, in case of complete dependence. In other words, (4.1) tells us how far away \mathbf{X} is from the complete dependence case.

In the bivariate case, Definition 4.1 is slightly different from the one proposed by Naveau et al. (2009). Here, we use the vector $\{F_1^{1/w_1}(X_1), F_2^{1/w_2}(X_2)\}$ instead of the couple $\{F_1^{w_1}(X_1), F_2^{w_2}(X_2)\}$. This new version has the advantage that the sample equivalent of (4.5) will automatically satisfy condition (C2), which is not necessarily for the original estimator by Naveau et al. (2009). We first assume that the marginal distributions F_1, \dots, F_d are known. Equation (4.1) suggests the statistic

$$\nu_n(\mathbf{w}) = \frac{1}{n} \sum_{m=1}^n \left(\bigvee_{i=1, \dots, d} \left\{ F_i^{1/w_i}(X_{m,i}) \right\} - \frac{1}{d} \sum_{i=1}^d F_i^{1/w_i}(X_{m,i}) \right), \quad (4.6)$$

and the Pickands dependence function can then be then estimated through

$$A_n^{\text{MD}}(\mathbf{w}) = \frac{\nu_n(\mathbf{w}) + c(\mathbf{w})}{1 - \nu_n(\mathbf{w}) - c(\mathbf{w})}, \quad \mathbf{w} \in \mathcal{S}_d. \quad (4.7)$$

Now, if the marginal distributions F_1, \dots, F_d are unknown, they can be estimated via the empirical distribution functions given by expression (3.9). Replacing F_i by $F_{n,i}$ in Equation (4.6) yields our proposed nonparametric estimators $\hat{\nu}_n$ and \hat{A}_n^{MD} of the multivariate

madogram and of the Pickands dependence function, respectively. Other estimators of the marginal distributions could be used instead. However, the use of the empirical distribution functions requires minimal assumptions and yields an estimator for A which is invariant under monotone transformations. The condition (C1) is not necessarily satisfied by $\widehat{A}_n^{\text{MD}}$. To ensure this condition, we propose a projection method based on Bernstein polynomials.

4.2 Estimation based on Bernstein Polynomials

4.2.1 Bernstein Polynomials on the Simplex \mathcal{S}_d

Multivariate Bernstein polynomials, defined on the d -dimensional hypercube and simplex have been widely discussed in mathematics and statistics, see for example [Ditzian \(1986\)](#) and the references therein, and for a comprehensive description see [Petrone \(2004\)](#). Here, our focus is on any bounded function $f(\mathbf{w})$ defined on the simplex \mathcal{S}_d . In the univariate case, the relevant shape features of the original function are preserved by its Bernstein polynomials approximation. For higher dimensions, shape properties like convexity may no longer be retained, e.g. [Sauer \(1991\)](#). The Bernstein-Bézier polynomials ([Sauer 1991](#)), under appropriate conditions (see [Li 2011](#), [Lai 1993](#)), solve this issue and preserve the desired shape properties. They are defined as follows.

Definition 4.2. Let $d \geq 2$ denote the dimension. For positive integer k , let Γ_k be the set of multi-indices $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{d-1}) \in \{0, 1, \dots, k\}^{d-1}$ such that $\alpha_1 + \dots + \alpha_{d-1} \leq k$. Define the Bernstein polynomial basis $b_{\boldsymbol{\alpha}}(\cdot; k)$ of degree k , on \mathcal{S}_d , by

$$b_{\boldsymbol{\alpha}}(\mathbf{w}; k) = \binom{k}{\boldsymbol{\alpha}} \mathbf{w}^{\boldsymbol{\alpha}}, \quad \mathbf{w} \in \mathcal{S}_d \quad (4.8)$$

where

$$\binom{k}{\boldsymbol{\alpha}} = \frac{k!}{\alpha_1! \dots \alpha_d!}, \quad \mathbf{w}^{\boldsymbol{\alpha}} = w_1^{\alpha_1} \dots w_d^{\alpha_d}.$$

The k -th degree Bernstein polynomial associated to A is defined as

$$B_A(\mathbf{w}; k) = \sum_{\boldsymbol{\alpha} \in \Gamma_k} A(\boldsymbol{\alpha}/k) b_{\boldsymbol{\alpha}}(\mathbf{w}; k), \quad \mathbf{w} \in \mathcal{S}_d. \quad (4.9)$$

Remark 4.3. The cardinality of Γ_k is equal to the number of multi-indices $\boldsymbol{\alpha} \in \{0, 1, \dots, k\}^d$ such that $\alpha_1 + \dots + \alpha_d = k$; just set $\alpha_d = k - \alpha_1 - \dots - \alpha_{d-1}$. Replacing each α_j by $\alpha_j + 1$, we find that the number of such multi-indices is also equal to the number of compositions of the integer $k + d$ into d positive integer parts. The number of such compositions is equal to

$$p_k = \binom{k + d - 1}{d - 1}, \quad (4.10)$$

and so is the cardinality of Γ_k .

Proposition 4.3. *For every Pickands dependence function A and $k = 1, 2, \dots$,*

$$\sup_{\mathbf{w} \in \mathcal{S}_d} |B_A(\mathbf{w}; k) - A(\mathbf{w})| \leq \frac{d}{2\sqrt{k}}.$$

Furthermore, if the Pickands dependence function A has bounded and continuous second-order partial derivatives, then for all k

$$\sup_{w \in \mathcal{S}_d} |B_A(\mathbf{w}; k) - A(\mathbf{w})| \leq \frac{cd}{4k}$$

for some positive constant c depending on A .

The aim of this work is to estimate the Pickands dependence function, thus we want to estimate the vector $A(\boldsymbol{\alpha}/k)$, $\boldsymbol{\alpha} \in \Gamma_k$, in the Bernstein representation (4.9), which we denote by $\boldsymbol{\beta} = \{\beta_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \Gamma_k\} \in [0; 1]^{p_k}$.

The family of Bernstein–Bézier polynomials of degree k is defined as the set

$$\mathcal{B}_k = \left\{ \sum_{\boldsymbol{\alpha} \in \Gamma_k} \beta_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\cdot; k) : \boldsymbol{\beta} \in [0, 1]^{p_k} \right\}.$$

For $\mathbf{w} \in \mathcal{S}_d$, let $\mathbf{b}_k(\mathbf{w})$ be the row vector $(b_{\boldsymbol{\alpha}}(\mathbf{w}; k), \boldsymbol{\alpha} \in \{0, 1, \dots, k\}^d : \alpha_1 + \dots + \alpha_d = k)$. In matrix notation, we have

$$B(\mathbf{w}; k) = \sum_{\boldsymbol{\alpha} \in \Gamma_k} \beta_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\mathbf{w}; k) = \mathbf{b}_k(\mathbf{w}) \boldsymbol{\beta}, \quad (4.11)$$

where $\boldsymbol{\beta}$ is viewed as a column vector.

Remark 4.4. If $d = 2$, then $\boldsymbol{\alpha} = (j)$, $j = 0, \dots, k$. Therefore the Bernstein basis polynomial of degree k , on \mathcal{S}_2 , simplifies to

$$b_j(w; k) = \binom{k}{j} w^j (1-w)^{k-j}, \quad w \in [0, 1]$$

and the polynomial representation of A becomes simply

$$B(w; k) = \sum_{j=0}^k \beta_j b_j(w; k) \quad w \in [0, 1],$$

with $\boldsymbol{\beta} = \{\beta_j, j = 0, \dots, k\} \in [0, 1]^{k+1}$.

Remark 4.5. If A is a polynomial in Bernstein form, then it is differentiable and its derivatives can also be expressed in Bernstein form as:

$$\begin{aligned} A'(w) &= k \sum_{j=0}^{k-1} (\beta_{j+1} - \beta_j) b_j(w; k-1) \\ A''(w) &= k(k-1) \sum_{j=0}^{k-2} (\beta_{j+2} - 2\beta_{j+1} + \beta_j) b_j(w; k-2). \end{aligned}$$

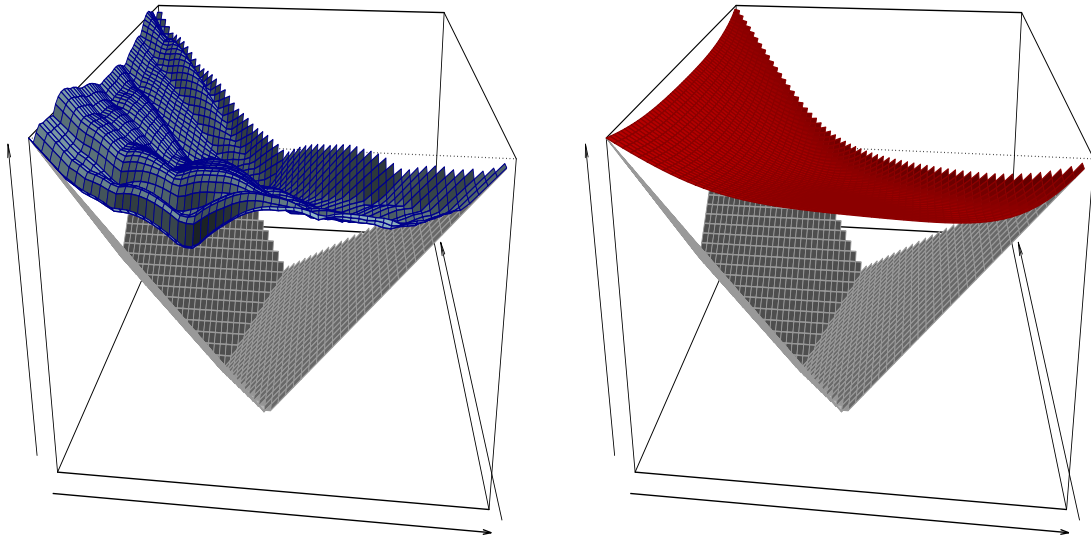


Figure 4.1: Visual illustration of the gain in implementing the Bernstein-Bézier projection of \widehat{A}_n

Note that the coefficients β_j , $j = 0, \dots, k$ are the same as in $A(w)$. Another important relation is given by the following remark.

Remark 4.6. For $j = 0, \dots, k$, $w \in [0, 1]$, the j -th Bernstein basis polynomial of degree $k \in \mathbb{N}$, can be rewritten in terms of a Beta density with parameters $\alpha_1 = j + 1$ and $\alpha_2 = k - j + 1$, as follows:

$$\begin{aligned}
 b_j(w; k) &= \frac{k!}{j! (k-j)!} w^j (1-w)^{k-j} \\
 &= \frac{\Gamma(k+1)}{\Gamma(j+1) \Gamma(k-j+1)} w^j (1-w)^{k-j} \\
 &= \frac{(k+1)^{-1} \Gamma(k+2)}{\Gamma(j+1) \Gamma(k-j+1)} w^j (1-w)^{k-j} \\
 &= \frac{1}{(k+1)} \frac{1}{B(j+1, k-j+1)} w^{(j+1)-1} (1-w)^{(k-j+1)-1} \\
 &= \frac{1}{(k+1)} \text{Be}(w; \alpha_1, \alpha_2).
 \end{aligned}$$

4.2.2 Shape Preserving Estimator

In this section, we describe how to use Bernstein-Bézier polynomials in order to obtain a projection estimator (Fils-Villetard et al. 2008) that satisfies conditions (C1)–(C2). The general idea is illustrated in Figure 4.1 where the madogram, on the left panel, clearly does not satisfy the convexity constraint. We want to use it as a starting point to obtain

a new function that fulfils all the required conditions (see right panel). The procedure can be applied to different functions, here we focus on all $A \in \mathcal{A}$. Given a pilot estimator, say \widehat{A}_n , we seek approximate solutions to the constrained optimization problem

$$\widetilde{A}_n = \arg \min_{A \in \mathcal{A}} \|\widehat{A}_n - A\|_2.$$

There is no closed-form solution to the above equation, and so an approximation based on the sieves method is explored. Consider a sequence $\mathcal{A}_k \subseteq \mathcal{A}$ of constrained multivariate Bernstein–Bézier polynomial families on \mathcal{S}_d given by

$$\mathcal{A}_k = \{\mathbf{w} \mapsto B(\mathbf{w}; k) = \mathbf{b}_k(\mathbf{w})\boldsymbol{\beta}_k : \boldsymbol{\beta}_k \in [0, 1]^{p_k} \text{ such that } \mathbf{R}_k\boldsymbol{\beta}_k \geq \mathbf{r}_k\}. \quad (4.12)$$

Here, $\mathbf{R}_k = [\mathbf{R}_k^{(1)}, \mathbf{R}_k^{(2)}, \mathbf{R}_k^{(3)}]^\top$ and $\mathbf{r}_k = [\mathbf{r}_k^{(1)}, \mathbf{r}_k^{(2)}, \mathbf{r}_k^{(3)}]^\top$ are a $(q \times p_k)$ full row rank matrix and a $(q \times 1)$ vector respectively such that the constraint $\mathbf{R}_k\boldsymbol{\beta}_k \geq \mathbf{r}_k$ on the coefficient vector $\boldsymbol{\beta}_k$ ¹ ensures that each member of \mathcal{A}_k satisfies (C1)–(C2). Details for deriving the block matrices and vectors of constraints are provided below.

R1) A sufficient condition to guarantee that the mapping $\mathbf{w} \mapsto B(\mathbf{w}; k)$ on \mathcal{S}_d is convex is that its Hessian matrix be positive semi-definite. In order to enforce the latter, we resort to the following relations. For $s \neq r \in \{0, \dots, d-1\}$ and two vectors \mathbf{v}_r and \mathbf{v}_s , where $\mathbf{v}_r = \mathbf{0}$ if $r = 0$ and $\mathbf{v}_r = \mathbf{e}_r$ if $r > 0$ with \mathbf{e}_r the canonical unit vector (analogously for \mathbf{v}_s), the directional derivative of B with respect to the direction $\overrightarrow{\mathbf{v}_r\mathbf{v}_s}$ is given by

$$D_{\mathbf{v}_s - \mathbf{v}_r} B(\mathbf{w}; k) = k \sum_{\boldsymbol{\alpha} \in \Gamma_{k-1}} \Delta_{s,r} \beta_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\mathbf{w}; k-1), \quad \mathbf{w} \in \mathcal{S}_d$$

where $\Delta_{s,r} \beta_{\boldsymbol{\alpha}} = (\beta_{\boldsymbol{\alpha} + \mathbf{v}_s} - \beta_{\boldsymbol{\alpha} + \mathbf{v}_r})$. The second directional derivative of B with respect to the directions $\overrightarrow{\mathbf{v}_r\mathbf{v}_s}$ and $\overrightarrow{\mathbf{v}_r\mathbf{v}_t}$ is

$$D'_{\mathbf{v}_s - \mathbf{v}_r, \mathbf{v}_t - \mathbf{v}_r} B(\mathbf{w}; k) = k(k-1) \sum_{\boldsymbol{\alpha} \in \Gamma_{k-2}} \Delta_{t,r} \Delta_{s,r} \beta_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\mathbf{w}; k-2), \quad \mathbf{w} \in \mathcal{S}_d.$$

Then, the Hessian matrix of $B(\mathbf{w}; k)$, $\mathbf{w} \in \mathcal{S}_d$, is $H_B = [D'_{\mathbf{v}_s, \mathbf{v}_t} B(\mathbf{w}; k)]_{s,t \in \{1, \dots, d-1\}, r=0}$, and it can be written as

$$H_B = k(k-1) \sum_{\boldsymbol{\alpha} \in \Gamma_{k-2}} \Sigma_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\mathbf{w}; k-2), \quad \mathbf{w} \in \mathcal{S}_d,$$

where, for all $\boldsymbol{\alpha} \in \Gamma_{k-2}$, $\Sigma_{\boldsymbol{\alpha}}$ is the symmetric $(d-1) \times (d-1)$ matrix

$$\Sigma_{\boldsymbol{\alpha}} = \begin{pmatrix} \Delta_{1,0}^2 \beta_{\boldsymbol{\alpha}} & \Delta_{1,0} \Delta_{2,0} \beta_{\boldsymbol{\alpha}} & \cdots & \cdots & \Delta_{1,0} \Delta_{d-1,0} \beta_{\boldsymbol{\alpha}} \\ & \Delta_{2,0}^2 \beta_{\boldsymbol{\alpha}} & \Delta_{2,0} \Delta_{3,0} \beta_{\boldsymbol{\alpha}} & \cdots & \Delta_{2,0} \Delta_{d-1,0} \beta_{\boldsymbol{\alpha}} \\ & & \vdots & \vdots & \vdots \\ & & & & \Delta_{d-1,0}^2 \beta_{\boldsymbol{\alpha}} \end{pmatrix}.$$

¹In this section $\boldsymbol{\beta}_k \equiv \boldsymbol{\beta}$ has subscript k to stress its dependence to the polynomial order. However, it is removed in the following part to simplify the notation.

Remark 4.7. When $d = 3$, the Hessian matrix takes the following form:

$$H_B(\mathbf{w}) = \begin{bmatrix} B''_1(w; k) = \frac{\partial^2}{\partial w^2} B(w; k) & B''_{12}(w) = \frac{\partial^2}{\partial w \partial 1-w} B(w; k) \\ B''_{12}(w; k) = \frac{\partial^2}{\partial w \partial 1-w} B(w; k) & B''_2(w; k) = \frac{\partial^2}{\partial 1-w^2} B(w; k) \end{bmatrix}$$

where the entries of the matrix are the second partial derivatives of $B(w; k)$ with respect to w and $1 - w$. Then, we can write

$$H B(w; k) = [\nabla_{i,j,0}^2 B_A(w; k)]_{i,j \in \{1,2\}}$$

where

$$\nabla_{i,j,0}^2 B(w; k) = k(k-1) \sum_{j=0}^{k-2} \Delta_{i,0} \Delta_{j,0} \beta_{\alpha} b_{\alpha}(w; k-2), \quad bw \in [0, 1].$$

Notice that, when considering $i = j = 1, 2$, the expression corresponds to the second pure partial derivatives, while for $i \neq j$ it refers to the second mixed partial derivatives. Furthermore, it is known that Bernstein polynomials basis are always non negative, and $k \in \mathbb{N}$. Therefore, it remains to show that the matrix Q_l is positive semi-definite.

By the weak diagonal dominance criterion (Lai 1993) in order to guarantee that Σ_{α} is positive semi-definite, it is sufficient to check, for all $\alpha \in \Gamma_{k-2}$ and $i \in \{1, \dots, d-1\}$, the conditions

$$\Delta_{i,0}^2 \beta_{\alpha} - \sum_{j \neq i} |\Delta_{i,0} \Delta_{j,0} \beta_{\alpha}| \geq 0.$$

This can be synthesized in matrix form as $\mathbf{R}_k^{(1)} \beta_k \geq \mathbf{r}_k^{(1)}$ where $\mathbf{R}_k^{(1)}$ is a matrix of dimension $(p_{k-2}(d-1)2^{d-2} \times p_k)$ and $\mathbf{r}_k^{(1)}$ is the corresponding null vector. For example, with $d = 3$ and $k = 3$,

$$\mathbf{R}_3^{(1)} = \begin{pmatrix} 0 & 1 & 0 & 0 & -1 & -1 & 0 & 1 & 0 & 0 \\ 2 & -1 & 0 & 0 & -3 & 1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 2 & -3 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & 2 & -1 & 0 & 0 & -3 & 1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 2 & -3 & 1 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 2 & -1 & 0 & -3 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 2 & -3 & 1 & -1 & 1 & 0 \end{pmatrix}, \quad \mathbf{r}_3^{(1)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

R2) B satisfies the upper bound condition in (C2) if $\beta_{\alpha} = 1$ for the set of coefficients $\{\beta_{\alpha} : \alpha = \mathbf{0} \text{ or } \alpha = k \mathbf{e}_i, \forall i = 1, \dots, d-1\}$. Thus, the $(2d \times p)$ matrix and

$2d$ -dimensional vector of restrictions are equal to

$$\mathbf{R}_k^{(2)} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ -1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & \cdots & -1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & -1 & \cdots & 0 \end{pmatrix}, \quad \mathbf{r}_k^{(2)} = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \\ 1 \\ -1 \end{pmatrix}.$$

R3) B satisfies the lower bound condition in (C2) if the restrictions R1)-R2) hold and the following constraints are fulfilled. Specifically, for all $(i, j) \in \{0, \dots, d-1\}^2$, $i \neq j$, the first directional derivatives with respect to $\overrightarrow{\mathbf{v}_i \mathbf{v}_j}$, evaluated at the vertices of the simplex, are compared with the first directional derivatives of the planes $z_0 = 1$, $z_1 = w_1$, $z_2 = w_2$, \dots , $z_d = 1 - w_1 - w_2 - \dots - w_{d-1}$, with respect to the same directions. So, it is sufficient to check the conditions

$$D_{\mathbf{v}_i - \mathbf{v}_j} B(\mathbf{v}_j; k) > -1, \quad \forall (i, j) \in \{0, \dots, d-1\}^2, i \neq j.$$

As a consequence, it is sufficient to check the conditions $\beta_{\boldsymbol{\alpha}} > 1 - 1/k$ for the set of coefficients $\{\beta_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} = \mathbf{e}_i \text{ or } \boldsymbol{\alpha} = (k-1)\mathbf{e}_i \text{ or } \boldsymbol{\alpha} = (k-1)\mathbf{e}_i + \mathbf{e}_j, \forall i, j = 1, \dots, d-1, i \neq j\}$. This can be synthesized in matrix form as $\mathbf{R}_k^{(3)} \boldsymbol{\beta}_k \geq \mathbf{r}_k^{(3)}$ where $\mathbf{R}_k^{(3)}$ is a $(d(d-1) \times p_k)$ -dimensional matrix and $\mathbf{r}_k^{(3)}$ is the corresponding vector of $1 - 1/k$ vaules. For example, when $d = 3$ and $k = 3$, the constraint matrix is the following:

$$\mathbf{R}_3^{(3)} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{r}_3^{(3)} = \begin{pmatrix} 1-1/k \\ 1-1/k \\ 1-1/k \\ 1-1/k \\ 1-1/k \\ 1-1/k \end{pmatrix}.$$

The use of the third restriction is justified by the following result.

Proposition 4.4. *Let B_A be the polynomial (4.9). Assume that B_A is convex on the simplex and $B_A(\mathbf{v}_j; k) = 1$ for all $j \in \{0, \dots, d-1\}$. Then, for all $\mathbf{w} \in \mathcal{S}_d$*

$$B_A(\mathbf{w}; k) \geq \max(w_1, \dots, w_d) \quad \iff \quad D_{\mathbf{v}_i - \mathbf{v}_j} B_A(\mathbf{v}_j; k) \geq -1,$$

for all $(i, j) \in \{0, \dots, d-1\}^2$, $i \neq j$.

Remark 4.8. If $d = 2$ and $k = 3$, the matrix of all constraints, and its corresponding vector, simplify to the following (8×4) matrix,

$$\mathbf{R}_3 = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{r}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 1 \\ -1 \\ 2/3 \\ 2/3 \end{pmatrix}.$$

The approximate projection estimator is given by the solution of the equation

$$\tilde{A}_{n,k} = \arg \min_{B \in \mathcal{A}_k} \|\hat{A}_n - B\|_2. \quad (4.13)$$

In practice, the estimator $\tilde{A}_{n,k}$ is evaluated on a finite set of points $\{\mathbf{w}_q, q = 1, \dots, Q\}$, with $Q \in \mathbb{N}$ and $\mathbf{w}_q \in \mathcal{S}_d$. Therefore, the discretized version of the above solution is given by

$$\tilde{A}_{n,k}(\mathbf{w}_q) = \mathbf{b}_k(\mathbf{w}_q) \hat{\boldsymbol{\beta}}_k, \quad \mathbf{w}_q \in \mathcal{S}_d, \quad q = 1, \dots, Q, \quad (4.14)$$

where $\hat{\boldsymbol{\beta}}_k$ is the minimizer of the constrained least-squares problem

$$\hat{\boldsymbol{\beta}}_k = \arg \min_{\boldsymbol{\beta}_k \in [0,1]^{p_k} : \mathbf{R}_k \boldsymbol{\beta}_k \geq \mathbf{r}_k} \frac{1}{Q} \sum_{q=1}^Q (\mathbf{b}_k(\mathbf{w}_q) \boldsymbol{\beta}_k - \hat{A}_n(\mathbf{w}_q))^2.$$

This is a quadratic programming problem and its solution, in matrix form, is:

$$\hat{\boldsymbol{\beta}}_k = \boldsymbol{\beta}'_k - (\mathbf{b}_k^\top \mathbf{b}_k)^{-1} \mathbf{R}_k^\top \boldsymbol{\gamma} \quad (4.15)$$

where $\boldsymbol{\gamma}$ is a vector of Lagrange multipliers and $\boldsymbol{\beta}'_k = (\mathbf{b}_k^\top \mathbf{b}_k)^{-1} \mathbf{b}_k^\top \hat{A}_n$ is the unconstrained least squares estimator. The elements of the vector $\boldsymbol{\gamma}$ determine how far the solution is from the unconstrained estimated parameter, and they take the value zero when the polynomial is approximating a linear function. If all constraints reduce to $\mathbf{R}_k \boldsymbol{\beta}_k = \mathbf{r}_k$, the results match the form of ordinary constrained least square

$$\hat{\boldsymbol{\beta}}_k = \boldsymbol{\beta}'_k - (\mathbf{b}_k^\top \mathbf{b}_k)^{-1} \mathbf{R}_k^T [\mathbf{R}_k (\mathbf{b}_k^T \mathbf{b}_k)^{-1} \mathbf{R}_k^T]^{-1} (\mathbf{R}_k \boldsymbol{\beta}'_k - \mathbf{r}_k).$$

The vectors $\hat{\boldsymbol{\beta}}_k$ and $\boldsymbol{\gamma}$ can be efficiently computed using an iterative quadratic programming algorithm (e.g. [Goldfarb and Idnani 1983](#)). A high resolution of (4.14) is obtained with increasing values of Q . Numerical experiments show that a close approximation of the true Pickands dependence function is obtained with moderate values of Q . However, Q should not be seen as an additional parameter of the projection estimator. The solution (4.14) provides better approximations of the true Pickands dependence function for increasing sample sizes and polynomial degrees.

4.2.3 Confidence Bands

To assess the uncertainty of the new estimates, we construct confidence bands using a resampling method. For $\mathbf{w} \in \mathcal{S}_d$ and $0 < \tilde{\alpha} < 1$, the bootstrap $(1 - \tilde{\alpha})$ pointwise confidence band, based on the estimates $\tilde{A}_{n,k}^{*(r)}(\mathbf{w})$, $r = 1, 2, \dots$, obtained from the bootstrapped sample $\mathbf{X}_n^{(r)} = (\mathbf{X}_1^{(r)}, \dots, \mathbf{X}_d^{(r)})$, has the drawback that the lower and upper limits of the band are rarely convex and continuous. To bypass this hurdle, we followed the strategy of working with the estimated Bernstein polynomial coefficients themselves. Specifically,

Estimator	Confident bands' type	Parameter α				
		0.3	0.5	0.7	0.9	1
BP-P	Pointwise	31.25	79.43	92.62	83.15	73.64
	Simultaneous	40.14	90.82	97.42	90.47	82.50
BP-Pw	Pointwise	92.43	93.97	91.48	89.64	72.67
	Simultaneous	98.63	93.63	93.67	93.87	81.48
BP-CFG	Pointwise	84.59	88.96	50.11	89.71	44.88
	Simultaneous	86.38	90.68	60.14	91.02	50.79
BP-CFGw	Pointwise	94.25	93.27	89.88	90.74	58.14
	Simultaneous	95.57	96.01	93.86	96.18	64.52
BP-HT	Pointwise	94.41	89.53	85.13	92.35	80.50
	Simultaneous	96.18	95.06	88.65	95.01	87.32
BP-MD	Pointwise	94.02	94.28	89.97	94.23	89.21
	Simultaneous	96.89	96.35	93.85	94.75	89.57

Table 4.1: Pointwise and simultaneous coverage probabilities for the BP method with nominal level 0.95, using four nonparametric estimators based on the symmetric logistic model for different parameter values, sample size equal to 100 and bootstrap resampling size of 500.

let $\widehat{\beta}_k^{*(r)}$ be the Bernstein polynomial coefficient estimator based on the bootstrap sample $\mathbf{X}_n^{(r)}$, $r = 1, 2, \dots$, we define a bootstrap simultaneous $(1 - \tilde{\alpha})$ confidence band by specifying the lower $\tilde{A}_{n,k}^L(\mathbf{w})$ and upper $\tilde{A}_{n,k}^U(\mathbf{w})$ limits as

$$\left[\sum_{\alpha \in \Gamma_k} \widehat{\beta}_\alpha^{*[r(\tilde{\alpha}/2)]} b_\alpha(\mathbf{w}; k); \sum_{\alpha \in \Gamma_k} \widehat{\beta}_\alpha^{*[r(1-\tilde{\alpha}/2)]} b_\alpha(\mathbf{w}; k) \right], \quad \mathbf{w} \in \mathcal{S}_d, \quad (4.16)$$

where $\widehat{\beta}_\alpha^{*[r(\tilde{\alpha}/2)]}$ and $\widehat{\beta}_\alpha^{*[r(1-\tilde{\alpha}/2)]}$, for all $\alpha \in \Gamma_k$, correspond to the $[r(\tilde{\alpha}/2)]$ and $[r(1 - \tilde{\alpha}/2)]$ order statistics respectively and $b_\alpha(\mathbf{w}; k)$ is the Bernstein basis polynomial of degree k , see (4.8). Although this approach does not guarantee convex confidence bands, it works very well in our simulations, where we find that the convexity is violated only when dependence is weak.

Simultaneous coverage probability is also taken into account. It refers to a collection of confidence intervals of each coefficient β_α , $\alpha \in \Gamma_k$, and it is the probability that all of the intervals cover the true values of the coefficients simultaneously. Because of the skewness of the bootstrap distribution, the percentile intervals are not quite symmetric around $\widehat{\beta}$. This can be seen in the boxplots of Figure 4.2 and compared to the “true” values $A(j/k)$ (red points) of the function $B_A(\mathbf{w}; k)$. We estimated the Pickands dependence functions of the logistic dependence model with different polynomial orders ($k = 23, 20, 16, 6, 3$) depending on the strength of dependence ($\alpha = 0.3, 0.5, 0.7, 0.9, 1$). The results for the 95%

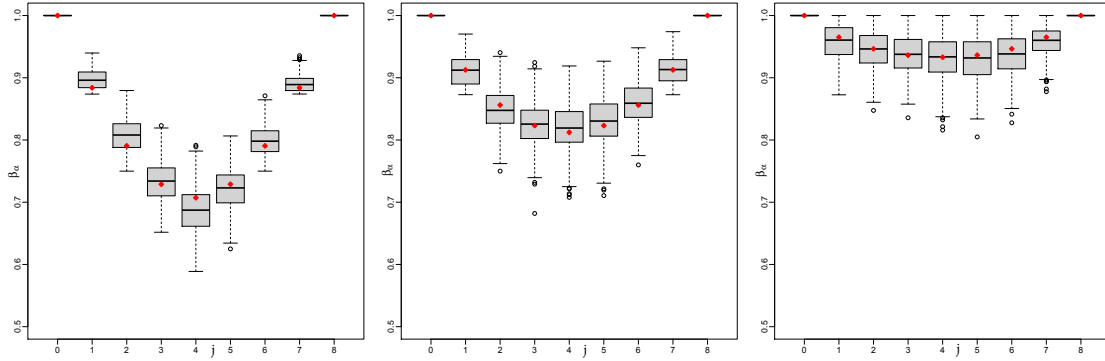


Figure 4.2: Confidence bands of $\hat{\beta}$. 500 replicates with respect to strong, mild and weak dependence of symmetric logistic model (red points).

confidence bands are reported in Table 4.1 and are based on a Monte Carlo sample of size 500. The rows labelled *BP* represent the coverage probabilities for the regularization method proposed in the previous section. The rows labelled *Pointwise* show the coverage probability referring to the standard pointwise method, while the rows labelled *Simultaneous* are computed using equation (4.16). We discover remarkable improvements in coverage when considering simultaneous bands in comparison to the usual pointwise approach; this holds for all estimators. Desirably, the coverage probabilities should be close to the nominal value, which we indeed observe, mainly for the CFGw and MD estimators.

4.3 Asymptotic Results

This section contains the asymptotic results for the madogram statistic and its regularised version. Starting from A_n^{MD} and \hat{A}_n^{MD} , the following theorem summarizes their asymptotic properties. The asymptotic normality requires a smoothness condition on the extreme-value copula C (Segers 2012).

Condition 4.1. For every $i \in \{1, \dots, d\}$, the partial derivative of C with respect to u_i exists and is continuous on the set $\{\mathbf{u} \in [0, 1]^d : 0 < u_i < 1\}$.

Let \mathcal{C} be a C -Brownian bridge, that is, a zero-mean Gaussian process on $[0, 1]^d$ with continuous sample paths and with covariance function given by

$$\text{Cov}(\mathcal{C}(\mathbf{u}), \mathcal{C}(\mathbf{v})) = C(\mathbf{u} \wedge \mathbf{v}) - C(\mathbf{u})C(\mathbf{v}), \quad \mathbf{u}, \mathbf{v} \in [0, 1]^d, \quad (4.17)$$

where the minimum is considered component-wise. Provided Condition 4.1 is satisfied, define the Gaussian process \mathbb{C} on $[0, 1]^d$ by

$$\mathbb{C}(\mathbf{u}) = \mathcal{C}(\mathbf{u}) - \sum_{i=1}^d \frac{\partial C}{\partial u_i}(\mathbf{u}) \mathcal{C}(1, \dots, 1, u_i, 1, \dots, 1), \quad \mathbf{u} \in [0, 1]^d. \quad (4.18)$$

Theorem 4.1. *Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be independent and identically distributed random vectors whose common distribution has continuous margins and extreme-value copula C with Pickands dependence function A . Then:*

a) $\|A_n^{\text{MD}} - A\|_\infty \xrightarrow{as} 0$ as $n \rightarrow \infty$. And in $\ell^\infty(\mathcal{S}_d)$, as $n \rightarrow \infty$,

$$\sqrt{n}(A_n^{\text{MD}} - A) \rightsquigarrow \left((1 + A(\mathbf{w}))^2 \frac{1}{d} \sum_{i=1}^d \int_0^1 (\mathcal{C}(1, \dots, 1, x^{w_i}, 1, \dots, 1) - \mathcal{C}(x^{w_1}, \dots, x^{w_d})) dx \right)_{\mathbf{w} \in \mathcal{S}_d};$$

b) $\|\widehat{A}_n^{\text{MD}} - A\|_\infty \xrightarrow{as} 0$ as $n \rightarrow \infty$. Moreover, if Condition 4.1 is satisfied, then, in $\ell^\infty(\mathcal{S}_d)$, as $n \rightarrow \infty$,

$$\sqrt{n}(\widehat{A}_n^{\text{MD}} - A) \rightsquigarrow \left(-(1 + A(\mathbf{w}))^2 \int_0^1 \mathbb{C}(x^{w_1}, \dots, x^{w_d}) dx \right)_{\mathbf{w} \in \mathcal{S}_d}.$$

In order to state the asymptotic results for the projection estimator the following result is required.

Proposition 4.5. \mathcal{A}_k , $k = 1, 2, \dots$ is a nested sequence in \mathcal{A} . Furthermore, if $A \in \mathcal{A}$ satisfies the condition

$$\Delta_{i,0}^2 A(\boldsymbol{\alpha}/k) - \sum_{j \neq i} |\Delta_{i,0} \Delta_{j,0} A(\boldsymbol{\alpha}/k)| \geq 0, \quad \forall k, \boldsymbol{\alpha} \in \Gamma_{k-2}, i \in \{1, \dots, d-1\}, \quad (4.19)$$

then there exist polynomials $A_k \in \mathcal{A}_k$ such that $\lim_{k \rightarrow \infty} \sup_{\mathbf{w} \in \mathcal{S}_d} |A_k(\mathbf{w}) - A(\mathbf{w})| = 0$.

The asymptotic distribution of the Bernstein-projection estimator based on our multivariate madogram estimator $\widehat{A}_n^{\text{MD}}$ is established in the following proposition.

Proposition 4.6. *Assume that the polynomial degree, k_n , increases with the sample size n in such a way that $k_n/n^{1/2} \rightarrow \infty$ as $n \rightarrow \infty$. If the Pickands dependence function A has bounded and continuous second-order partial derivatives and satisfies the condition (4.19), then*

$$\sqrt{n}(\widetilde{A}_{n,k_n}^{\text{MD}} - A) \rightsquigarrow \left(-\mathbf{b}_k^{*2}(\mathbf{w}) (1 + A(\mathbf{w}))^2 \int_0^1 \mathbb{C}(x^{w_1}, \dots, x^{w_d}) dx \right)_{\mathbf{w} \in \mathcal{S}_d},$$

in $L^2(\mathcal{S}_d)$, where $\mathbf{b}_k^*(\mathbf{w})$ is the row vector $\mathbf{b}_k(\mathbf{w}) (\mathbf{b}_k^\top(\mathbf{w}) \mathbf{b}_k(\mathbf{w}))^{-1} \mathbf{b}_k^\top(\mathbf{w})$.

This follows from the work of [Fils-Villetard et al. \(2008\)](#). Specifically, from Theorem 4.1 we have that $\sqrt{n}(\widehat{A}_n^{\text{MD}} - A) \rightsquigarrow Z$ in $L^2(\mathcal{S}_d)$ as $n \rightarrow \infty$ where Z is a Gaussian process. From Proposition 4.3 we have that $\|B_A - A\|_\infty = O(k^{-1})$ as $k \rightarrow \infty$. Then, using the

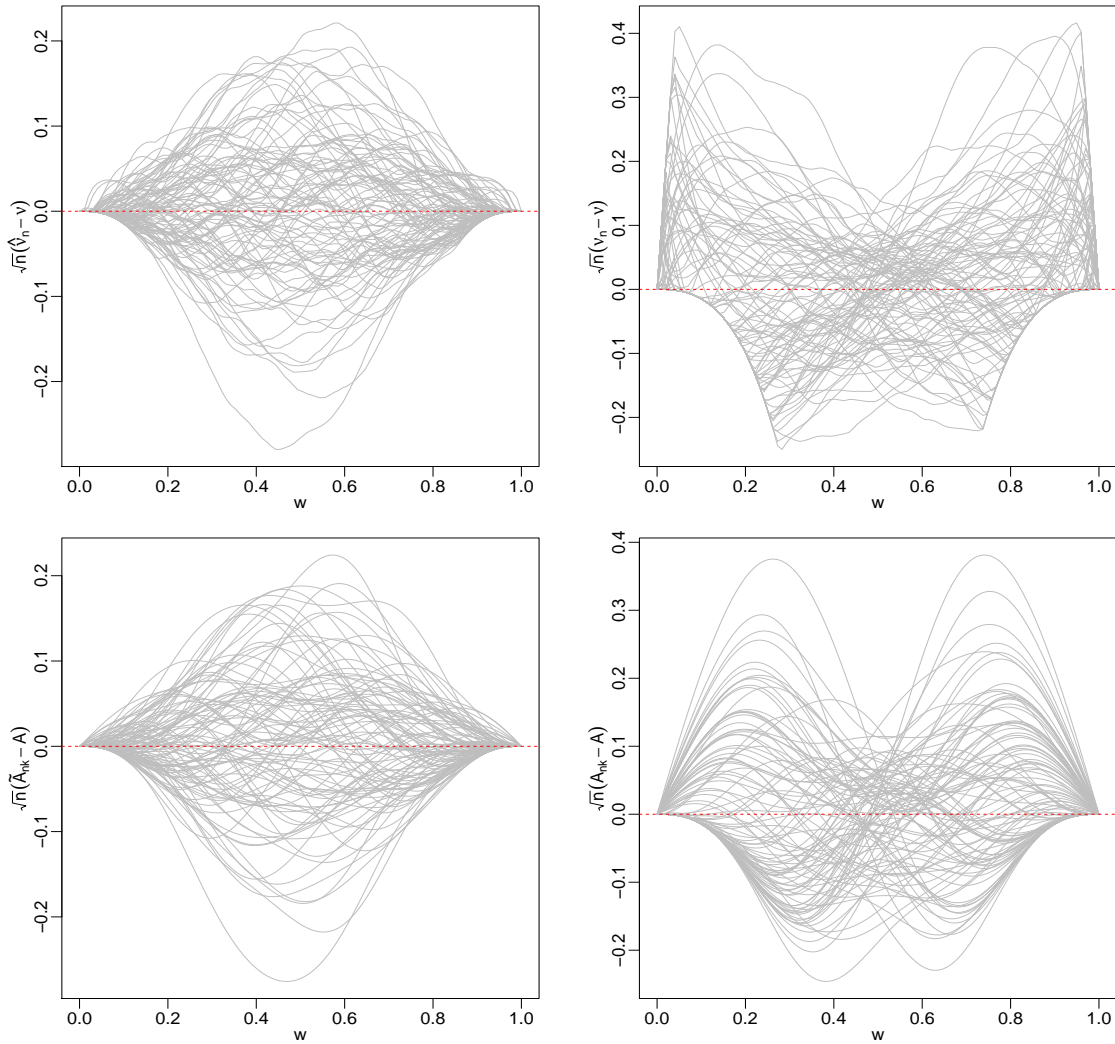


Figure 4.3: Asymptotic convergence of the madogram (top) and its Bernstein projection (bottom) to a Brownian Bridge. Empirical (left) and Frèchet (right) marginal distributions are considered

hypothesis on k_n and applying Lemma 1 in [Fils-Villetard et al. \(2008\)](#) we have that $\|\tilde{A}_{n,k_n}^{\text{MD}} - \tilde{A}_n^{\text{MD}}\| = o_p(n^{1/2})$. As a consequence

$$\sqrt{n}(\tilde{A}_{n,k_n}^{\text{MD}} - A) = \sqrt{n}(\tilde{A}_n^{\text{MD}} - A) + o_p(1) \quad n \rightarrow \infty$$

and the asymptotic distribution of the approximate Bernstein-projection estimator follows from that of our madogram estimator. By applying Theorem 1 in [Fils-Villetard et al. \(2008\)](#) it follows that

$$\sqrt{n}(\tilde{A}_n^{\text{MD}} - A) \rightsquigarrow \arg \min_{Z' \in T_{\mathcal{A}}(A)} \|Z' - Z\|_2, \quad n \rightarrow \infty$$

in $L^2(\mathcal{S}_d)$, where $T_{\mathcal{A}}(A)$ is the tangent cone of \mathcal{A} at A , given by the set of limits of all the sequences $a_n(A_n - A)$, $a_n \geq 0$ and $A_n \in \mathcal{A}$. A definition of Gaussian process projection

can be found, for example, in [Lin and Dunson \(2014\)](#). Let $\{Z(t)\}_{t \in \mathcal{X}}$, with $\mathcal{X} \subseteq \mathbb{R}^d$, denote a Gaussian process with mean $\mu(t) \in \mathbb{R}$ and covariance function $c(h) \in \mathbb{R}$, where $h \in \mathbb{R}^d$ is the separation between two index points. Let \mathcal{M} be a subset of the space of continuous functions mapping from \mathcal{X} to \mathbb{R} having some constraints. The projective Gaussian process P_Z is defined on the constrained space \mathcal{M} as

$$P_Z(t) = \arg \min_{F \in \mathcal{M}} \int_{\mathcal{X}} \{Z(t) - F(t)\}^2 dt, \quad t \in \mathcal{X}.$$

Concluding, the approximated projection estimator of the Pickands dependence function takes the form,

$$\begin{aligned} \tilde{A}_{n,k}^{MD} &= \mathbf{b}_k(\mathbf{w}) \hat{\boldsymbol{\beta}}_k \\ &= \mathbf{b}_k(\mathbf{w}) \boldsymbol{\beta}'_k - \mathbf{b}_k(\mathbf{w}) (\mathbf{b}_k^\top(\mathbf{w}) \mathbf{b}_k(\mathbf{w}))^{-1} \mathbf{R}_k^\top \boldsymbol{\gamma} \\ &= \mathbf{b}_k(\mathbf{w}) (\mathbf{b}_k^\top(\mathbf{w}) \mathbf{b}_k(\mathbf{w}))^{-1} \mathbf{b}_k^\top(\mathbf{w}) \hat{A}_n^{MD} - \mathbf{b}_k(\mathbf{w}) (\mathbf{b}_k^\top(\mathbf{w}) \mathbf{b}_k(\mathbf{w}))^{-1} \mathbf{R}_k^\top \boldsymbol{\gamma} \end{aligned}$$

where \hat{A}_n^{MD} is a column vector of length equal to n , $\mathbf{b}_k(\mathbf{w})$ is the row vector $(b_\alpha(\mathbf{w}; k), \boldsymbol{\alpha} \in \{0, 1, \dots, k\}^d : \alpha_1 + \dots + \alpha_d = k)$ and $\hat{\boldsymbol{\beta}}_k$ is given by expression (4.15). Notice that, in Proposition 4.6, the term $b_k^*(\mathbf{w}) := \mathbf{b}_k(\mathbf{w}) (\mathbf{b}_k^\top(\mathbf{w}) \mathbf{b}_k(\mathbf{w}))^{-1} \mathbf{b}_k^\top(\mathbf{w})$ is around zero, which would decrease the variance of the processes associated to the shape preserving estimator.

It remains an open problem to establish the asymptotic behaviour of the projection estimator without the condition in Proposition 4.5 and the smoothness and boundedness conditions in Proposition 4.6.

Theorem 4.1 and Proposition 4.6 state the asymptotic behaviour of the madogram and its regularised version, respectively. We performed a visual check via Monte Carlo simulation, see Figure 4.3, where the curves of the variances of the two versions are drawn as a function of the argument of the Pickands dependence function. It seems that the limit distribution of $\sqrt{(n)}(\tilde{A}_n - A)$, or $\sqrt{(n)}(\hat{\nu}_n - \nu)$, (using the empirical distributions) has smaller variance than the distribution of $\sqrt{(n)}(A_n - A)$, or $\sqrt{(n)}(\nu_n - \nu)$, (using the theoretical distributions), although restrained. The convergence to a Brownian bridge is almost clear for both the madogram and the Bernstein projection. Using the empirical distribution functions (left panels) rather than the true Frèchet distribution functions (right panels) actually yields a smaller asymptotic variance. We have proved the weak convergence of an empirical copula process indexed by a class of functions. The convergence to Brownian bridges highlights how both the madogram and the Bernstein projection satisfy the constraints at the vertices.

4.4 Numerical Results

In order to visually illustrate the gain in implementing our Bernstein-Bézier projection approach, Figure 4.5 compares some basic estimators \hat{A}_n , reviewed in Section 3.2 (NP) with their Bernstein-Bézier-projection (BP) version defined by (4.14), with $k = 20$, for the special case of the symmetric logistic model (SL, Tawn 1990) with $d = 3$ and $\alpha = 0.3$. For all sample sizes ($n = 10, 20, 50$), an improvement can be observed by comparing the estimated contour lines (black) in the upper and lower panels. This is particularly true for small sample sizes, e.g. for $n = 10$, the corrected version provides much smoother and realistic contour lines.

Proposition 4.6 suggests the use of a large polynomial degree k for large sample sizes, to guarantee a good approximation of A with $\tilde{A}_{n,k}$ (see also Fils-Villetard et al. (2008), Gudendorf and Segers (2011), Gudendorf and Segers (2012)), but computational time limits restrict the choice of k . Figure 4.4 explores this issue for the logistic model with $\alpha = 0.3$ and $n = 100$. As expected from the theory, the choice of k is not anecdotal. A shift in the contour lines appears for the small value $k = 5$, see the left panel of Figure 4.4. This undesirable feature disappears for a moderate value of k (see the right panel with $k = 14$).

To go beyond these visual checks, we also estimate the mean integrated squared error

$$\text{MISE}(\hat{A}_n, A) = \mathbb{E} \left\{ \int_{S_d} \left(\hat{A}_n(\mathbf{w}) - A(\mathbf{w}) \right)^2 d\mathbf{w} \right\},$$

for a variety of setups. The MISE is obtained by repeating 1000 times, with different samples, a given inference method for three different sample sizes $n = 50, 100, 200$.

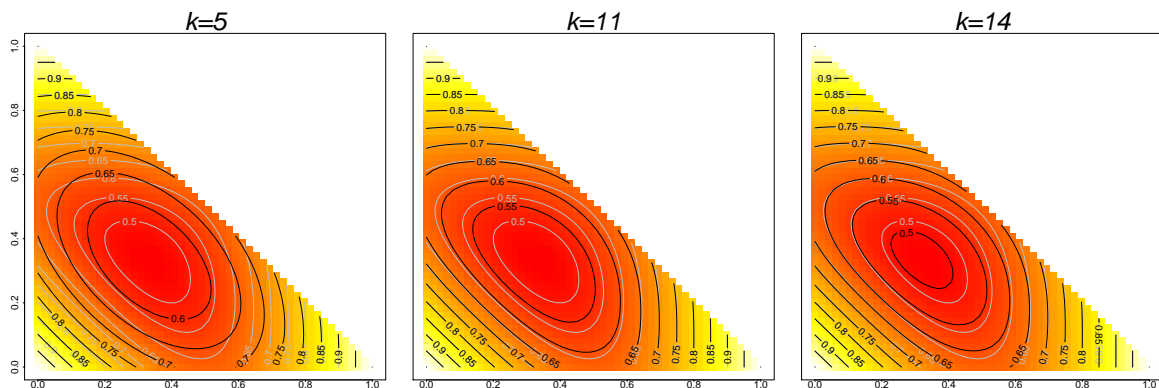


Figure 4.4: Estimates (black lines) of the Pickands dependence function of the symmetric logistic model with $\alpha = 0.3$ (grey line) and sample size 100, obtained with the Bernstein projection of the madogram, with three different values of the polynomial order.

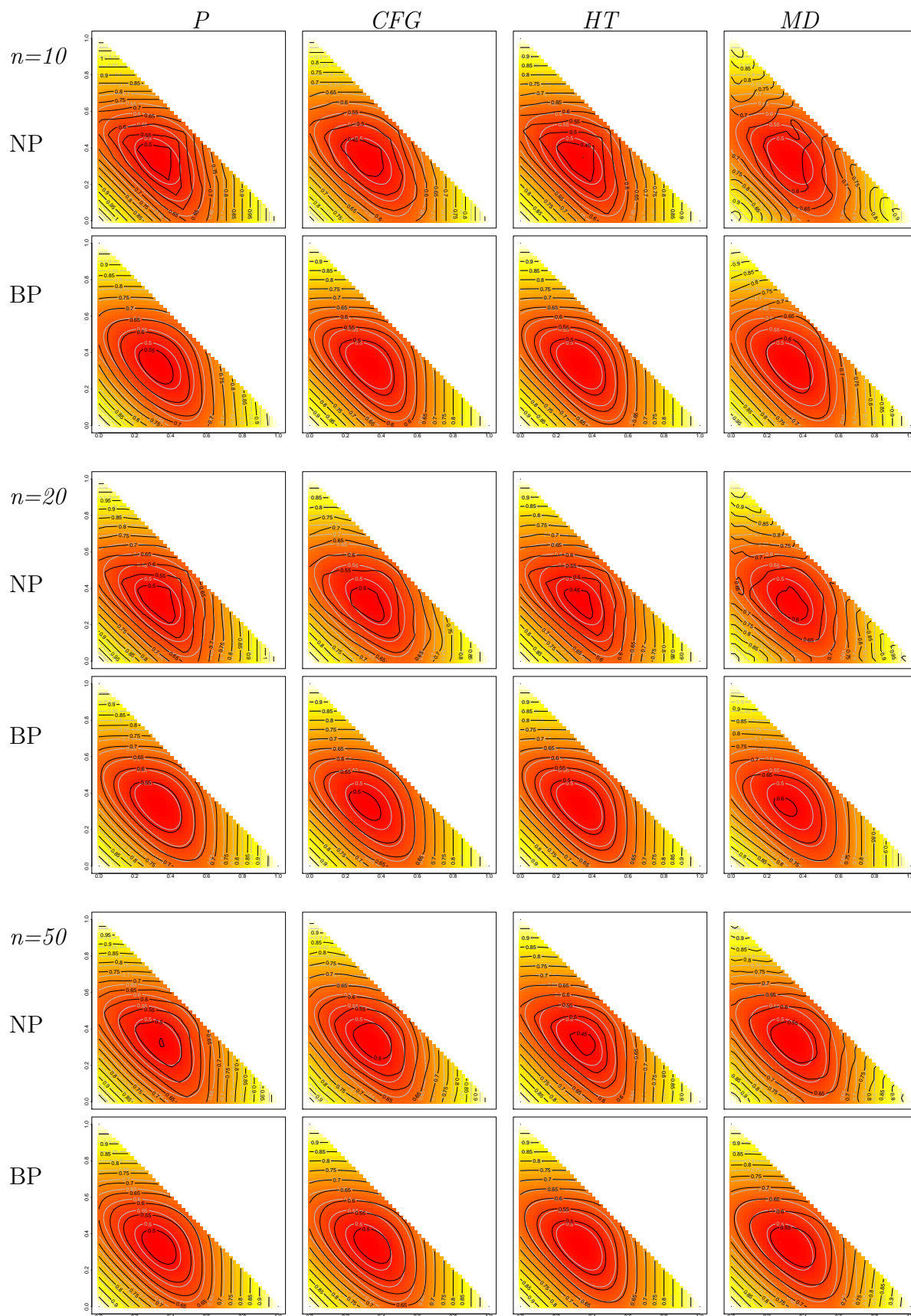


Figure 4.5: Estimates (black lines) of the Pickands dependence function for the symmetric logistic model with $\alpha = 0.3$ (grey line), obtained with four nonparametric estimators (NP) and their BP version with polynomial's degree $k = 20$, respectively for sample sizes $n = 10, 20, 50$ (top-bottom).

The trivariate parametric models considered are: the symmetric logistic (SL) and asymmetric logistic (AL) with dependence parameters $(\alpha, \theta, \phi, \psi) \in (0, 1] \times [0, 1]^3$, (Tawn 1990); the Hüsler–Reiss model (HR) with dependence parameter $(\lambda_1, \lambda_2, \lambda_3) \in (0, \infty)^3$ (Hüsler and Reiss 1989) and the extremal skew- t (EST) with dependence parameters $(\alpha_1, \alpha_2, \alpha_3, \omega_1, \omega_2, \omega_3, \nu) \in (-\infty, \infty)^3 \times (-1, 1)^3 \times (0, \infty)$, (Padoan 2011).

Different dependence strengths have been explored for each of these models. The results summarised in Tables 4.2 and 4.3 are obtained with different values of the parameters. In Tables 4.2, α varies between 0.3 (strong dependence) and 1 (independence) for the SL model. Additionally, in Table 4.3, (i)-(v) duplicate the same strengths of dependence with different parameter settings: α varies between 0.3 and 1, $\beta_1 = 0.2$ and $\beta_2 = 0.8$, for the AL model; the parameter setting for the HR model is $(\lambda_{1,2}, \lambda_{1,3}, \lambda_{2,3}) = \{(0.5, 0.9, 0.8), (0.7, 0.8, 0.33), (0.2, 0.3, 0.13), (0.5, 0.3, 0.29), (0.2, 0.9, 0.88)\}$; the parameters of the extremal skew- t model vary within $(\alpha_1 = 1, \alpha_2 = 1, \omega = -0.2, \nu = 3)$, $(\alpha_1 = -2, \alpha_2 = 2, \omega = -0.2, \nu = 3)$, $(\alpha_1 = -2, \alpha_2 = 2, \omega = 0.5, \nu = 5)$, $(\alpha_1 = 3, \alpha_2 = -6, \omega = 0.6, \nu = 3)$ and $(\alpha_1 = 3, \alpha_2 = -6, \omega = 0.9, \nu = 5)$.

We compare the four non-parametric estimators introduced in Chapter 3: the MD, HT, P and CFG. For comparison purposes we have considered the weighted and endpoint-corrected versions of the P and CFG estimators, Pw and CFGw respectively, as discussed in Gudendorf and Segers (2012), whereas the MD, HT, P and CFG estimators do not have any correction.

We can see that with a strong dependence ($\alpha = 0.3$), the CFG estimator provides the best results, especially for small sample size ($n = 50$). With a medium dependence ($\alpha = 0.5, 0.7$), the four estimators provide similar results. With a weak dependence or in the independence case ($\alpha = 0.9, 1$), the MD estimator provides the best results, especially for small and moderate sample sizes ($n = 50, 100$). The same conclusions arise when considering different models which allow for asymmetric dependence structures (see table 4.3).

Tables 4.4 and 4.5 show how an initial estimate of the Pickands dependence function improves using the projection method. The improvement is computed as

$$\frac{MISE_{NP} - MISE_{BP}}{MISE_{NP}} \times 100,$$

(reported in columns 3–6 in Table 4.4), where $MISE_{NP}$ and $MISE_{BP}$ are the MISE obtained with a non-parametric estimator and its projection respectively. As before, MISE provides a Monte-Carlo approximation of (4.4) obtained from 1000 random samples. The true dependence structure is the symmetric logistic model, α denotes the model parameter, and n and k are the sample size and the polynomial degree respectively. Estimates obtained with the the four non-parametric methods are regularized using the BP method. The order, k , of the polynomial exploited is an “optimal” value in the sense that the

Sample size n	Estimator	Parameter α				
		0.3	0.5	0.7	0.9	1
50	P	4.25×10^{-4}	8.06×10^{-4}	1.47×10^{-3}	2.45×10^{-3}	2.50×10^{-3}
	Pw	1.45×10^{-4}	5.13×10^{-4}	1.26×10^{-3}	2.53×10^{-3}	2.81×10^{-3}
	CFG	2.36×10^{-4}	6.92×10^{-4}	1.87×10^{-3}	4.07×10^{-3}	5.02×10^{-3}
	CFGw	9.17×10^{-5}	4.45×10^{-4}	1.24×10^{-3}	2.66×10^{-3}	3.07×10^{-3}
	HT	2.64×10^{-4}	8.54×10^{-4}	2.59×10^{-3}	5.13×10^{-3}	5.65×10^{-3}
	MD	1.80×10^{-4}	8.66×10^{-4}	1.91×10^{-3}	3.02×10^{-3}	2.87×10^{-3}
100	P	1.53×10^{-4}	3.16×10^{-4}	6.98×10^{-4}	1.20×10^{-3}	1.39×10^{-3}
	Pw	6.36×10^{-5}	2.38×10^{-4}	6.51×10^{-4}	1.25×10^{-3}	1.51×10^{-3}
	CFG	9.54×10^{-5}	3.27×10^{-4}	8.66×10^{-4}	1.78×10^{-3}	2.15×10^{-3}
	CFGw	4.32×10^{-5}	2.21×10^{-4}	6.35×10^{-4}	1.24×10^{-3}	1.39×10^{-3}
	HT	2.61×10^{-4}	7.66×10^{-4}	2.16×10^{-3}	4.24×10^{-3}	5.27×10^{-3}
	MD	7.02×10^{-5}	3.18×10^{-4}	7.91×10^{-4}	1.19×10^{-3}	1.09×10^{-3}
200	P	5.87×10^{-5}	1.54×10^{-4}	3.40×10^{-4}	6.25×10^{-4}	7.24×10^{-4}
	Pw	3.01×10^{-5}	1.31×10^{-4}	3.28×10^{-4}	6.60×10^{-4}	7.59×10^{-4}
	CFG	3.87×10^{-5}	1.58×10^{-4}	4.00×10^{-4}	8.31×10^{-4}	8.52×10^{-4}
	CFGw	2.12×10^{-5}	1.23×10^{-4}	3.24×10^{-4}	6.36×10^{-4}	5.90×10^{-4}
	HT	2.55×10^{-4}	7.31×10^{-4}	2.05×10^{-3}	3.82×10^{-3}	5.85×10^{-3}
	MD	3.17×10^{-5}	1.58×10^{-4}	3.70×10^{-4}	5.81×10^{-4}	4.91×10^{-4}

Table 4.2: MISE of four estimators of the Pickands dependence function based on a symmetric logistic dependence model for different parameter values and sample sizes.

MISE does not decrease significantly for larger values of k . It turns out that with a weak dependence a small value of k is sufficient, while with a strong dependence a large value of k is needed. This makes sense if we view a dependence structure as an added complexity, especially with respect to the independence case, the simplest possible model. In such a framework, the polynomial degree has to be higher in order to capture this additional information. For the symmetric logistic model, in Table 4.4, the improvements by regularizing each of the four estimators, sorted from largest to smallest, are: HT, P, MD and CFG. For each estimator, the improvements, sorted from largest to smallest, are obtained with: independence ($\alpha = 1$), weak dependence ($\alpha = 0.9$), strong dependence ($\alpha = 0.3$) and medium dependence ($\alpha = 0.5, 0.7$). These results are compared with those provided by [Gudendorf and Segers \(2012\)](#), which are obtained using the discrete spectral measure projection method proposed by the authors (see columns 7,8). Again, the same considerations can be made for the asymmetric logistic, Hüsler-Reiss and extremal skew- t models. The gain of the Bernstein projection method is stressed especially for the case of independence where the pilot estimator may be much less smooth. In this case, no comparisons could be done, with respect to the discrete spectral measure projection

Model	Estimator	Parameter				
		(i)	(ii)	(iii)	(iv)	(v)
AL	P	9.58×10^{-4}	1.08×10^{-3}	1.28×10^{-3}	1.40×10^{-3}	1.44×10^{-3}
	P _w	9.82×10^{-4}	1.06×10^{-3}	1.32×10^{-3}	1.50×10^{-3}	1.59×10^{-3}
	CFG	1.20×10^{-3}	1.24×10^{-3}	1.71×10^{-3}	1.97×10^{-3}	2.16×10^{-3}
	CFG _w	7.90×10^{-4}	8.16×10^{-4}	1.15×10^{-3}	1.30×10^{-3}	1.41×10^{-3}
	HT	1.03×10^{-3}	1.06×10^{-3}	1.29×10^{-3}	1.30×10^{-3}	1.29×10^{-3}
	MD	8.09×10^{-4}	9.59×10^{-4}	1.07×10^{-3}	1.16×10^{-3}	1.14×10^{-3}
HR	P	8.46×10^{-3}	5.57×10^{-3}	1.81×10^{-3}	2.73×10^{-3}	7.53×10^{-3}
	P _w	7.10×10^{-3}	4.82×10^{-3}	1.51×10^{-3}	2.27×10^{-3}	6.32×10^{-3}
	CFG	5.75×10^{-3}	4.07×10^{-3}	1.07×10^{-3}	2.31×10^{-3}	5.30×10^{-3}
	CFG _w	6.99×10^{-3}	4.81×10^{-3}	1.57×10^{-3}	2.42×10^{-3}	6.16×10^{-3}
	HT	6.67×10^{-3}	4.44×10^{-3}	1.39×10^{-3}	2.21×10^{-3}	5.94×10^{-3}
	MD	9.41×10^{-3}	6.18×10^{-3}	1.69×10^{-3}	3.02×10^{-3}	8.05×10^{-3}
EST	P	8.25×10^{-3}	5.74×10^{-3}	1.74×10^{-3}	2.85×10^{-3}	7.45×10^{-3}
	P _w	6.93×10^{-3}	4.83×10^{-3}	1.46×10^{-3}	2.47×10^{-3}	6.26×10^{-3}
	CFG	5.64×10^{-3}	4.17×10^{-3}	1.03×10^{-3}	2.57×10^{-3}	5.21×10^{-3}
	CFG _w	6.78×10^{-3}	4.97×10^{-3}	1.43×10^{-3}	2.67×10^{-3}	6.14×10^{-3}
	HT	4.47×10^{-3}	3.07×10^{-3}	1.18×10^{-3}	1.72×10^{-3}	3.37×10^{-3}
	MD	7.21×10^{-3}	6.38×10^{-3}	1.36×10^{-3}	3.15×10^{-3}	2.47×10^{-3}

Table 4.3: MISE of four estimators of the Pickands dependence function based on a asymmetric logistic, Hüsler-Reiss and extremal skew- t dependence models for different parameter values and sample sizes equal to 100.

method. We can conclude that overall the BP method provides a better percentage improvement.

A further issue of this simulation study concerns the choice of the grid size. In general, any estimator of the Pickands dependence function is evaluated on a grid of a finite set of points $\{\mathbf{w}_q, q = 1, \dots, Q\}$, where Q determines how dense is the grid. One naturally expects to improve the solution when increasing the number of points of the grid. However, the goodness of our projection method mainly depends on sample size, rather than the grid size. Let consider the expectation of $\frac{1}{Q} \sum_{q=1}^Q \left(\mathbf{b}_k(\mathbf{w}_q) \hat{\boldsymbol{\beta}}_k - \hat{A}_n(\mathbf{w}_q) \right)^2$ for increasing values of Q , see Table 4.6. The procedure has been repeated for 1000 sample. In average, the expectation does not change remarkably, which stress the fact that even a small grid is enough in order to have a good performance.

We also explore the validity of our procedure by deriving bootstrap pointwise and simultaneous $(1 - \alpha)$ confidence bands, see Table 4.1, in Section 4.2.3, which displays 95%

			Projection method									
			Bernstein-Bézier				Discrete spectral measure					
			% Improvement									
			Estimator					Estimator				
n	α	k	P	Pw	CFG	CFGw	HT	MD	Pw	CFGw		
50	0.3	23	18.11	13.34	76.84	18.97	51.53	8.50	2.14	0.82		
	0.5	20	8.19	5.44	13.98	1.46	12.52	2.22	5.51	1.17		
	0.7	16	15.60	11.01	4.43	2.10	9.05	6.48	11.03	3.17		
	0.9	6	44.70	25.92	3.98	6.51	16.93	48.72	22.39	4.37		
	1	3	69.95	34.53	4.92	9.04	34.68	93.60	29.07	4.89		
100	0.3	23	16.59	13.36	59.75	13.43	45.45	7.41	1.27	0.40		
	0.5	20	5.85	3.83	7.59	0.63	9.78	1.23	3.52	0.84		
	0.7	16	9.89	8.15	2.21	0.95	6.42	2.74	7.51	1.37		
	0.9	6	34.95	23.98	3.48	6.50	8.33	26.72	23.13	4.07		
	1	3	68.00	39.35	5.93	11.50	19.22	87.46	36.10	8.11		
200	0.3	23	15.16	10.63	37.73	5.66	44.72	5.05	0.60	0		
	0.5	20	3.06	2.51	3.80	0.41	9.06	0.13	2.10	0		
	0.7	16	5.70	5.22	0.90	0	5.60	0.76	4.85	0.88		
	0.9	6	25.22	20.48	3.43	6.07	5.53	13.39	18.52	3.28		
	1	3	69.17	46.32	8.63	16.06	10.88	81.99	40.99	11.89		

Table 4.4: Percentage improvement of the MISE gained with the projection method for the symmetric logistic model.

coverage probabilities from 1000 independent samples and $r = 500$ bootstrap resampling. The parametric setup is identical to the one used in Table 4.4, but with fixed sample size equal to $n = 100$. Overall, excluding the independence case, the coverage probabilities are close to the nominal value of 0.95 with both methods. However, the simultaneous method (4.16) outperforms the pointwise method, since the coverage probabilities are always larger.

To close this small simulation study, we illustrate the bivariate and trivariate confidence bands extending the class of parametric families, see Figures ?? and 4.7. The setting considered is: the asymmetric logistic model (AL, Tawn 1990) with $\beta_1 = 0.6$, $\beta_2 = 0.3$, the Hüsler-Reiss model (HR, Hüsler and Reiss 1989) with three parameters settings $(\lambda_1 = 0.8, \lambda_2 = 0.3, \lambda_3 = 0.7)$, $(\lambda_1 = 0.49, \lambda_2 = 0.51, \lambda_3 = 0.03)$, $(\lambda_1 = 0.24, \lambda_2 = 0.23, \lambda_3 = 0.11)$ and the extremal skew- t model (EST, Padoan 2011) with three setups $(\alpha = 7, -10, 1, \nu = 3, \omega = 0.9)$, $(\alpha = -2, 9, -15, \nu = 2, \omega = 0.9)$, $(\alpha = -0.5, -0.5, -0.5, \nu = 3, \omega = 0.9)$. Figure 4.7 shows that, for all these cases, the lower and upper limits of the variability bands are always convex functions and they always contain the true Pickands dependence function. The variability bands of weaker dependence structures

% Improvement Bernstein-Bézier Projection method						
Model	Estimator	Parameter α				
		(i)	(ii)	(iii)	(iv)	(v)
AL	P	17.25	27.49	34.53	59.74	66.68
	Pw	11.20	20.15	24.17	34.43	39.39
	CFG	2.94	6.55	4.93	5.86	6.77
	CFGw	4.16	7.00	7.94	10.18	12.05
	HT	10.67	17.00	19.38	20.30	17.21
	MD	9.95	17.20	28.39	69.61	86.13
HR	P	1.59	0.15	8.96	9.39	7.61
	Pw	3.37	0.84	7.69	18.21	9.80
	CFG	0.09	19.84	21.90	8.77	11.76
	CFGw	0.28	3.65	23.01	9.65	11.92
	HT	1.70	3.84	16.18	17.33	11.26
	MD	0.17	1.84	18.03	0.00	6.81
EST	P	3.08	3.36	7.23	46.30	22.75
	Pw	1.54	3.43	2.53	5.10	5.08
	CFG	0.01	17.38	17.32	1.71	8.18
	CFGw	1.07	0.09	7.23	1.18	6.20
	HT	2.58	1.91	6.03	8.22	6.34
	MD	0.62	4.50	11.92	0.02	7.56

Table 4.5: Percentage improvement of the MISE when using the projection method, for the asymmetric logistic, Hüsler-Reiss and extremal skew- t models.

are typically wider than those of stronger dependence structures. The same is true for asymmetric versus symmetric dependence structures.

Concluding the analysis of the performance of the Bernstein-projection estimator proposed, let consider its computational complexity. The convex quadratic programming has a polynomial time complexity. For example, [Ye and Tse \(1989\)](#) showed that a convex quadratic problem that creates a sequence of interior feasible points, converges to the optimal feasible solution in $O(p_k)$ iterations, where p_k is the length of the β_k vector of coefficients which depends on both the order of the polynomials k and d . Each iteration can be computed in $O(p_k^3)$ arithmetic operations. About quadprog's inner implementation and time-complexity, [Table 4.7](#) resumes computational times for the estimation of the Pickands dependence function $\tilde{A}_{n,k}$, when increasing the order of the polynomials k and d .

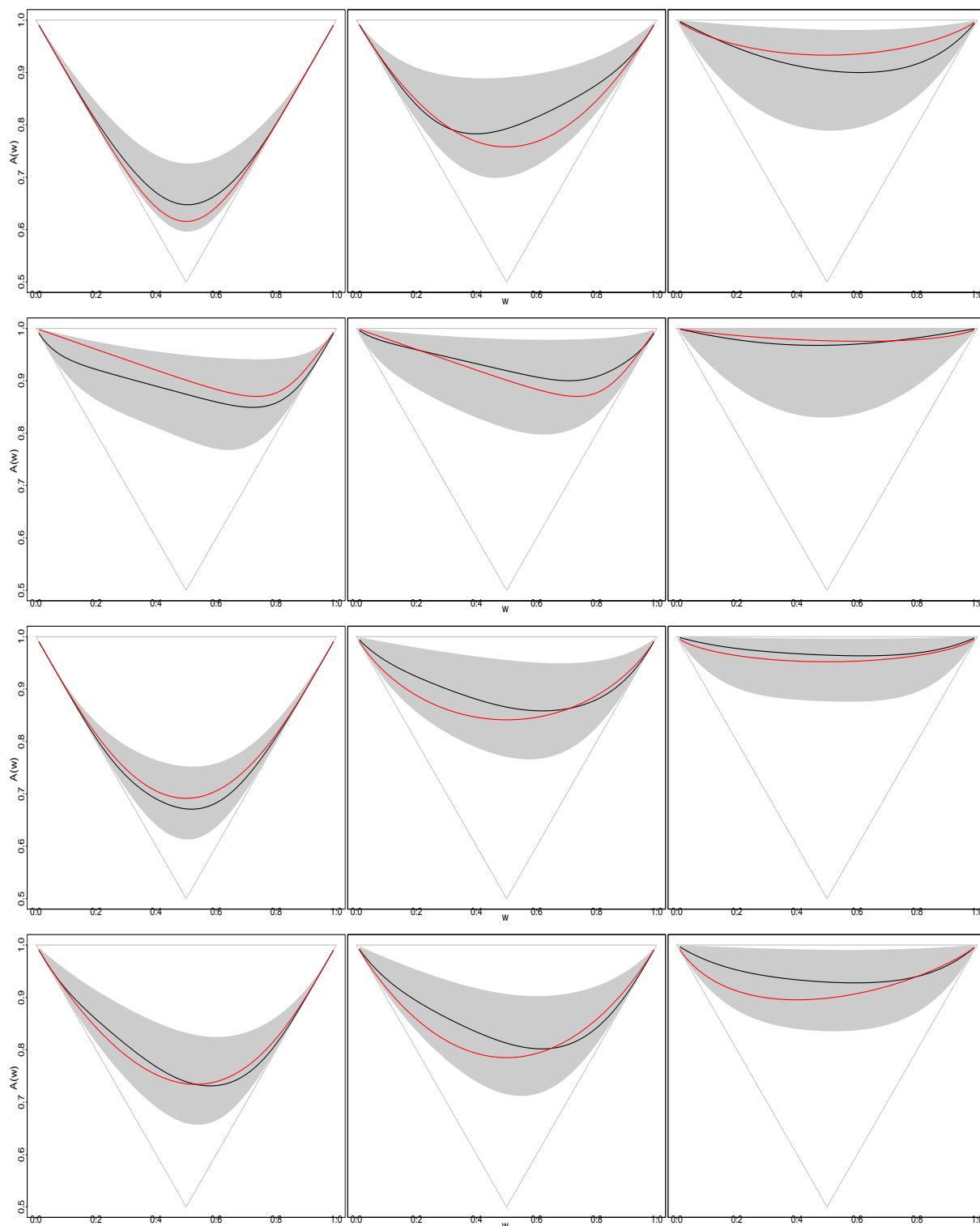


Figure 4.6: Estimates of the Pickands dependence functions $A(w)$ (black line) with bootstrap variability bands (grey shade). Strong, mild and weak dependencies (left-right) are modelled with polynomial orders $k = 23, 16, 6$, respectively, for (top-bottom) the Symmetric Logistic ($\alpha = 0.3, 0.6, 0.9$), Asymmetric Logistic ($\alpha = 0.3, 0.6, 0.9, \beta_1 = 0.8, \beta_2 = 0.2$), Hüsler-Reiss ($\lambda = 2, 1, 0.6$) and Tilted Dirichlet models ($\alpha = 3, 0.7, 0.1, \beta = 0.7$)

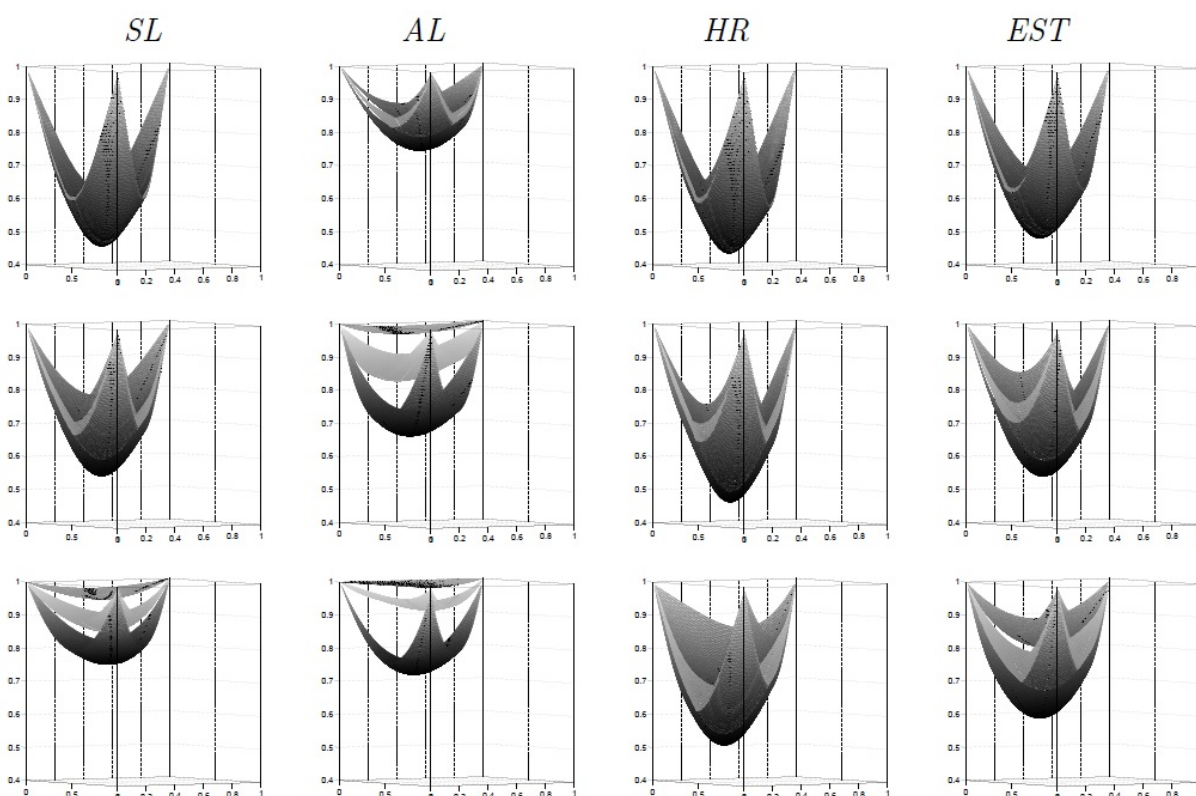


Figure 4.7: Estimates of trivariate Pickands (light grey shade) and bootstrap variability bands (dark grey shade) for the SL, AL, HR, EST (left-right) models with strong, mild and weak dependence (top-bottom)

d	Q					
	10	15	20	30	50	100
2	0.088	0.093	0.093	0.095	0.087	0.097
3	0.184	0.173	0.183	0.176	0.175	0.176

Table 4.6: Expectation of $\frac{1}{Q} \sum_{q=1}^Q \left(\mathbf{b}_k(\mathbf{w}_q) \hat{\boldsymbol{\beta}}_k - \hat{A}_n(\mathbf{w}_q) \right)^2$ for increasing values of $Q = 10, 15, 20, 30, 50, 100$ for bivariate and trivariate maxima vectors.

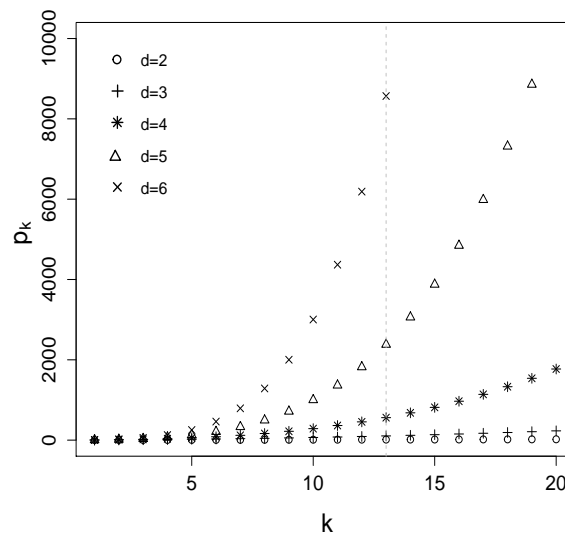


Figure 4.8: Length of the $\boldsymbol{\beta}$ vector of coefficients, p_k , when increasing the order of the polynomial k and the dimension d

k	d					
	2	3	4	5	6	7
3	0.0003	0.0068	0.14	1.84	2.09	2.47
6	0.0005	0.0075	0.26	7.35	22.21	133.09
16	0.0005	0.0170	15.52	NA	NA	NA
20	0.0005	0.0367	75.17	NA	NA	NA
23	0.0005	0.0708	218.76	NA	NA	NA

Table 4.7: Computational time (in minutes) for the estimation of the Pickands dependence function by means of the projection estimator $\tilde{A}_{n,k}$, with respect to a processor Intel(R) Core(TM) i5-4440 CPU 3.10 GHz. NA entries are determined by lack of memory in the processor for data storage.

4.5 Weekly Maxima of Hourly Rainfall in France

The methodology discussed so far is applied in order to estimate the extremal dependence of French heavy rainfalls. The data are provided by the French meteorological service, Météo-France. Recalling Figure 1.1, introduced in Chapter 1, our goal here is to measure the dependence within each cluster of size $d = 7$. This dataset was studied by Bernard et al. (2013) who performed a cluster analysis for detecting clusters of locations with similar spatial dependence structures over the French territory. 7 clusters have been identified with about 8-20 climatologically homogeneous stations and with similar topographical features. Thus, the focus on clusters is motivated by the aim of describing the dependence of locations with homogeneous climatology characteristics, within a cluster, and heterogeneous between clusters. Climatologically, extreme precipitation that affect the Mediterranean coast in the fall are caused by the interaction of southern and mountains winds coming from the Pyrénées, Cévennes and Alps regions. In the north of France, heavy rainfall is often produced by mid-latitude perturbations in Brittany or in the north of France and Paris. It can be verified that, indeed, extremes within clusters are strongly dependent. See Bernard et al. (2013) for a complete description. Here, we consider the same clusters but, within each of them, we select 7 random stations. Figure 1.1 shows the selected stations and those belonging to the same cluster are depicted with the same colour. There are $n = 228$ data points available for 49 stations. Therefore, the dimension of our multivariate vector of weekly maxima of hourly rainfall is seven and it would be of interest for hydrologists to infer and compare the dependence among these multivariate vectors of maxima with seven components. Such an endeavour represents the main motivation of this work.

For each cluster, we compute our Bernstein-projection estimator based on the madogram. To summarize this seven dimensional dependence structure, we take advantage of the extremal coefficient (2.13) defined in Chapter 2. It satisfies the condition $1 \leq \vartheta \leq d$, where the lower and upper bounds represent the cases of complete dependence and independence among the extremes, respectively. Thus, in each cluster, the extremal coefficient is estimated using the equation $\hat{\vartheta} = 7 \tilde{A}_{n,k}^{\text{MD}}(1/7, \dots, 1/7)$ and so $\hat{\vartheta}$ always belongs to the interval $[1, 7]$. The range of the estimated coefficients is around 3.5, which corresponds to strong dependence, and 4.6 corresponding to medium dependence.

As climatologically expected, we can see in Figure 1.1 a latitudinal gradient in the estimated extremal coefficients. They are smaller in the northern regions and higher in the south. This can be explained by westerly fronts above the 46^{e} latitude that affect large regions, while extreme precipitation in the south is more likely driven by localised convective storms with weak spatial dependence structures. Finally, in the centre of the country, away from the costs, we observe the highest degree of dependence among extremes, as they are the result of the meeting between different densities of air masses.

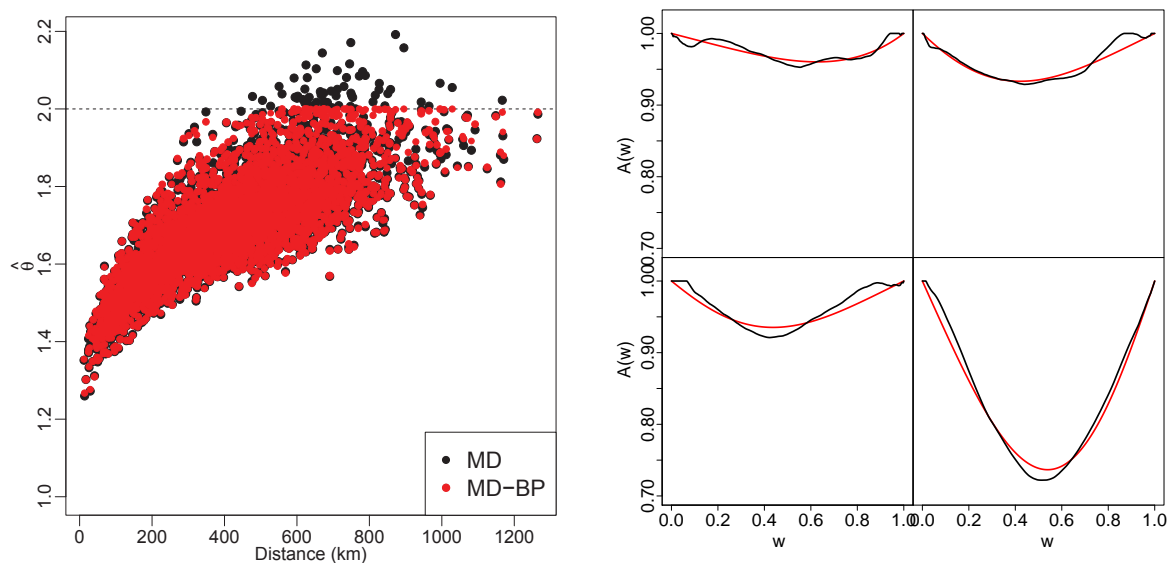


Figure 4.9: Estimated pairwise extremal coefficients and Pickands dependence functions of French weekly precipitation maxima

For all possible pairs of locations, we have also estimated the bivariate Pickands dependence function using the madogram estimator and its projection. The left panel of Figure 4.9 shows a scatterplot of the pairwise extremal coefficients, computed through the estimated Pickands dependence functions, versus the Euclidean distance between sites. Notice that $\hat{\vartheta} \leq 1.5$ for the locations that are less than 200 km apart, meaning that the extremes are strongly or at least mildly dependent, while for sites more than 200 km apart, $\vartheta > 1.5$ meaning that the extremes are independent or at most mildly dependent. The graph also shows the benefit of using the projected version of an unconstrained estimator. Indeed, in this case the extremal coefficients fall within the admissibility range $[1, 2]$, while they can be greater than 2 when they are derived with the unconstrained estimator.

The right plot of Figure 4.9 shows four examples of estimated Pickands dependence functions obtained with pairs of sites whose distances are 979.8, 505.9, 390.1 and 158.1 km, respectively (top-left to bottom-right panels). We see that the madogram estimator provides estimates (black lines) that are not convex functions and, hence, are invalid Pickands dependence functions. Contrarily, the estimates obtained with the projection estimator (red lines) are valid Pickands dependence functions.

Tesi di dottorato "Non-Parametric Inference and Simulation in Extreme Value Theory"
di MARCON GIULIA

discussa presso Università Commerciale Luigi Bocconi-Milano nell'anno 2016

La tesi è tutelata dalla normativa sul diritto d'autore (Legge 22 aprile 1941, n.633 e successive integrazioni e modifiche).

Sono comunque fatti salvi i diritti dell'università Commerciale Luigi Bocconi di riproduzione per scopi di ricerca e didattici, con citazione della fonte.

Chapter 5

Simulation based on Bernstein polynomial representations

*The one “chance in a million” will occur,
with no less and no more than its appropriate frequency,
however surprised we may be that it should occur to us*

(R. Fisher)

Beyond the estimation aspect, another aspect of fundamental importance is the simulation of excesses on the base of the estimated dependence structure. By means of the simulation we are able to assess return levels, which are extreme events even more extreme than those that have already been observed, or return periods. In applications, in the multivariate context, being able to simulate extreme events is useful when there are missing values, for example, if we consider heavy rainfall recorded in different weather stations spread over a region. However, the real problem is to compute the return levels associated to the return periods for the entire region, including the areas where there are not stations.

We propose a new simulation method to simulate observations from a bivariate max-stable random vector. According to our knowledge, there are very few nonparametric simulation methods available in the literature, one example is [Ghoudi et al. \(1998\)](#). Our simulation method relies on the following two steps: firstly, observations are generated with a given nonparametric dependence determined by the “angular” density; secondly, the intensities of the events are generated using the “radial” density. The first step is carried out through the use of the nonparametric Pickands estimator based on the Bernstein Polynomials. At the present, for an easier discussion and on the basis of the available results, we concentrate on the bivariate case. We discuss the properties of the proposed simulation method and illustrate its performance through a simulation study. We show the applicability of this framework in bivariate problems.

5.1 Motivation

Max-stable distributions are important in extreme value theory as the Gaussian distribution is fundamental for the classical theory. Modelling and estimating the dependence structure is crucial in order to perform a reliable statistical analysis of multivariate extremes. One fundamental scope of extreme value analyses is to be able to estimate how often a very extreme event will be exceeded in the future. Such a specific event, which can be more extreme than those which have already been observed, this is known as return level associated to the return period. The latter is the average inter-arrival time with which a return level and or more extreme event occur.

The problem of estimating the return levels using parametric max-stable models, in the univariate and multivariate case, has been extensively studied. In the multivariate case, there are different ways to compute the return levels, for a simple approach see for example [Beranger and Padoan \(2015\)](#). When the dependence structure is modelled by a nonparametric approach, it is not obvious at all how to derive the return levels. In this context, an easy way to derive return levels is by simulation. With our approach, the computation of return levels via simulation is computationally feasible and provides reliable estimates.

The problem of assessing the extremal dependence, especially in the bivariate case, has been widely studied. Chapter 4 proposed a nonparametric method to model the extremal dependence by means of the Pickands dependence function of multivariate max-stable vectors. With the aim of computing return levels, once an estimate of the extremal dependence structure is available, it is of crucial importance to be able to simulate observations from a max-stable distribution with the estimated nonparametric dependence form. This is the motivation of the present Chapter.

Essentially, our simulation method consists in modelling nonparametrically the Pickands dependence function using the Bernstein polynomials and hence exploiting the functional expression for it. More specifically, we show that the first derivative of a Bernstein polynomial representation of the Pickands dependence function is nothing but a mixture of Beta densities. Therefore, exploiting the relationship between the Pickands dependence function and the angular density, we are able to simulate from the latter using the mixture of Beta densities. By adding the simulation of the marginal component, the overall objective is attained.

5.2 Extremal Dependence Functions in Bernstein Form

Let H be a continuous univariate c.d.f. on \mathbb{R} . Then, there is an associated convex function to H , given by the relation (Bassett Jr 1997),

$$A(w) = \int_{-\infty}^w H(x) dx, \quad w \in \mathbb{R},$$

where A is a convex function but not necessarily a Pickands dependence function. We know that H is a non-decreasing function, thus, the function

$$H(w) := \frac{\partial}{\partial w} A(w) \tag{5.1}$$

is non-decreasing. An important remark is that A is not just any convex function, its derivative satisfies the boundary conditions of a c.d.f, that is,

$$\lim_{w \rightarrow -\infty} H(w) = 0$$

and

$$\lim_{w \rightarrow \infty} H(w) = 1.$$

In general, it is clear for any element of the class of convex functions with derivatives satisfying the above boundaries conditions, there is an associated c.d.f., H .

Now, assume that A is a differentiable Pickands dependence function, thus convex. Then, we can write

$$A(w) = \int_0^w H^*(x) dx, \quad w \in [0, 1], \tag{5.2}$$

where H^* is a non-decreasing function in $[0, 1]$ which does not necessarily satisfy the boundary conditions. However, we can always find a normalizing constant c such that the function such that H satisfies the boundaries conditions. In particular, if we consider the following standardization, we obtain a function which is a proper cumulative distribution function,

$$H(w) = \frac{A'(w) - A'(0)}{A'(1) - A'(0)}. \tag{5.3}$$

Thus, we can derive the density h by

$$h(w) = \frac{\partial}{\partial w} H(w)$$

and so by (5.3) we obtain

$$h(w) = \frac{\partial^2}{\partial^2 w} \frac{A(w)}{A'(1) - A'(0)}.$$

We observe that the above statements are in accordance with the general theory of extreme values (Coles and Tawn 1996). Indeed, given

$$V(z_1, z_2) = \int_0^1 \max\left(\frac{t}{z_1}, \frac{1-t}{z_2}\right) H^*(dt),$$

we have that

$$\begin{aligned} A(w) &= \int_0^1 \max\{(1-w)t, w(1-t)\} H^*(dt) \\ &= w \int_0^w (1-t) H^*(dt) + (1-w) \int_w^1 t H^*(dt), \end{aligned} \quad (5.4)$$

since $(1-w)t > w(1-t)$ if and only if $0 < w < t < 1$ (and vice versa). Therefore $A'(w) = H^*(w) - 1$, from which it follows that $H^*(w) = A'(w) + 1$ and hence

$$h^*(w) = \frac{\partial}{\partial w} H^*(w) = \frac{\partial^2}{\partial^2 w} A(w).$$

In other terms, using Coles and Tawn's result we have

$$h^*(w) = -(z_1 + z_2)^3 \frac{\partial^2}{\partial z_1 \partial z_2} V(z_1, z_2) \Big|_{\substack{z_1=rw \\ z_2=r(1-w)}}.$$

Now, if we compute the derivatives of V we obtain

$$\frac{\partial^2}{\partial z_1 \partial z_2} V(z_1, z_2) \Big|_{\substack{z_1=rw \\ z_2=r(1-w)}} = -\frac{1}{r^3} \frac{\partial^2}{\partial^2 w} A(w)$$

and therefore

$$h^*(w) = \frac{\partial^2}{\partial^2 w} A(w).$$

To conclude, by normalizing h we obtain the p.d.f

$$h(w) = h^*(w) \times c'.$$

In the previous Chapter we showed that it is possible to represent the estimate of the Pickands function by means of a Bernstein polynomial form of order k ,

$$A(w) = \sum_{j=0}^k \beta_j b_j(w; k)$$

where $b_j(w; k) = \binom{k}{j} w^j (1-w)^{k-j}$ as in Remark 4.4.

The strength of the proposed method is that this kind of polynomials maintain the shape constraints required, by choosing particular vectors of coefficients β_j which reflect such restrictions. In the bivariate case, the estimated function $\tilde{A}_n(w)$ is a valid Pickands

dependence function, so starting from it we may obtain estimate of the angular density distribution, h , by using a relation proposed in (Coles and Tawn 1996).

Let H represented as in (5.3), then, by using the results given in 4.5 and 4.6, we obtain the following result:

$$\begin{aligned}
h(w) &= \frac{A''(w)}{A'(1) - A'(0)} \\
&= \frac{k(k-1)}{k(\beta_k - \beta_{k-1}) - k(\beta_1 - \beta_0)} \sum_{j=0}^{k-2} (\beta_{j+2} - 2\beta_{j+1} + \beta_j) b_j(w; k-2) \\
&= \sum_{j=0}^{k-2} \frac{\beta_{j+2} - 2\beta_{j+1} + \beta_j}{\beta_k - \beta_{k-1} - \beta_1 + \beta_0} \text{Be}(w; j+1, k-j-1) \\
&= \sum_{j=0}^{k-2} \gamma_j \text{Be}(w; j+1, k-j-1)
\end{aligned} \tag{5.5}$$

where $\gamma := (\beta_{j+2} - 2\beta_{j+1} + \beta_j) / (\beta_k - \beta_{k-1} - \beta_1 + \beta_0)$. It is straightforward to check that the last expression in (5.5) is a mixture of Beta density functions, with parameters $j+1$ and $k-j-1$, as the sum of the coefficients γ_j , $j = 0, \dots, k-2$, is equal to one. Once the coefficients β_j , and consequently γ_j , are estimated, hence it is possible to simulate w from $\tilde{h}(w)$.

5.3 Simulation Scheme

Let $M := \max(X_1, X_2)$, then

$$\begin{aligned}
\mathbb{P}\{M \leq x\} &= \mathbb{P}\{\max(X_1, X_2) \leq x\} \\
&= \mathbb{P}\{X_1 \leq x, X_2 \leq x\} \\
&= \mathbb{P}\{X_1 \leq x\}^\vartheta \\
&= \exp\left\{-\frac{V(1,1)}{x}\right\} \\
&= \exp\left\{-\frac{2A(1/2)}{x}\right\},
\end{aligned}$$

therefore M has Fréchet distribution with location parameter equal to zero and scale parameter equal to $\vartheta = 2A(1/2)$, that is, the extremal coefficient. From the relation with the pseudo-polar coordinates, we have that $X_{\ell i} = r_\ell w_i$, $i = 1, \dots, n$, $\ell = 1, 2$ and hence we can write

$$\begin{aligned}
M_i &= \max\{r_i w_i, r_i(1 - w_i)\} \\
&= r_i \max\{w_i, 1 - w_i\}.
\end{aligned}$$

Thus, if we want to simulate the intensity component r , in order to get the vector $(\tilde{x}_1, \tilde{x}_2)$ from $r w$ and $r(1 - w)$, we have

$$r_i = \frac{M_i}{\max\{w_i, 1 - w_i\}}.$$

From these considerations we derive the following simulation scheme.

Algorithm 5.1. For $i = 1, \dots, n$:

1. Compute $\tilde{A}(w)$ and plug $\tilde{\beta}$, through $\tilde{\gamma}$, into the expression

$$\tilde{h}(w) = \sum_{j=0}^{k-2} \tilde{\gamma}_j \text{Be}(w; j + 1, k - j - 1);$$

2. Draw w_i from $\tilde{h}(w)$:
 - (a) Generate a random variable $U \sim \text{Unif}(0, 1)$
 - (b) If $U \in [\sum_{h=0}^j \tilde{\gamma}_h, \sum_{h=0}^{j+1} \tilde{\gamma}_h)$, then simulate w from $\text{Be}(w; j + 1, k - j - 1)$;
3. Draw $M_i \sim \text{Frechét}(1, \tilde{\vartheta})$;
4. Compute $r_i = M_i / \max\{w_i, 1 - w_i\}$;
5. Then $\tilde{x}_{1,i} = r_i w_i$ and $\tilde{x}_{2,i} = r_i(1 - w_i)$.

The results obtained for two examples are illustrated in Figures 5.1 and 5.2-???. The first correspond to the dataset presented in Chapter 1 of heavy rainfall in France considering two pairs of stations in the south, at Auvergne, and the Rhône-Alpes and Rhône-Alpes and Corse regions. Due to the distance and morphological differences, the former (top panels) shows much stronger dependence than the latter (bottom panels). The left-panels compares the real data (black points) with the simulated data (red points). It can be seen that the general pattern is reproduced by the simulation scheme and the Q-Qplots on the right-panels fairly confirm that simulated data presents the same dependence structure as the original.

The second example illustrates simulations from the symmetric and asymmetric, Hüsler-Reiss and tilted Dirichlet models, respectively. The estimates of the Pickands dependence function (black line) are displayed on top panels. We consider samples of size $n = 100$ and polynomial orders corresponding to different parameter settings. The grey shades show the simultaneous bootstrap confidence bands. Second-row panels illustrate the estimates of the angular distribution (black line) given by (5.5). Density estimations for $S = 500$ samples from \tilde{h} , are represented with grey lines. Bottom panels show the marginal behaviour of the simulated data, in comparison to the real data, through the Q-Qplots.

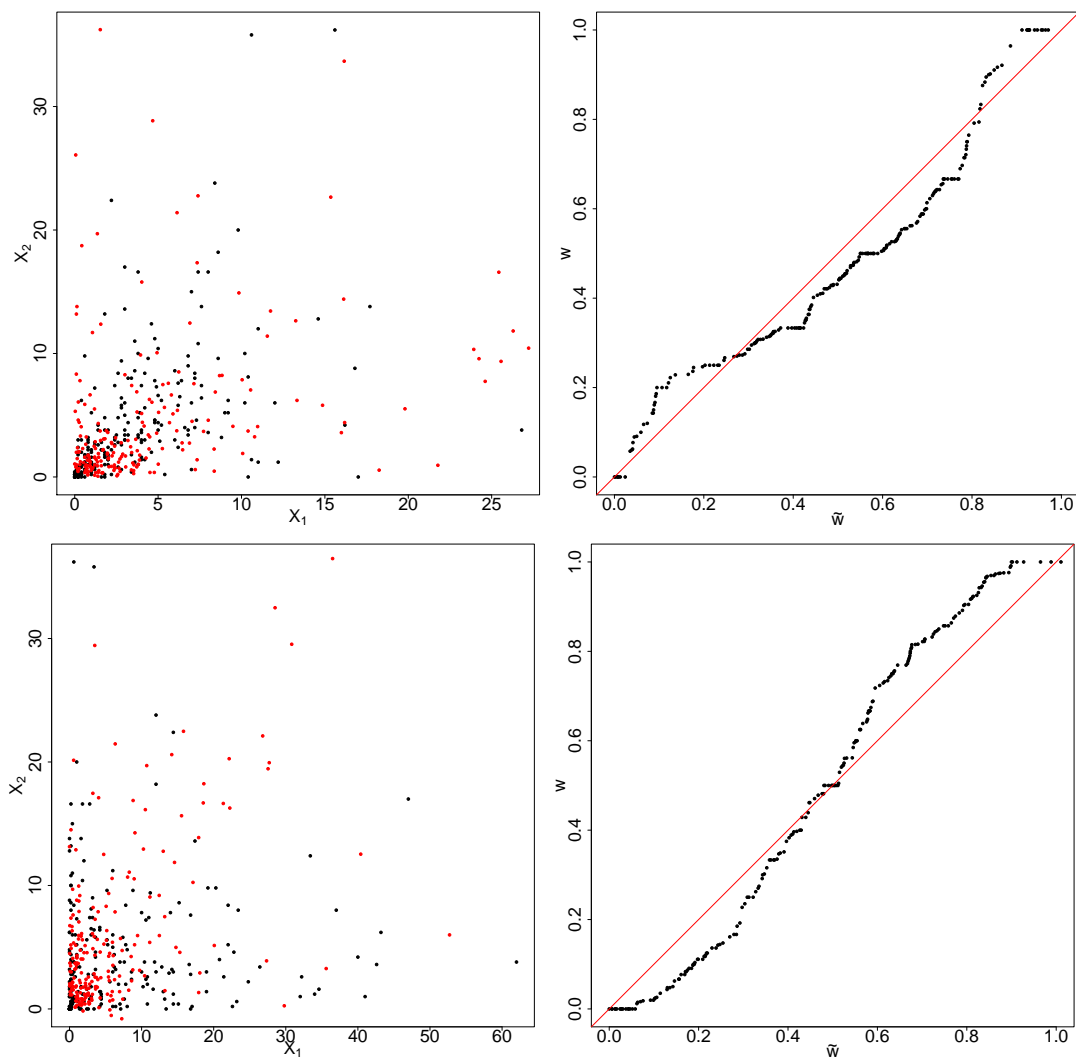


Figure 5.1: (left) Simulated bivariate vectors (red points) and real data (black points) from heavy rainfall dataset and the corresponding Q-Qplots (right).

Strong dependences are well reproduced for all models, while asymmetries are not fully recovered. We do not capture all the mass close to zero and one, which is particularly evident at the tails of the distributions.

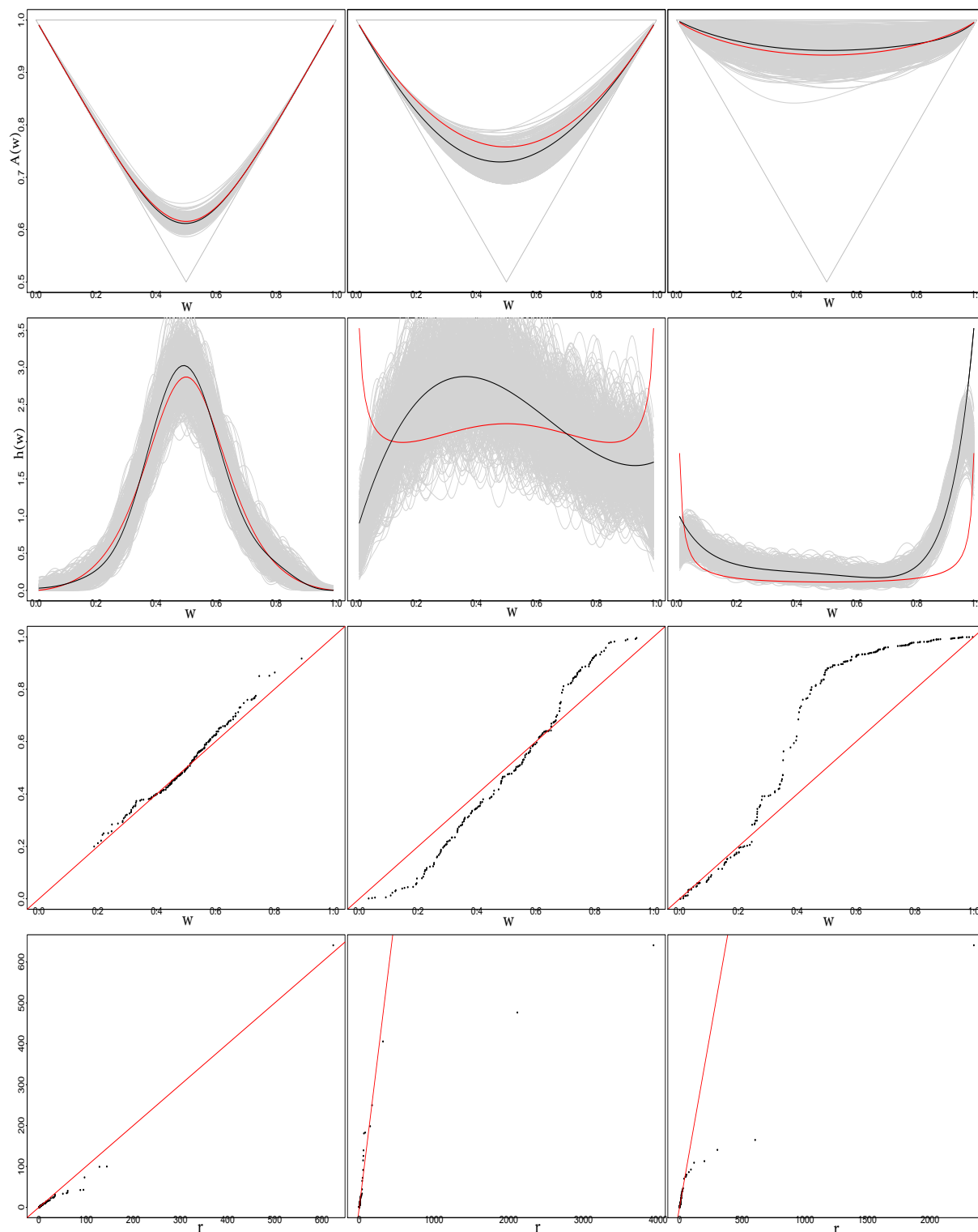


Figure 5.2: Estimates of the Pickands dependence functions $A(w)$ (top panels) and angular density functions $h(w)$ (second-row panels) for the Symmetric Logistic model with strong, mild and weak dependence ($\alpha = 0.3, 0.6, 0.9$) and polynomial orders $k = 23, 5, 15$. Third-row panel shows the Q-Qplots corresponding to simulated angular component \tilde{w} and real angular component w . Bottom panels illustrates the Q-Qplots corresponding to simulated radial component \tilde{r} and real radial component r .

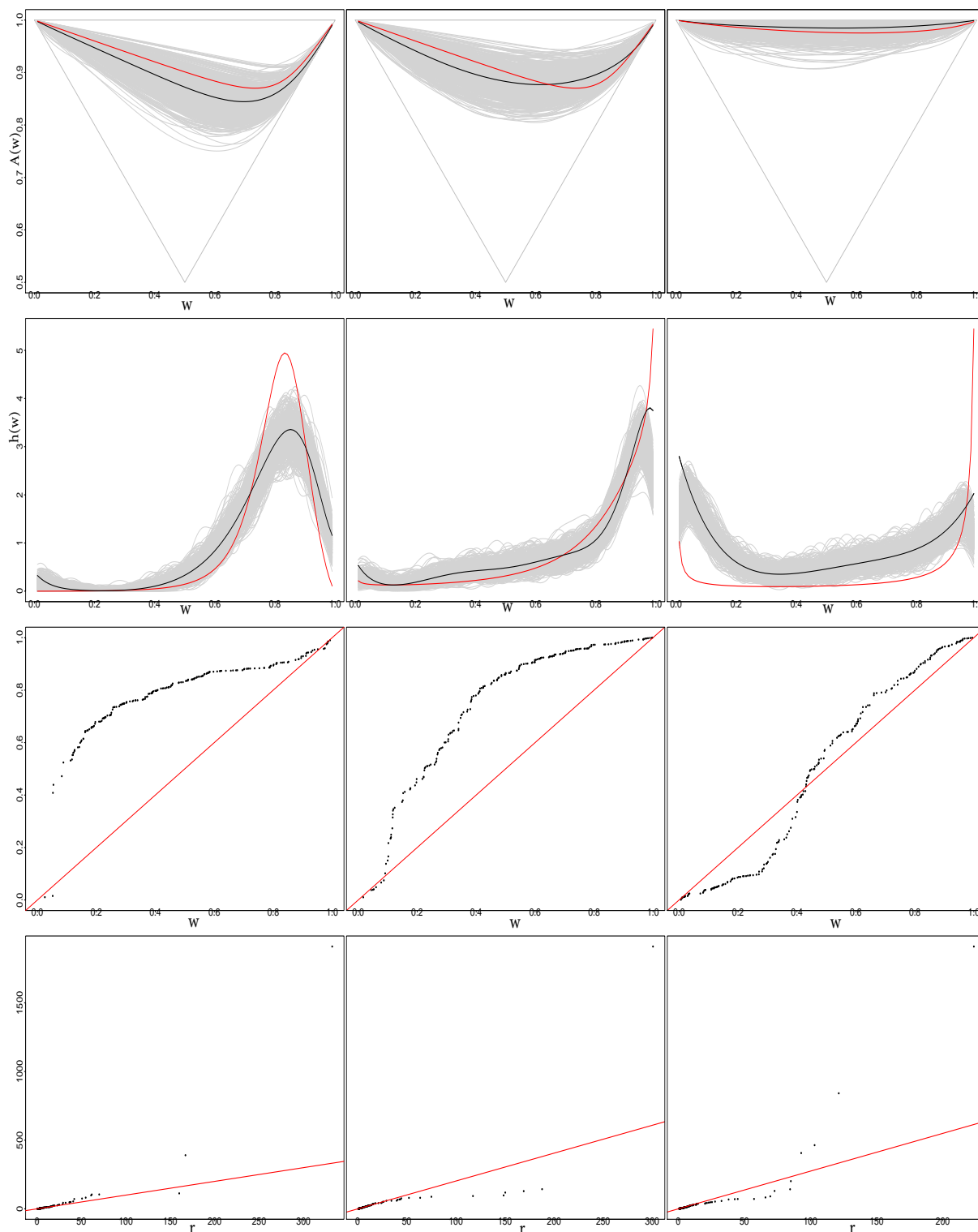


Figure 5.3: Estimates of the Pickands dependence functions $A(w)$ (top panels) and angular density functions $h(w)$ (second-row panels) for the Asymmetric Logistic model with strong, mild and weak dependence ($\alpha = 0.3, 0.6, 0.9, \beta_1 = 0.8, \beta_2 = 0.2$) and polynomial orders $k = 23, 23, 10$. Third-row panel shows the Q-Qplots corresponding to simulated angular component \tilde{w} and real angular component w . Bottom panels illustrates the Q-Qplots corresponding to simulated radial component \tilde{r} and real radial component r .

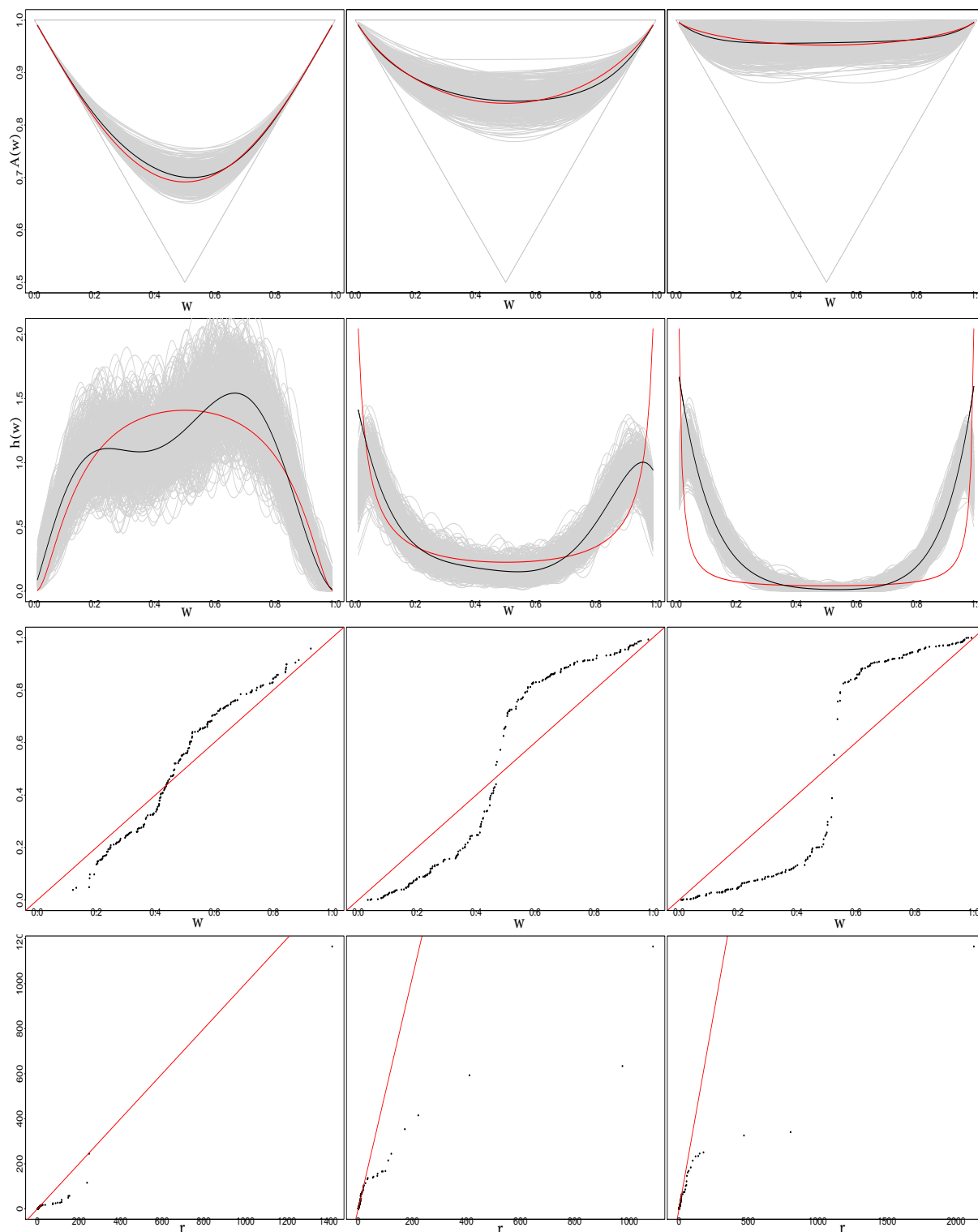


Figure 5.4: Estimates of the Pickands dependence functions $A(w)$ (top panels) and angular density functions $h(w)$ (second-row panels) for the Hüsler-Reiss model with strong, mild and weak dependence ($\lambda = 2, 1, 0.6$) and polynomial orders $k = 12, 12, 12$. Third-row panel shows the Q-Qplots corresponding to simulated angular component \tilde{w} and real angular component w . Bottom panels illustrates the Q-Qplots corresponding to simulated radial component \tilde{r} and real radial component r .

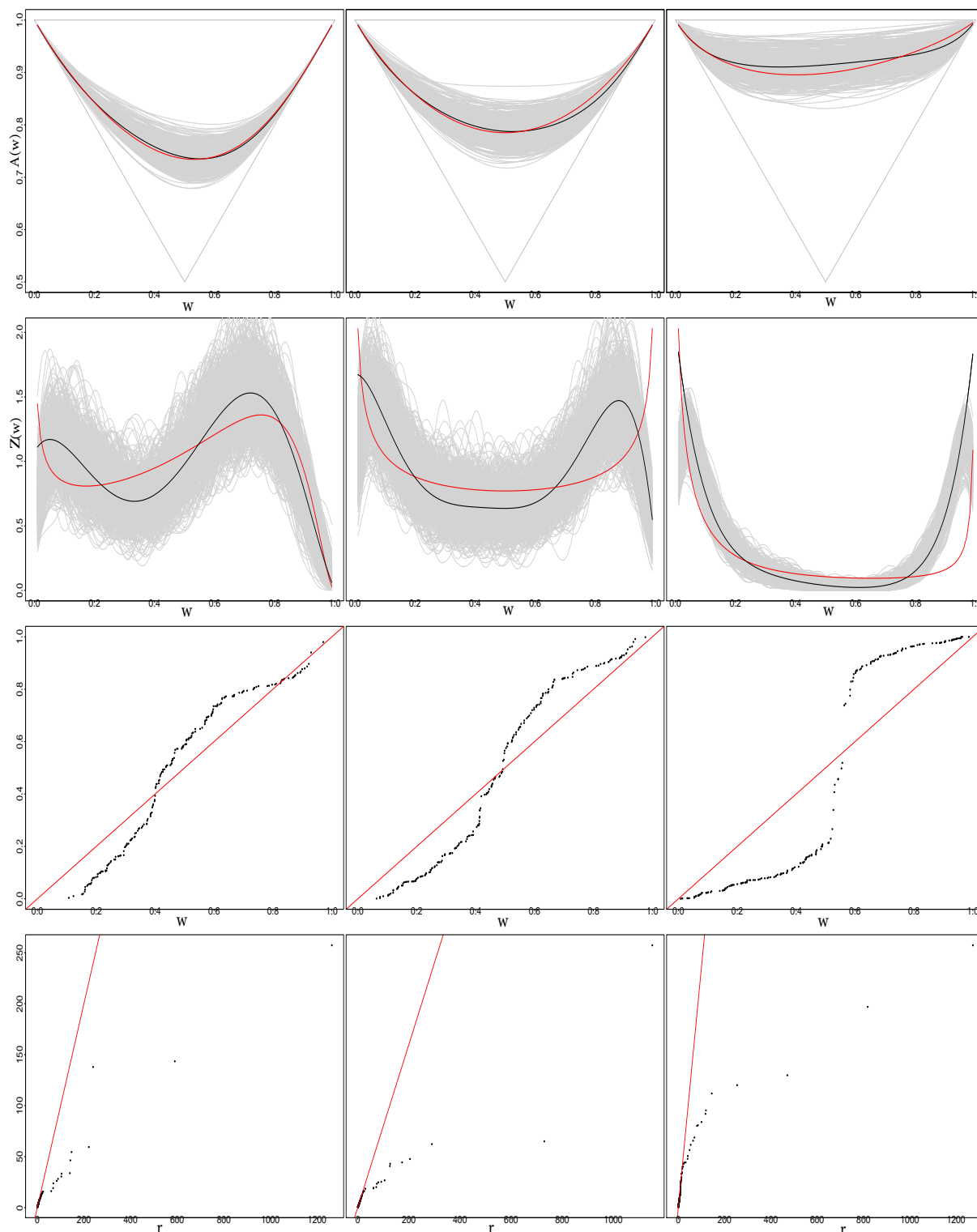


Figure 5.5: Estimates of the Pickands dependence functions $A(w)$ (top panels) and angular density functions $h(w)$ (second-row panels) for the Tilted Dirichlet model with strong, mild and weak dependence ($\alpha = 3, 0.7, 0.1, \beta = 0.7$) and polynomial orders $k = 10, 12, 15$. Third-row panel shows the Q-Qplots corresponding to simulated angular component \tilde{w} and real angular component w . Bottom panels illustrates the Q-Qplots corresponding to simulated radial component \tilde{r} and real radial component r .

Tesi di dottorato "Non-Parametric Inference and Simulation in Extreme Value Theory"
di MARCON GIULIA

discussa presso Università Commerciale Luigi Bocconi-Milano nell'anno 2016

La tesi è tutelata dalla normativa sul diritto d'autore (Legge 22 aprile 1941, n.633 e successive integrazioni e modifiche).

Sono comunque fatti salvi i diritti dell'università Commerciale Luigi Bocconi di riproduzione per scopi di ricerca e didattici, con citazione della fonte.

Chapter 6

Bayesian Inference

*Inside every nonBayesian
there is a Bayesian struggling to get out*

(D. Lindley)

Chapter 4 discusses how to derive an estimator of the Pickands function as a solution of an optimization problem subject to some constraints such that the final estimator fully satisfies the required properties. Such technique allows us to model the existing extremal dependence among a multivariate vector of componentwise maxima and to represent such dependence. We recall that the estimation technique is a two steps procedure: firstly a pilot estimate is computed, secondly it is transformed into a valid estimate. The asymptotic properties of the final estimator are established in the L^2 metric while similar results on L^∞ are still an open question. In addition, for dimension $d > 2$, quite stringent assumptions are required in order to demonstrate the limiting behaviour of the estimator. We also remark that there are not much guidelines for the choice of the polynomial's degree aside from asymptotic arguments that suggest it should be as large as possible. For these reasons, we derive another estimation procedure for inferring the extremal dependence structure, this time in a Bayesian framework.

Our interest still focuses on making inference for extreme values by assessing their extremal dependence. The marginal components are, for the moment, considered as known. As already seen in previous Chapters, there are different approaches to inferring the extremal dependence, such as estimating the stable tail or the Pickands dependence functions or the so-called spectral or angular measure (see Section 6.1), but not both at the same time. This is commonly done nonparametrically and most of the literature considers likelihood-based methods, see for example [Genest and Segers \(2009\)](#), [Marcon et al. \(2015\)](#) and [Boldi and Davison \(2007\)](#). In this Chapter we focus on the bivariate case and propose a nonparametric Bayesian model that allows the simultaneous estimation of the angular measure and the Pickands dependence function through a polynomial in the Bernstein

form. As stated in previous Chapters, in order to derive a valid extremal dependence structure, the angular measure must satisfy what is known as the mean condition; analogously, the Pickands dependence function is constrained by a set of shape restrictions. With our proposal such constraints are addressed by placing a prior on the coefficients of the Bernstein polynomials which gives probability one to the set of valid functions. The prior is completed by a prior on the polynomial degree, making our approach a fully Bayesian nonparametric model. Additionally, the estimation of marginal distributions would be possible through a unique inferential procedure, but this remains an ongoing work. Although the analytical expression of the posterior distribution of the model parameters is unknown, inference is possible via a trans-dimensional MCMC scheme.

6.1 Extremal Dependence

Chapter 2 provides a brief overview of the dependence measures that describe in different forms the dependence structure of multivariate vector of componentwise maxima. For simplicity, we focus here on the dependence modelling in the bivariate case. Assume that $\mathbf{Y} = (Y_1, Y_2)$ is a bivariate random vector of maxima with an asymptotically bivariate max-stable distribution G (e.g., de Haan and Ferreira 2006, Ch. 6). Specifically, the marginal distributions of G are members of the Generalised Extreme Value (GEV) distribution, i.e.

$$F_i(y_i|\mu_i, \sigma_i, \xi_i) = \exp \left\{ - \left(1 + \xi_i \frac{y_i - \mu_i}{\sigma_i} \right)_+^{-1/\xi_i} \right\}, \quad (6.1)$$

where $y_i, \mu_i, \xi_i \in \mathbb{R}$, $\sigma_i > 0$ for $i = 1, 2$ and $(x)_+ = \max(0, x)$.

Furthermore, recall from Chapter 2 that after standardising the marginal distribution to unit Fréchet, i.e. $\mathbb{P}\{Y_i \leq y\} = e^{-1/y}$, for all $y > 0$, the joint distribution G admits the following form

$$G(y_1, y_2) = \exp\{-\ell(1/y_1, 1/y_2)\}, \quad y_1, y_2 > 0, \quad (6.2)$$

where $\ell : [0, \infty)^2 \rightarrow [0, \infty)$, called the stable tail dependence function, is equal to

$$\ell(x_1, x_2) = 2 \int_{\mathcal{S}} \max\{x_1 u, x_2(1-u)\} H(du), \quad x_1, x_2 \geq 0. \quad (6.3)$$

$\mathcal{S} = [0, 1]$ and H , called the spectral or angular distribution, is a probability distribution on \mathcal{S} (e.g., de Haan and Ferreira 2006, Ch. 6). For brevity, hereafter we only use the term angular. The angular distribution, which must satisfy the following condition

$$\int_{\mathcal{S}} u H(du) = \int_{\mathcal{S}} (1-u) H(du) = 1/2, \quad (C3)$$

Clearly $H(\mathcal{S}) = 1$. Formally, for each max-stable distribution G there exists a finite measure H^* , called the angular measure such that the mean conditions $\int_{\mathcal{S}} u H^*(du) = \int_{\mathcal{S}} (1-u) H^*(du) = 1$ are satisfied and this implies $H^*(\mathcal{S}) = 2$, see [de Haan and Ferreira \(2006, Ch. 6\)](#). The angular distribution is obtained by the normalization $H := H^*/H^*(\mathcal{S})$. As a consequence, any probability distribution satisfying (C3) provides a valid bivariate extreme value distribution. Consider the partition $(\{0\}, \overset{\circ}{\mathcal{S}}, \{1\})$ of the set \mathcal{S} , where $\overset{\circ}{\mathcal{S}} = (0, 1)$. The angular distribution H can place mass on the subset $\overset{\circ}{\mathcal{S}}$ as well as on the vertices $\{0\}$ and $\{1\}$ and if H is absolutely continuous on $\overset{\circ}{\mathcal{S}}$, then we have

$$H(u) = \mathbb{I}_{\mathcal{S}}(u) h_0 + \mathbb{I}_{\overset{\circ}{\mathcal{S}}}(u) \overset{\circ}{H}(u) + \mathbb{I}_1(u) h_1, \quad u \in \mathcal{S} \quad (6.4)$$

where $\mathbb{I}_{\mathcal{S}}$ is the indicator function of the set \mathcal{S} ,

$$h(u) = \frac{d\overset{\circ}{H}(u)}{du}, \quad \overset{\circ}{H}(u) = \int_0^u h(t)dt,$$

and

$$h_1 = 1/2 - \int_0^1 u h(u) du, \quad h_0 = 1/2 - \int_0^1 (1-u) h(u) du. \quad (6.5)$$

The last expressions, regarding the point masses on the vertices of the simplex, are derived from identity (C3).

The stable tail dependence function (6.3) has the following properties:

- a) it is homogeneous of order 1, that is $\ell(vx_1, vx_2) = v\ell(x_1, x_2)$ for all $v, x_1, x_2 > 0$;
- b) $\ell(x, 0) = \ell(0, x) = x$ for all $x > 0$;
- c) it is continuous and convex, i.e. $\ell(v(x_1, x_2) + (1-v)(x'_1, x'_2)) \leq v\ell(x_1, x_2) + (1-v)\ell(x'_1, x'_2)$ for all $x_1, x_2, x'_1, x'_2 \geq 0$ and $v \in \mathcal{S}$;
- d) $\max(x_1, x_2) \leq \ell(x_1, x_2) \leq x_1 + x_2$ for all $x_1, x_2 \geq 0$ ([de Haan and Ferreira 2006, Ch. 6](#)).

The lower and upper bounds of the last condition represent the cases of complete dependence and complete independence, respectively. By the homogeneity of ℓ we have that for all $x_1, x_2 \geq 0$

$$\ell(x_1, x_2) = (x_1 + x_2)A(w), \quad A(w) = 2 \int_{\mathcal{S}} \max\{w(1-u), (1-w)u\} H(du), \quad (6.6)$$

where $w = x_2/(x_1 + x_2) \in \mathcal{S}$, and the function A is the Pickands dependence function. From the properties of ℓ it follows that A satisfies the conditions (C1)–(C2) introduced in Section 2.2.1, namely:

- (C1) $A(w)$ is convex, i.e., $A(aw + (1-a)w') \leq aA(w) + (1-a)A(w')$, for $a \in \mathcal{S}$ and $w, w' \in \mathcal{S}$;

(C2) $A(w)$ has lower and upper bounds

$$1/2 \leq \max(w, 1 - w) \leq A(w) \leq 1,$$

for any $w \in \mathcal{S}$.

In condition (C2), the lower and upper bounds represent the cases of complete dependence and complete independence, respectively.

Let \mathcal{A} be the family of functions $A : \mathcal{S} \rightarrow [1/2, 1]$ that satisfy the above conditions, then any Pickands dependence function belongs to the class \mathcal{A} (Falk, Hüsler, and Reiss, 2010, Ch. 4). In addition, there is a one to one relationship between A and H , thus if A is a twice differentiable function on $\mathring{\mathcal{S}}$, then

$$A(w) = 1 + 2 \int_0^w H(u) du - w$$

and therefore $A'(w) = -1 + 2H(w)$ and $A''(w) = 2h(w)$. Finally, from (6.5) it follows that the mass on the vertices can be written in terms of the Pickands dependence function as:

$$h_0 = \frac{1 + A'(0)}{2} \quad \text{and} \quad h_1 = \frac{1 - A'(1)}{2}. \quad (6.7)$$

6.2 Bayesian Modelling of H and A

The basic idea is to define the angular distribution function and the Pickands dependence function through polynomial functions of the form $\sum_{j=0}^k a_j b_j(x)$, where a_j and $b_j(\cdot)$, with $x \in \mathcal{S}$, $j = 0, \dots, k$, are coefficients and basis functions respectively. Let \mathcal{P}_k , \mathcal{H} and \mathcal{A} be the spaces of order k polynomials, angular distributions and Pickands dependence functions, respectively.

In connection with the previous section we denote, as before, here $p_0 = H(\{0\})$ and $p_1 = H(\{1\})$ so that we have $\mathring{H}(1) = 1 - p_0 - p_1$ and we recall that $\mathring{H}(0) = 0$.

We focus on the class of bivariate Bernstein polynomial basis on \mathcal{S} of degree k , denoted by $b_j(\cdot; k)$ with $k = 1, 2, \dots$ (see Remarks 4.4 and 4.6). Then, we start by representing an angular distribution \mathring{H} related to (6.3) as a polynomial of order k in Bernstein form, for some $k = 1, 2, \dots$, i.e.

$$\mathring{H}_k(u) := \sum_{j=0}^k \eta_j b_j(u; k), \quad u \in \mathcal{S}. \quad (6.8)$$

By taking the first derivative of \mathring{H}_k with respect to u we obtain

$$\mathring{H}'_k(u) = \sum_{j=0}^{k-1} (\eta_{j+1} - \eta_j) \text{Be}(u; j+1, k-j) =: h_{k-1}(u), \quad u \in \mathcal{S}, \quad (6.9)$$

where $\text{Be}(a, b)$ indicates the beta density function with parameters (a, b) and so (6.9) turns out to be a finite linear combination of beta densities (see Remark 4.6). It can be seen that

$$\int_0^1 h_{k-1}(u) du = \sum_{j=0}^{k-1} (\eta_{j+1} - \eta_j) = \eta_k - \eta_0 \quad (6.10)$$

and

$$\int_0^1 u h_{k-1}(u) du = \sum_{j=0}^{k-1} (\eta_{j+1} - \eta_j) \frac{j+1}{k+1}, \quad (6.11)$$

$$\int_0^1 (1-u) h_{k-1}(u) du = \sum_{j=0}^{k-1} (\eta_{j+1} - \eta_j) \frac{k-j}{k+1}. \quad (6.12)$$

In order to define a valid angular distribution H , the vector of coefficients (η_0, \dots, η_k) must be such that the integrals (6.10)–(6.12) equal $1 - p_0 - p_1$, $1/2 - p_1$ and $1/2 - p_0$, respectively. These conditions hold under the following restrictions:

$$(R4) \quad \eta_0 \leq \eta_1 \leq \dots \leq \eta_k;$$

$$(R5) \quad \eta_0 = 0 \text{ and } \eta_k = 1 - p_0 - p_1;$$

$$(R6) \quad \eta_0 + \dots + \eta_k = (k+1)(1/2 - p_0);$$

Similarly, for any $r = 1, 2, \dots$ we define the Pickands dependence function as an r -th degree polynomial in Bernstein form, i.e.

$$A_r(w) := \sum_{j=0}^r \beta_j b_j(w; r), \quad w \in \mathcal{S}. \quad (6.13)$$

Function (6.13) must satisfy conditions (C1)–(C2) in order to be a genuine Pickands dependence function. This is guaranteed if the restrictions (R1)–(R3) on the polynomial coefficients explored in Section 4.2 are verified. For the bivariate case, they simplify to

$$(R1) \quad \beta_0 = \beta_r = 1;$$

$$(R2) \quad \beta_1 = \beta_{r-1} = 1 - 1/r;$$

$$(R3) \quad \beta_{j+2} - 2\beta_{j+1} + \beta_j \geq 0, \quad j = 0, \dots, r-2.$$

Notice that condition (R2) slightly differs from the one stated in 4.2.2 because, here, the equality ensures that the relationship with the angular density representation is valid.

Taking the second derivative of A_r with respect to w we obtain

$$A''_{r-2}(w) = \sum_{j=0}^{r-2} (\beta_{j+2} - 2\beta_{j+1} + \beta_j) \text{Be}(w; j+1, r-j-1) =: h_{r-2}(w), \quad w \in \mathcal{S}. \quad (6.14)$$

It can be verified that

$$\int_0^1 h_{r-2}(w)dw = r(2 - \beta_1 - \beta_{r-1}) \equiv 2 - \{1 + A'_{r-1}(0)\} - \{1 - A'_{r-1}(1)\}, \quad (6.15)$$

where A'_{r-1} is the first derivative of (6.13) with respect to w , and corresponds to a polynomial of order $r - 1$. Then,

$$\int_0^1 wh_{r-2}(w)dw = r(1 - \beta_{r-1}) \equiv 1 - \{1 - A'_{r-1}(1)\}, \quad (6.16)$$

$$\int_0^1 (1 - w)h_{r-2}(w)dw = r(1 - \beta_1) \equiv 1 - \{1 + A'_{r-1}(0)\} \quad (6.17)$$

and these results are consistent with the theoretical arguments discussed in Section 6.1. A benefit of defining the angular distribution function and the Pickands dependence function through polynomial functions is given by the following result.

Proposition 6.1.

i) Let H be defined by (6.4) with point masses $0 \leq p_0, p_1 < 1, p_0 + p_1 < 1$ and \mathring{H} given by (6.8). Then A is of the form (6.13) with $r = k + 1$ and

$$\beta_t = \frac{2}{k+1} \left(\sum_{i=0}^{t-1} \eta_i + t p_0 + \frac{k+1-t}{2} \right), \quad t = 0, \dots, k+1 \quad (6.18)$$

Conversely, define A as (6.13). Then H takes the form (6.4), where \mathring{H} is given by expression 6.8 with $k = r - 1$ and coefficients:

$$\eta_t = \frac{k+1}{2} \left(\beta_{t+1} - \beta_t + \frac{1-2p_0}{k+1} \right), \quad t = 0, \dots, k. \quad (6.19)$$

for any $0 \leq p_0 < 1$ and $p_1 = 1 - p_0 - \int_0^1 \mathring{H}(du)$.

ii) Restrictions (R1)-(R3) are satisfied and A meets condition (C1)-(C2), if and only if restrictions (R4)-(R6) are verified and H meets conditions (C3).

This result tells us that the representation of the extremal dependence obtained by modelling the angular distribution with Bernstein polynomials coincides with the analogous modelling of the Pickands dependence function. This is because the eta coefficients in equation (6.8) can be re-expressed in terms of the beta coefficients of equation (6.13) and when the restrictions (R1)-(R3) are satisfied, so are (R4)-(R6). The crucial consequence is that there is not need to choose between different characterizations of the dependence.

6.3 Model Formulation and Algorithms

In this section two approaches are described for modelling the extremal dependence. First, for k fixed, we consider the estimation of the coefficients $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{k+1})$ of the representation of the Pickands dependence function as the polynomial in Bernstein form (6.13). An MCMC simulation scheme for posterior inference is proposed in Algorithm 6.1. Secondly, we move to the coefficients $\boldsymbol{\eta} = (\eta_0, \dots, \eta_k)$ of the angular distribution in (6.8) and extend the study by considering k to be random and learning about it from the data (see algorithm 6.3).

The first idea is related to the approximated projection method proposed in Chapter 4, but we substitute the use of the preliminary estimate \hat{A}_n with the definition of a uniform prior on $\boldsymbol{\beta}$, satisfying all necessary constraints,

$$\Pi(\boldsymbol{\beta}) \propto \mathbb{I}\{\boldsymbol{\beta} \in \mathcal{B}_k\}$$

where \mathcal{B}_k is the set of all $\boldsymbol{\beta} \in \mathcal{S}^{k+1}$ which satisfy constraints (C1)–(C2), for a fixed order of the polynomial k .

Regarding the likelihood, we consider two approaches. One uses the cumulative distribution function $G(y_1, y_2; \pi)$, and define a pseudo-likelihood; the other is based on the density distribution function. The pseudo-likelihood can be rewritten in terms of the coefficients β_j as

$$\begin{aligned} L^*(\mathbf{y}; k, \boldsymbol{\beta}, \pi) &= \prod_{i=1}^n G(y_{1,i}, y_{2,i}; \pi) = \exp \left\{ - \sum_{i=1}^n \left(\frac{1}{y_{1,i}} + \frac{1}{y_{2,i}} \right) A(w_i) \right\} \\ &= \exp \left\{ - \sum_{i=1}^n \left(\frac{1}{y_{1,i}} + \frac{1}{y_{2,i}} \right) \sum_{j=0}^{k+1} \beta_j b_j(w_i; k+1) \right\}, \end{aligned}$$

where $\pi = (\mu_1, \xi_1, \sigma_1, \mu_2, \xi_2, \sigma_2)$ denotes the vector of known parameters for the marginal Fréchet distributions of

$$Y_1 = \left(1 + \frac{\xi_1(X_1 - \mu_1)}{\sigma_1} \right)_+^{1/\xi_1} \quad \text{and} \quad Y_2 = \left(1 + \frac{\xi_2(X_2 - \mu_2)}{\sigma_2} \right)_+^{1/\xi_2}.$$

This corresponds to conditioning on the event $(Y_1 \leq y_1, Y_2 \leq y_2)$ rather than on the observations y_1, y_2 themselves. The loss of information is justified by the gain in computational time, and avoiding the difficulties involved in calculating the likelihood function.

The probability density function of the joint distribution (6.2) is equal to

$$g(y_1, y_2) = G(y_1, y_2) \left[\frac{\{A(w) - w A'(w)\} \{A(w) + (1-w) A'(w)\}}{(y_1 y_2)^2} + \frac{A''(w)}{(y_1 + y_2)^3} \right],$$

for all $y_1, y_2 > 0$, and is obtained by

$$g(y_1, y_2) = |J(y_1, y_2)| \frac{\partial^2}{\partial x_1 \partial x_2} G(1/x_1, 1/x_2) \Big|_{x_1=1/y_1, x_2=1/y_2},$$

where $J(y_1, y_2) = (y_1 y_2)^{-2}$ and $A(w)$ is given in the expression (6.6). Let $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ be iid copies of a bivariate max-stable random vector with unit Fréchet marginal distributions. If the dependence structure of joint distribution is expressed through the Pickands dependence function and the latter is represented by (6.13), then the likelihood function is equal to

$$\begin{aligned}
L(\mathbf{y}; k, \boldsymbol{\beta}, \pi) &= \prod_{i=1}^n g(y_{1,i}, y_{2,i}; \pi) \\
&= \prod_{i=1}^n \frac{G(y_{1,i}, y_{2,i})}{(r_i w_i)^2} \left\{ \frac{[A(w_i) - w_i A'(w_i)][A(w_i) + (1 - w_i) A'(w_i)]}{(r_i (1 - w_i))^2} + \frac{w_i^2}{r_i} A''(w_i) \right\} \\
&= \exp \left\{ - \sum_{i=1}^n \left(\frac{1}{y_{1,i}} + \frac{1}{y_{2,i}} \right) \sum_{j=0}^{k+1} \beta_j b_j(w_i; k+1) \right\} \prod_{i=1}^n \left\{ \frac{1}{(r_i w_i)^2} \right. \\
&\quad \times \left[\sum_{j=0}^{k+1} \beta_j b_j(w_i; k+1) - w_i (k+1) \sum_{j=0}^k (\beta_{j+1} - \beta_j) b_j(w_i; k) \right] \\
&\quad \times \left[\sum_{j=0}^{k+1} \beta_j b_j(w_i; k+1) + (1 - w_i) (k+1) \sum_{j=0}^k (\beta_{j+1} - \beta_j) b_j(w_i; k) \right] \frac{1}{(r_i (1 - w_i))^2} \\
&\quad \left. + \frac{w_i^2}{r_i} k (k+1) \sum_{j=0}^{k-1} (\beta_{j+2} - 2\beta_{j+1} + \beta_j) b_j(w_i; k-1) \right\},
\end{aligned}$$

where $w = y_1/(y_1 + y_2)$ and $r = y_1 + y_2$.

In other words, we base the inference on the *pseudo*-posterior distribution function

$$\Pi^{*n}(\boldsymbol{\beta} | \mathbf{Y}_1 \leq \mathbf{y}_1, \mathbf{Y}_2 \leq \mathbf{y}_2, k, \pi) \propto \Pi(\boldsymbol{\beta}) L^*(\mathbf{y}; k, \boldsymbol{\beta}, \pi),$$

and similarly, on the posterior distribution,

$$\Pi^n(\boldsymbol{\beta} | \mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2, k, \pi) \propto \Pi(\boldsymbol{\beta}) L(\mathbf{y}; k, \boldsymbol{\beta}, \pi).$$

When considering the pseudo-likelihood, MCMC inference is greatly simplified, as the full conditionals for each β_j correspond to truncated exponentials of the form

$$\Pi^n(\beta_j | \mathbf{Y}_1 \leq \mathbf{y}_1, \mathbf{Y}_2 \leq \mathbf{y}_2, \pi) \propto \exp\{-\lambda_j \beta_j\} \mathbb{1}\{\beta_j \in [a_j, b_j]\}$$

for $j = 1, \dots, k - 1$, with $\lambda_j = \sum_{i=1}^n (1/y_{1,i} + 1/y_{2,i}) b_j(w_i; k)$. Notice that $\beta_0, \beta_1, \beta_k$ and β_{k+1} are not involved in the algorithm because they are constrained by (R1) and (R2). Lower and upper bounds, a_j and b_j , are determined, depending on the value of $j = 2, \dots, k - 2$, as:

$$a_j = \frac{\beta_{j-1} + \beta_{j+1}}{2}$$

$$b_j = \max\{2\beta_{j-1} - \beta_{j-2}, 2\beta_{j+1} - \beta_{j+2}\}.$$

The algorithm, based on the pseudo-likelihood, can be synthesised in the following way,

Algorithm 6.1.

1. Define a starting value, $\boldsymbol{\beta}^{(0)} \in \mathcal{B}_k$;
2. For $s = 1, \dots, S$, update each beta sequentially from its full-conditional distribution:
 - (a) Sample $\beta_j^{(s)} \sim t\text{Exp}(\lambda_j; a_j, b_j)$;

Here, $t\text{Exp}(\lambda_j, a_j, b_j)$ denotes an exponential distribution with parameter λ_j , truncated on the interval $[a_j, b_j]$. In this way, we obtain a new Markov chain state $\boldsymbol{\beta}^{(s)}$ (notice that (a_j, b_j) depend on the previously updated $\beta_{j-1}^{(s-1)}, \beta_{j-1}^{(s-1)}, \dots$). After an appropriate burn-in period, this will reproduce a sample $\boldsymbol{\beta}^{(1)}, \dots, \boldsymbol{\beta}^{(S)}$ from $\Pi^n(\boldsymbol{\beta}|\mathbf{Y}, k, \pi)$.

On the other hand, when we consider the likelihood, the algorithm is given by:

Algorithm 6.2.

1. Define a starting value, $\boldsymbol{\beta}^{(0)} \in \mathcal{B}_k$;
2. For $s = 1, \dots, S$, sample $\boldsymbol{\beta}^s \sim \Pi^n(\boldsymbol{\beta}|\mathbf{y}, k)$. This requires a Metropolis-Hastings:
 - (a) Generate a candidate $\boldsymbol{\beta}^*$ for the next sample by picking from the proposal distribution $\Pi(\boldsymbol{\beta}^*|\boldsymbol{\beta}^{s-1})$;
 - (b) Calculate the acceptance ratio $p = \Pi^n(\boldsymbol{\beta}^*|\mathbf{y}, k)/\Pi^n(\boldsymbol{\beta}^{s-1}|\mathbf{y}, k)$, which will be used to decide whether to accept or reject the candidate;
 - (c) Accept the candidate, by setting $\boldsymbol{\beta}^s = \boldsymbol{\beta}^*$, with probability p ; if the candidate is rejected, set $\boldsymbol{\beta}^s = \boldsymbol{\beta}^{s-1}$, instead.
3. Return to step 2 and repeat to produce a Markov chain of length S.

Figure 6.1 shows an application of algorithm 6.1 where the symmetric logistic model is considered with a mild dependence ($\alpha = 0.6$) and the polynomial degree is set to $k = 10$. The number of iterations involved in the MCMC is $S = 5000$ and the first 3000 observations form the burn-in period. In the left panel, the median curve (black line) of the MCMC posterior sampled is shown along with the 95% credible bands (light grey region, computed by (4.16)). Boxplots of posterior samples of each coefficient β_j are illustrated in the right-side panel, together with the “true” coefficients $A(j/k)$ (red points), $j = 0, \dots, k$.

The following simulation study shows the performance of our proposed methods with respect to different dependence structures. Figures 6.2–6.5 illustrate the Pickands dependence function (top panels) and the angular distribution function (bottom panels) for the symmetric and asymmetric logistic, Hüsler-Reiss and Tilted Dirichlet models, with strong, mild and weak dependence. The red line is the true function; the black line is the mean curve of the MCMC posterior sample; the green line is the mean curve of the MCMC *pseudo*-posterior sample. Closed to the limiting cases of strong and weak dependence, the loss of information in the likelihood is partly, while it becomes much more evident when the dependence is mild. The method proposed provides good results for the symmetric logistic model and almost for the Hüsler-Reiss and Tilted Dirichlet models. Instead, the mean curves of the MCMC posterior distributions for the asymmetric logistic model, do not capture the asymmetry. However, a convenience of using this approach is the fast computational time; for example, it takes only 10 seconds to run $S = 5000$ iterations when $k = 25$, with respect to a processor Intel(R) Core(TM) i5-4440 CPU 3.10 GHz.

We now introduce the second approach, which involves the angular distribution. In this case we estimate also to the order of the polynomial, by defining a prior on k . Notice that, in this case, a trans-dimensional MCMC is needed to simulate from the posterior distribution of the model parameters, because the size of the vector of coefficients, at each updating step, depends on the value of k .

In order to construct a prior on the space \mathcal{H} of valid angular distributions, we consider the subspace $\mathcal{H}_p \in \mathcal{H}$ of polynomial angular distributions. Then, by using the representation in (6.8), any $H \in \mathcal{H}_p$ can be written as

$$H(w) := \sum_{j=0}^k \eta_j b_j(w; k),$$

for some $k = 1, 2, \dots$ where the vector $\boldsymbol{\eta}_{0:k} \in \mathcal{E}_k$, which is the set of all $\boldsymbol{\eta}_{0:k} \in \mathcal{S}^{k+1}$ satisfying all conditions (R4)-(R6), analogously to \mathcal{B}_k .

The prior on \mathcal{H}_p is induced by a prior on the vector of coefficients $\boldsymbol{\eta}_k \equiv \boldsymbol{\eta}_{0:k}$ and k ,

$$\Pi(\boldsymbol{\eta}_k, k) = \Pi(\boldsymbol{\eta}_k | k) \Pi(k).$$

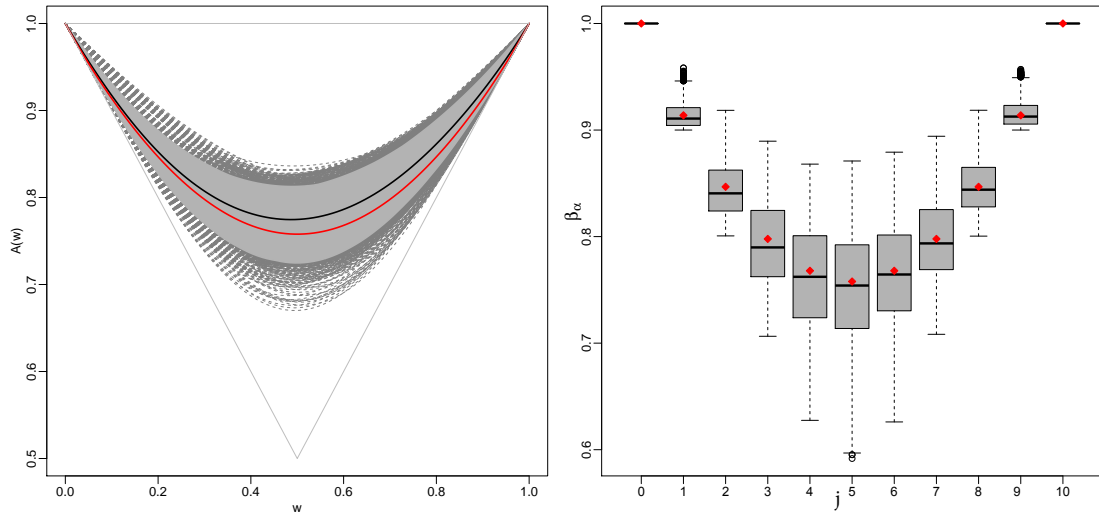


Figure 6.1: (left) Bivariate Pickands dependence function (red line). The black line is the median curve of the MCMC sample from the *pseudo*-posterior distribution, along with 95% credible bands (light grey region). The boxplots in the right-panel show the posterior samples of each coefficient β_j (right), and true coefficients $A(j/k)$ (red), $j = 0, \dots, k$.

Since \mathcal{H}_p is dense in \mathcal{H} , the induced prior will have full support as long as $\Pi(k)$ has full support on \mathbb{Z} and $\Pi(\boldsymbol{\eta}_k|k)$ has full support on $\mathcal{E}_k, \forall k$.

Now, we define a prior on the polynomial order. Specifically, let $k = \tilde{k} + 1$ and define,

$$\Pi(\tilde{k}) = \text{Poisson}(\zeta) \quad (6.20)$$

where the rate of the Poisson distribution ζ can be chosen empirically based on the data correlation. In particular, when $\zeta = 0$ then $k = 1$, and as ζ goes to 1, k goes to infinity. We propose using $\zeta = \exp(-|\rho(y_1, y_2)| + 2.5)$, with ρ denoting the sample correlation of the data.

For the prior on the vector of coefficients $\boldsymbol{\eta}_k$, we consider,

$$\begin{aligned} \Pi(\boldsymbol{\eta}_k|k) &= \mathbb{I}_{\{\eta_0=0\}} \mathbb{I}_{\{\eta_k=1-p_0-p_1\}} \times \prod_{j=1}^{k-1} \left[\frac{1}{k-j} \left\{ \frac{k-1}{2} - kp_0 + p_1 - \sum_{i=1}^{j-1} \eta_i \right\} \right. \\ &\quad \left. - \max \left\{ \eta_{j-1}; \frac{2j+1-k(2p_0+1)+2p_1}{2} - \sum_{i=1}^{j-1} \eta_i \right\} \right]^{-1} \mathbb{I}_{\{\eta_j \in \mathcal{E}_j\}}. \end{aligned} \quad (6.21)$$

Notice that this is a well defined density function for $\boldsymbol{\eta}_k$, for each k , on the set \mathcal{E}_k of vector satisfying all required conditions.

Our goal is to draw samples from the posterior distribution $\Pi^n(k, \boldsymbol{\eta}_k|\mathbf{Y}, \varpi)$, where $\varpi = (p_0, p_1, \mu_1, \sigma_1, \xi_1, \mu_2, \sigma_2, \xi_2)$ is the vector of known parameters for the Fréchet marginal distributions of Y_1 and Y_2 , and p_0 and p_1 are the point masses, here assumed known.

Regarding the likelihood and *pseudo*-likelihood, they can be rewritten in terms of the coefficients η_j , from the relationship between $\boldsymbol{\beta}_{k+1} = (\beta_0, \dots, \beta_{k+1})$ and $\boldsymbol{\eta}_k = (\eta_0, \dots, \eta_k)$ proved in (6.19), as:

$$\begin{aligned}
L(\mathbf{y}; k, \boldsymbol{\eta}_k, \varpi) &= \exp \left\{ - \sum_{i=1}^n \left(\frac{1}{y_{1,i}} + \frac{1}{y_{2,i}} \right) \sum_{j=0}^{k+1} \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 + \frac{k+1-j}{k+1} \right) \right. \\
&\times b_j(w_i; k+1) \left. \right\} \prod_{i=1}^n \frac{1}{(r_i w_i)^2} \left\{ \left[\sum_{j=0}^{k+1} \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 + \frac{k+1-j}{k+1} \right) \right. \right. \\
&\times b_j(w_i; k+1) - w_i (k+1) \sum_{j=0}^k \left\{ \left(\frac{2}{k+1} \sum_{i=0}^j \eta_i + \frac{2(j+1)}{k+1} p_0 + \frac{k-j}{k+1} \right) \right. \\
&- \left. \left. \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 + \frac{k+1-j}{k+1} \right) \right\} b_j(w_i; k) \right] \\
&\times \left[\sum_{j=0}^{k+1} \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 + \frac{k+1-j}{k+1} \right) b_j(w_i; k+1) \right. \\
&+ (1-w_i) (k+1) \sum_{j=0}^k \left\{ \left(\frac{2}{k+1} \sum_{i=0}^j \eta_i + \frac{2(j+1)}{k+1} p_0 + \frac{k-j}{k+1} \right) \right. \\
&- \left. \left. \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 + \frac{k+1-j}{k+1} \right) \right\} b_j(w_i; k) \right] \frac{1}{(r_i (1-w_i))^2} \\
&+ \frac{w_i^2}{r_i} k (k+1) \sum_{j=0}^{k-1} \left\{ \left(\frac{2}{k+1} \sum_{i=0}^{j+1} \eta_i + \frac{2(j+2)}{k+1} p_0 + \frac{k-j-1}{k+1} \right) \right. \\
&- 2 \left(\frac{2}{k+1} \sum_{i=0}^j \eta_i + \frac{2(j+1)}{k+1} p_0 + \frac{k-j}{k+1} \right) + \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 \right. \\
&\left. \left. + \frac{k+1-j}{k+1} \right) \right\} b_j(w_i; k-1) \left. \right\}.
\end{aligned}$$

and

$$\begin{aligned}
L^*(\mathbf{y}; k, \boldsymbol{\eta}_k, \varpi) &= \exp \left\{ - \sum_{i=1}^n \left(\frac{1}{y_{1,i}} + \frac{1}{y_{2,i}} \right) \sum_{j=0}^{k+1} \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 \right. \right. \\
&\left. \left. + \frac{k+1-j}{k+1} \right) b_j(w_i; k+1) \right\}.
\end{aligned}$$

Notice that $L(\mathbf{y}; \boldsymbol{\eta}_k, k, \varpi)$ is much simpler than $L^*(\mathbf{y}; \boldsymbol{\eta}_k, k, \varpi)$.

The posterior distribution, and similarly the *pseudo*-posterior distribution, is then

given by

$$\begin{aligned}
\Pi^n(k, \boldsymbol{\eta}_k | \mathbf{Y}, \varpi) &\propto \Pi(k) \times \Pi(\boldsymbol{\eta}_k | k, \varpi) \times L(\mathbf{y}; k, \boldsymbol{\eta}_k, \varpi) \\
&\propto \mathbb{I}_{\{\eta_0=0\}} \mathbb{I}_{\{\eta_k=1-p_0-p_1\}} \times \prod_{j=1}^{k-1} \left[\frac{1}{k-j} \left\{ \frac{k-1}{2} - k p_0 + p_1 - \sum_{i=1}^{j-1} \eta_i \right\} \right. \\
&\quad \left. - \max \left\{ \eta_{j-1}; \frac{2j+1-k(2p_0+1)+2p_1}{2} - \sum_{i=1}^{j-1} \eta_i \right\} \right]^{-1} \mathbb{I}_{\{\eta_j \in \mathcal{E}_j\}} \\
&\times \exp \left\{ - \sum_{i=1}^n \left(\frac{1}{y_{1,i}} + \frac{1}{y_{2,i}} \right) \sum_{j=0}^{k+1} \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 + \frac{k+1-j}{k+1} \right) \right. \\
&\quad \left. \times b_j(w_i; k+1) \right\} \prod_{i=1}^n \frac{1}{(r_i w_i)^2} \left\{ \left[\sum_{j=0}^{k+1} \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 + \frac{k+1-j}{k+1} \right) \right. \right. \\
&\quad \left. \left. \times b_j(w_i; k+1) - w_i(k+1) \sum_{j=0}^k \left\{ \left(\frac{2}{k+1} \sum_{i=0}^j \eta_i + \frac{2(j+1)}{k+1} p_0 + \frac{k-j}{k+1} \right) \right. \right. \right. \\
&\quad \left. \left. \left. - \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 + \frac{k+1-j}{k+1} \right) \right\} b_j(w_i; k) \right] \right. \\
&\quad \left. \times \left[\sum_{j=0}^{k+1} \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 + \frac{k+1-j}{k+1} \right) b_j(w_i; k+1) \right. \right. \\
&\quad \left. \left. + (1-w_i)(k+1) \sum_{j=0}^k \left\{ \left(\frac{2}{k+1} \sum_{i=0}^j \eta_i + \frac{2(j+1)}{k+1} p_0 + \frac{k-j}{k+1} \right) \right. \right. \right. \\
&\quad \left. \left. \left. - \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 + \frac{k+1-j}{k+1} \right) \right\} b_j(w_i; k) \right] \frac{1}{(r_i(1-w_i))^2} \right. \\
&\quad \left. + \frac{w_i^2}{r_i} k(k+1) \sum_{j=0}^{k-1} \left\{ \left(\frac{2}{k+1} \sum_{i=0}^{j+1} \eta_i + \frac{2(j+2)}{k+1} p_0 + \frac{k-j-1}{k+1} \right) \right. \right. \\
&\quad \left. \left. - 2 \left(\frac{2}{k+1} \sum_{i=0}^j \eta_i + \frac{2(j+1)}{k+1} p_0 + \frac{k-j}{k+1} \right) + \left(\frac{2}{k+1} \sum_{i=0}^{j-1} \eta_i + \frac{2j}{k+1} p_0 \right. \right. \right. \\
&\quad \left. \left. \left. + \frac{k+1-j}{k+1} \right) \right\} b_j(w_i; k-1) \right\} \times \frac{e^{-\zeta} \zeta^{k-1}}{(k-1)!}
\end{aligned} \tag{6.22}$$

Notice that, given k , η_0 and η_k are not involved in the computation because they are constrained to be equal to zero and $1 - p_0 - p_1$, respectively.

We stress that an important step of the MCMC algorithm consists in changing the size of the vector $\boldsymbol{\eta}_k$ when a change of k is proposed. For this reason we consider a trans-

dimensional MCMC scheme (Godsill 2001). Following the ideas in Antoniano-Villalobos and Walker (2013) we extend $\Pi(\boldsymbol{\eta}_k, k)$ to

$$\Pi(\boldsymbol{\eta}_\infty, k) = \Pi(\boldsymbol{\eta}_k|k) \Pi(k) \prod_{j>k} \Pi(\eta_j),$$

where $\boldsymbol{\eta}_\infty$ denotes the coefficients of an hypothetical infinite polynomial representation, i.e. the vector $(\eta_0, \dots, \eta_\infty)$. We then construct a proposal distribution for a Metropolis-Hastings algorithm which would update $(k^{(s)}, \boldsymbol{\eta}_\infty^{(s)})$. We consider the following,

$$q(k, \boldsymbol{\eta}_\infty|k^{(s)}, \boldsymbol{\eta}_\infty^{(s)}, \varpi) = q_k(k|k^{(s)}) q_\eta(\boldsymbol{\eta}_k|k, k^{(s)}, \boldsymbol{\eta}_{k^{(s)}}, \varpi) \prod_{j>k} \Pi(\eta_j),$$

where $q_\eta(\boldsymbol{\eta}_k|k, k^{(s)}, \boldsymbol{\eta}_{k^{(s)}}, \varpi)$ refers to the prior $\Pi(\boldsymbol{\eta}_k|k)$. $\Pi(\eta_j)$ is a fully specified density on \mathcal{S} ,

$$q_k(k = k^{(s)} + 1|k^{(s)}) = \begin{cases} 1 & \text{if } k = 3 \\ 1/2 & \text{if } k > 3 \end{cases}$$

and

$$q_k(k^{(s+1)} = k^{(s)} - 1|k^{(s)}) = \begin{cases} 0 & \text{if } k = 3 \\ 1/2 & \text{if } k > 3. \end{cases}$$

Notice that, $k = 3$ is the minimum value that ensures the relation stated in Proposition 6.1.

We proceed by sampling k^* and the corresponding $\boldsymbol{\eta}_{k^*}^*$ from the proposal distribution and then calculate the acceptance probability $p^* = \min\{p(k, k^*, \boldsymbol{\eta}_k, \boldsymbol{\eta}_{k^*}^*), 1\}$. We also stress that the η_j for $j > \max(k, k^*)$ are not needed, so a finite number of parameters is sufficient for each update. This because, in the updating step of the Metropolis-Hastings algorithm, the expression of the ratio simplifies the product terms $\prod_{j>k} \Pi(\eta_j)$ and $\prod_{j>k^*} \Pi(\eta_j^*)$. Specifically we have

$$\begin{aligned} p(k, k^*, \boldsymbol{\eta}_k, \boldsymbol{\eta}_{k^*}^*) &= \frac{\Pi^n(k^*, \boldsymbol{\eta}_\infty^*|\mathbf{Y}, \varpi) q(k, \boldsymbol{\eta}_\infty|k^*, \boldsymbol{\eta}_\infty^*, \varpi)}{\Pi^n(k, \boldsymbol{\eta}_\infty|\mathbf{Y}, \varpi) q(k^*, \boldsymbol{\eta}_\infty^*|k, \boldsymbol{\eta}_\infty, \varpi)} \\ &= \frac{\Pi(k^*) \times \Pi(\boldsymbol{\eta}_{k^*}^*|k^*, \varpi) \prod_{j>k^*} \Pi(\eta_j^*) \times L(\mathbf{y}; k^*, \boldsymbol{\eta}_{k^*}^*, \varpi)}{\Pi(k) \times \Pi(\boldsymbol{\eta}_k|k, \varpi) \prod_{j>k} \Pi(\eta_j) \times L(\mathbf{y}; k, \boldsymbol{\eta}_k, \varpi)} \\ &\times \frac{q_k(k|k^*) \times q_\eta(\boldsymbol{\eta}_k|k, k^*, \boldsymbol{\eta}_{k^*}^*, \varpi) \times \prod_{j>k} \Pi(\eta_j)}{q_k(k^*|k) \times q_\eta(\boldsymbol{\eta}_{k^*}^*|k^*, k, \boldsymbol{\eta}_k, \varpi) \times \prod_{j>k^*} \Pi(\eta_j^*)}. \end{aligned} \tag{6.23}$$

This, for any $k > 3$, can be summarised in the following,

$$p(k, k^*, \boldsymbol{\eta}_k, \boldsymbol{\eta}_{k^*}) = \begin{cases} \frac{\zeta}{k} \frac{L(\mathbf{y}; k+1, \boldsymbol{\eta}_{k+1}, \varpi)}{L(\mathbf{y}; k, \boldsymbol{\eta}_k, \varpi)} & \text{if } k^* = k+1 \\ \frac{(k-1)}{\zeta} \frac{L(\mathbf{y}; k-1, \boldsymbol{\eta}_{k-1}, \varpi)}{L(\mathbf{y}; k, \boldsymbol{\eta}_k, \varpi)} & \text{if } k^* = k-1 \end{cases}$$

Thus, the MCMC scheme only consider a finite number of eta coefficients. The complete algorithm is given below.

Algorithm 6.3. Let $\phi = (k, \boldsymbol{\eta}_k)$ be the vector of all parameters. For $s = 1, \dots, S$,

1. Pick a starting value, $\phi^{(0)} = (k^{(0)}, \boldsymbol{\eta}_{k^{(0)}}) \in \mathcal{E}_{k^{(0)}})$;
2. Update each variable in turn:
 - (a) Sample $(k^{(s)}, \boldsymbol{\eta}_{k^{(s)}}) \sim \Pi^n(k, \boldsymbol{\eta}_k | k^{(s-1)}, \boldsymbol{\eta}_{k^{(s-1)}}, \mathbf{Y}, \varpi)$. This requires a Metropolis-Hastings update:
 - i. ‘‘Propose’’ a new value for k , say k^* , according to the proposal distribution q_k along with the corresponding $\boldsymbol{\eta}_{k^*}$;
 - ii. Compute the Metropolis-Hastings acceptance probability,
$$p^* = \min \left\{ p(k^{(s-1)}, k^*, \boldsymbol{\eta}_{k^{(s)}}, \boldsymbol{\eta}_{k^*}), 1 \right\}$$
with $p(k^{(s-1)}, k^*, \boldsymbol{\eta}_{k^{(s)}}, \boldsymbol{\eta}_{k^*})$ as in (6.23);
 - iii. Accept the new values $(k^*, \boldsymbol{\eta}_{k^*})$ with probability p^* , thus set $k^{(s)} = k^*$ and $\boldsymbol{\eta}_{k^{(s)}} = \boldsymbol{\eta}_{k^*}$. Otherwise reject k^* , i.e., the next value of k remains the same as before, $k^{(s-1)}$, and so does $\boldsymbol{\eta}_{k^{(s)}}$.
 - (b) We now have a new Markov chain state $\phi^{(s)} = (k^{(s)}, \boldsymbol{\eta}_{k^{(s)}})$;

3. Return to step 2 and repeat to produce a Markov chain of length S .

A simulation study is in progress for inclusion in a paper which, as often happens in research, will be ready after the submission of this thesis.

Some extensions are possible and would be interesting to explore. Firstly, we could also infer the parameters of the marginal distributions, which could be done by extending the prior to include ϖ and then incorporate the additional parameters into the MCMC scheme. Similarly, involving the probability masses p_0 and p_1 at the extreme points of the unit simplex. Secondly, the multivariate extension of this methodology sounds fascinating, especially from a computational point of view.

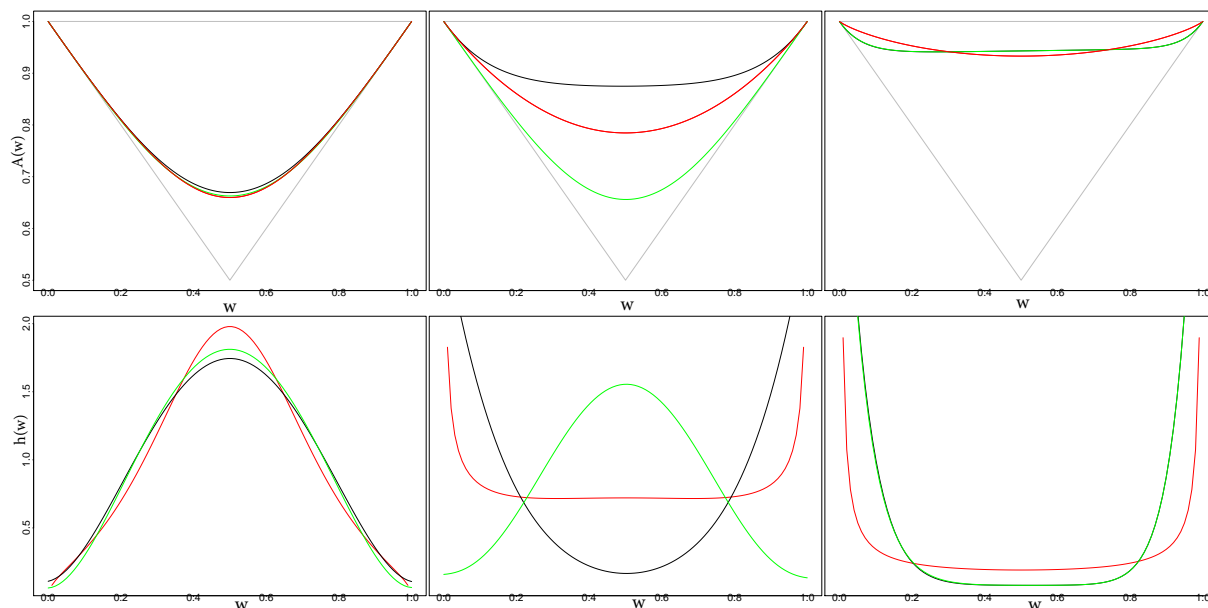


Figure 6.2: Pickands dependence function (top panels) and angular distribution function (bottom panels) of the Symmetric Logistic model with strong, mild and weak dependence ($\alpha = 0.4, 0.7, 0.9$) and polynomial orders $k = 5, 7, 18$ respectively. The red line is the true function; the black line is the mean curve of the MCMC posterior sample; the green line is the mean curve of the MCMC *pseudo*-posterior sample.

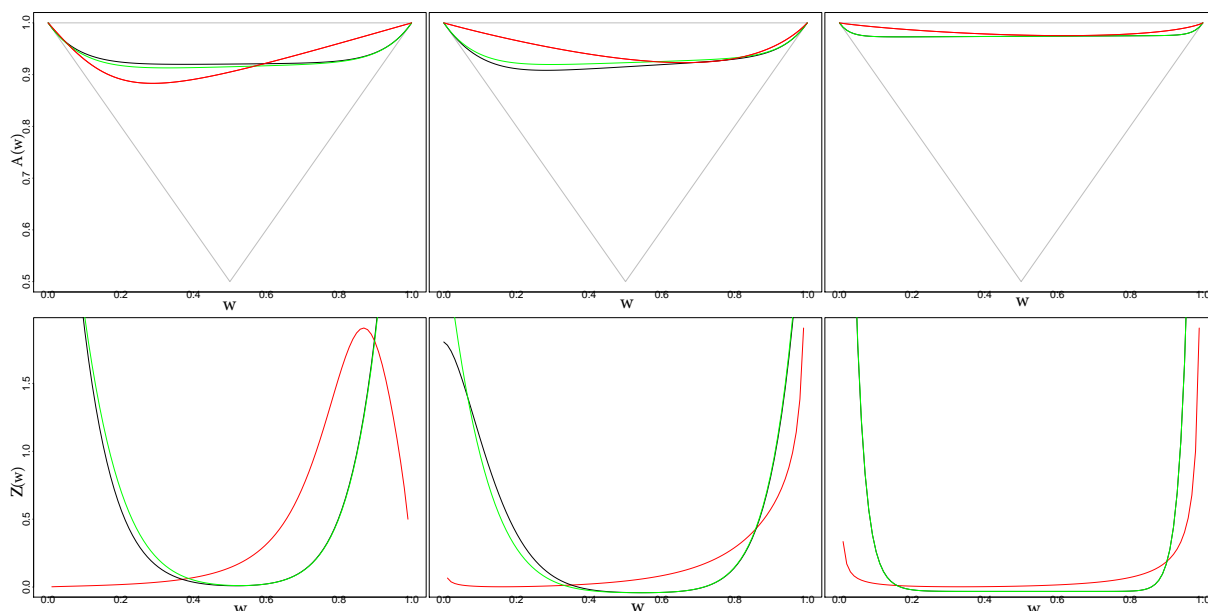


Figure 6.3: Pickands dependence function (top panels) and angular distribution function (bottom panels) of the Asymmetric Logistic model with strong, mild and weak dependence ($\alpha = 0.4, 0.7, 0.9$, $\beta_1 = 0.8$, $\beta_2 = 0.2$) and polynomial orders $k = 12, 14, 40$ respectively. The red line is the true function; the black line is the mean curve of the MCMC posterior sample; the green line is the mean curve of the MCMC *pseudo*-posterior sample.

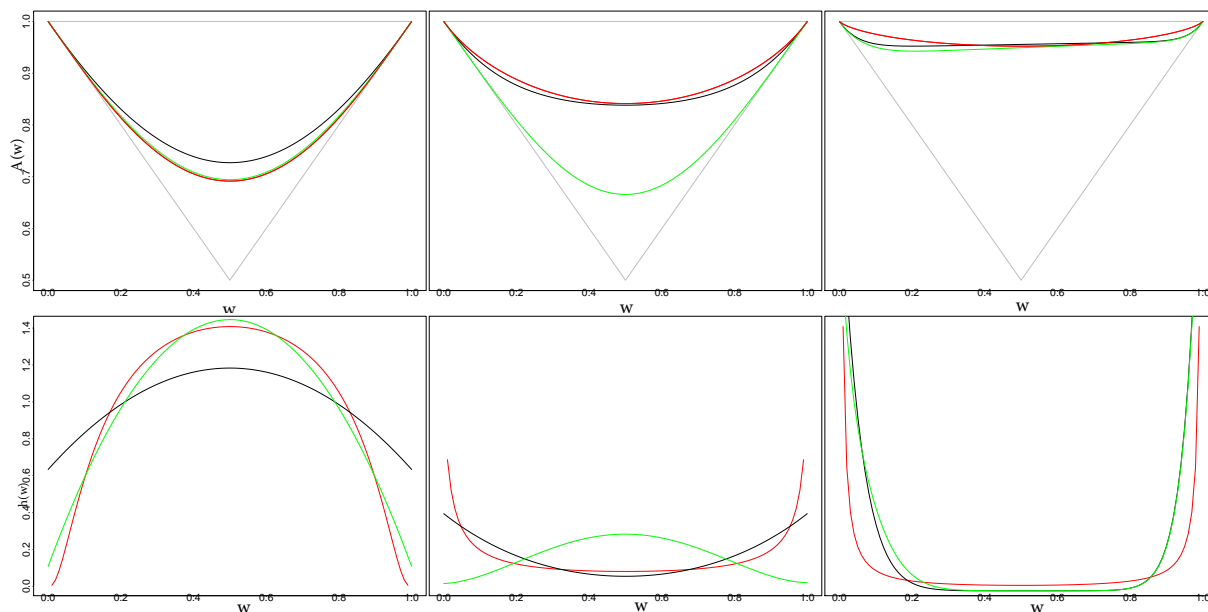


Figure 6.4: Pickands dependence function (top panels) and angular distribution function (bottom panels) of the Hüsler-Reiss model with strong, mild and weak dependence ($\lambda = 2, 1, 0.6$) and polynomial orders $k = 3, 5, 25$ respectively. The red line is the true function; the black line is the mean curve of the MCMC posterior sample; the green line is the mean curve of the MCMC *pseudo*-posterior sample.

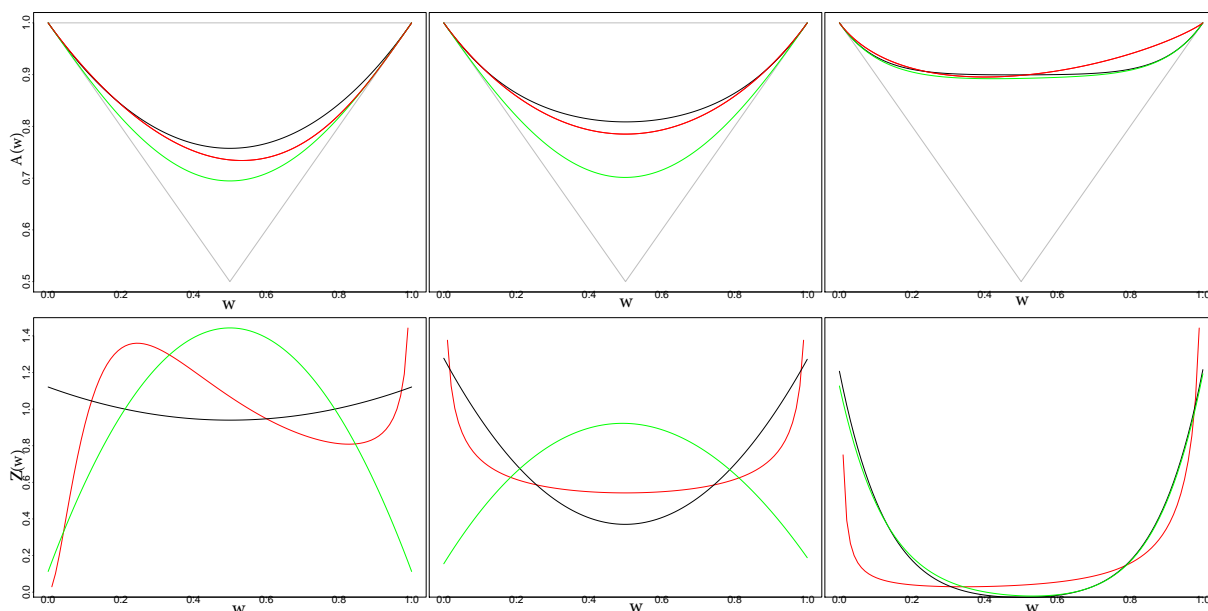


Figure 6.5: Pickands dependence function (top panels) and angular distribution function (bottom panels) of the Tilted Dirichlet model with strong, mild and weak dependence ($\alpha = 3, 0.7, 0.1, \beta = 0.7$) and polynomial orders $k = 3, 4, 9$ respectively. The red line is the true function; the black line is the mean curve of the MCMC posterior sample; the green line is the mean curve of the MCMC *pseudo*-posterior sample.

Tesi di dottorato "Non-Parametric Inference and Simulation in Extreme Value Theory"
di MARCON GIULIA

discussa presso Università Commerciale Luigi Bocconi-Milano nell'anno 2016

La tesi è tutelata dalla normativa sul diritto d'autore (Legge 22 aprile 1941, n.633 e successive integrazioni e modifiche).

Sono comunque fatti salvi i diritti dell'università Commerciale Luigi Bocconi di riproduzione per scopi di ricerca e didattici, con citazione della fonte.

Appendix: Proofs

For $\mathbf{w} \in \mathcal{S}_d$, define the function $\nu_{\mathbf{w}} : [0, 1]^d \rightarrow [0, 1]$ by

$$\nu_{\mathbf{w}}(\mathbf{u}) = \bigvee_{i=1, \dots, d} (u_i^{1/w_i}) - \frac{1}{d} \sum_{i=1}^d u_i^{1/w_i}, \quad \mathbf{u} \in [0, 1]^d, \quad (\text{A.1})$$

where, by convention, $u^{1/w} = 0$ whenever $w = 0$ and $u \in [0, 1]$.

Lemma A.1. *For any cumulative distribution function H on $[0, 1]^d$ and for any $\mathbf{w} \in \mathcal{S}_d$, we have*

$$\int_{[0,1]^d} \nu_{\mathbf{w}}(\mathbf{u}) \, dH(\mathbf{u}) = \frac{1}{d} \sum_{i=1}^d \int_0^1 H(1, \dots, 1, x^{w_i}, 1, \dots, 1) \, dx - \int_0^1 H(x^{w_1}, \dots, x^{w_d}) \, dx.$$

Proof. Fix $\mathbf{w} \in \mathcal{S}_d$. For every $\mathbf{u} \in [0, 1]^d$ we have

$$\begin{aligned} \bigvee_{i=1, \dots, d} u_i^{1/w_i} &= 1 - \int_0^1 \mathbb{I}(\forall i = 1, \dots, d : u_i^{1/w_i} \leq x) \, dx \\ &= 1 - \int_0^1 \mathbb{I}(\forall i = 1, \dots, d : u_i \leq x^{w_i}) \, dx \end{aligned}$$

and

$$\frac{1}{d} \sum_{i=1}^d u_i^{1/w_i} = 1 - \frac{1}{d} \sum_{i=1}^d \int_0^1 \mathbb{I}(u_i \leq x^{w_i}) \, dx.$$

Subtracting both expressions and integrating over H yields

$$\begin{aligned} \int_{[0,1]^d} \nu_{\mathbf{w}}(\mathbf{u}) \, dH(\mathbf{u}) &= \frac{1}{d} \sum_{i=1}^d \int_{[0,1]^d} \int_0^1 \mathbb{I}(u_i \leq x^{w_i}) \, dx \, dH(\mathbf{u}) \\ &\quad - \int_{[0,1]^d} \int_0^1 \mathbb{I}(\forall i = 1, \dots, d : u_i \leq x^{w_i}) \, dx \, dH(\mathbf{u}). \end{aligned}$$

Applying Fubini's theorem to both double integrals yields the stated formula. \square

Proof of Proposition 4.1

The first part of the property 1 stems from the inequality $\bar{U}_d \leq M_d$ and hence taking the expectation. For the second part, it must only be shown that $\rho(\mathbf{U}) = 0$ implies $U_i = U_j$, for $i, j \in I$, since the converse is clear. If an inequality $U_j < U_k$, for $j \in I$ and $k \in I \setminus \{j\}$, would exist, then as a consequence $U_j < M_d$ and $U_k \leq M_d$, $k \neq j$. This implies $\bar{U}_d < M_d$ involving the contradiction $\rho(\mathbf{U}) > 0$. Property 2 is straightforward. Inequalities of property 3 rely on

$$M_d + Y \leq \sum_{i=1}^d \max\{U_i, Y\}. \quad (\text{A.2})$$

Let $m_d = \min_{i \in I} \{U_i\}$. If $Y < m_d$, then $U_i - Y > 0$, for all $i \in I$. Therefore, for all $i \in I$

$$U_i \leq U_i + \sum_{j \neq i} U_j - Y = \sum_{i \in I} \max\{U_j, Y\} - Y,$$

and hence (A.2) is satisfied. If $Y \geq m_d$, then the above inequality is still valid since $\sum_{i \in I} \max\{U_i, Y\} - Y = \sum_{j \in I \setminus \{k\}} \max\{U_j, Y\} \geq U_i$, for all $i \in I$ and $k \in I : m_d = U_k$ and therefore (A.2) is satisfied. In conclusion, inequality (4.3) stems from the inequality

$$M_d - \bar{U}_d \leq \sum_{i \in I} \left(\max\{U_i, Y\} - \frac{U_i + Y}{2} \right) - Y - \bar{U}_d + \sum_{i \in I} \frac{U_i + Y}{2}.$$

Inequality (4.4) is obtained by noting that in virtue of (A.2) it follows that

$$M_d - \bar{U}_d \leq \sum_{i \in I} \max\{U_j, Y\} - \frac{1}{d} \sum_{i \in I} (U_i + Y),$$

thus adding and subtracting $\max\{\mathbf{U} + Y\}$ on the right-hand side and taking the expectation on both sides, the result follows. \square

Proof of Proposition 4.2

By the definition of the $\nu(\mathbf{w})$ in (4.1) it follows that

$$\nu(\mathbf{w}) = \mathbb{E}(M_d) - \mathbb{E}(\bar{U}_d),$$

where $M_d = \max\{F^{1/w_1}(X_1), \dots, F^{1/w_d}(X_d)\}$, $\bar{U}_d = d^{-1} \sum_{i \in I} F^{1/w_i}(X_i)$, $0 < w_i < 1$ with $i \in I$. The expression of $\nu(\mathbf{w})$ follows by noting that

$$\mathbb{E}\{M_d\} = \frac{V(1/w_1, \dots, 1/w_d)}{1 + V(1/w_1, \dots, 1/w_d)}, \quad \mathbb{E}\{F^{1/w_i}(X_i)\} = \frac{w_i}{1 + w_i}, \quad i \in I,$$

where the left-hand side is easily obtained knowing the distribution of M_d that is

$$\mathbb{P}\{M_d \leq u\} = u^{V(1/w_1, \dots, 1/w_d)},$$

whereas the right-hand side is obtained from the facts that $U = F(X_i)$ is uniformly distributed in $[0, 1]$ and $\mathbb{E}\{U^{1/a}\} = a/(1+a)$. Putting all together we obtain

$$\nu(\mathbf{w}) = \frac{V(1/w_1, \dots, 1/w_d)}{1 + V(1/w_1, \dots, 1/w_d)} - \frac{1}{d} \sum_{i \in I} \frac{w_i}{1 + w_i}.$$

Finally, the expression of $V(1/w_1, \dots, 1/w_d)$ is simply obtained inverting the above expression.

The same conclusion arises if considering copulas. The joint distribution $G(\mathbf{x})$ can be written in terms of uniform margins as

$$G(\mathbf{x}) = C\{F_1(x_1), \dots, F_d(x_d)\}$$

where C is the extreme-value copula,

$$C(u_1, \dots, u_d) = \exp(-\ell(-\log u_1, \dots, -\log u_d)), \quad \mathbf{u} \in (0, 1]^d, \quad (\text{A.3})$$

with $\ell : \mathbb{R}_+^d \rightarrow [0, \infty)$ the so-called stable tail dependence function. The latter function is homogeneous of order one and is therefore determined by the Pickands dependence function. Formally, we have

$$\ell(\mathbf{z}) = (z_1 + \dots + z_d) A(\mathbf{w}), \quad \mathbf{z} \in \mathbb{R}_+^d, \quad (\text{A.4})$$

where $w_j = z_j/(z_1 + \dots + z_d)$ for $j \in I$. We view A as a function defined on the d -dimensional unit simplex.

The marginal distribution functions being continuous, the copula C is the joint distribution function of the random vector $(F_1(X_1), \dots, F_d(X_d))$. For $\mathbf{w} \in \mathcal{S}_d$, the multivariate \mathbf{w} -madogram can thus be written as

$$\nu(\mathbf{w}) = \int_{[0,1]^d} \nu_{\mathbf{w}}(\mathbf{u}) dC(\mathbf{u}).$$

Next, apply Lemma A.1. Since C is an extreme-value copula with Pickands dependence function A , we find, after some elementary calculations using (A.3) and (A.4),

$$C(x^{w_1}, \dots, x^{w_d}) = x^{A(\mathbf{w})}$$

for all $x \in (0, 1)$. We obtain

$$\begin{aligned} \nu(\mathbf{w}) &= \frac{1}{d} \sum_{i=1}^d \int_0^1 C(1, \dots, 1, x^{w_i}, 1, \dots, 1) dx - \int_0^1 C(x^{w_1}, \dots, x^{w_d}) dx \\ &= \frac{1}{d} \sum_{i=1}^d \int_0^1 x^{w_i} dx - \int_0^1 x^{A(\mathbf{w})} dx, \end{aligned} \quad (\text{A.5})$$

yielding the formula stated in the proposition. Finally, solve for $A(\mathbf{w})$. \square

Proof of Proposition 4.3

For the first statement, by the Cauchy–Schwarz inequality we have $|B_A(\mathbf{w}; k) - A(\mathbf{w})| \leq \mathbb{E}|A(\mathbf{Y}_k/k) - A(\mathbf{w})|$, where $\mathbf{Y}_k = (Y_{k,i}, i = 1, \dots, d)$ with $Y_{k,i}$ is Binomial with k trials and success probability w_i . Since the Pickands dependence function is Lipschitz-1, then $|B_A(\mathbf{w}; k) - A(\mathbf{w})| \leq \sum_{i=1}^d \mathbb{E}|Y_{k,i}/k - w_i|$. From this follows that $|B_A(\mathbf{w}; k) - A(\mathbf{w})| \leq \sum_{i=1}^d (\mathbb{E}(Y_{k,i}/k - w_i)^2)^{1/2}$ and therefore the result $|B_A(\mathbf{w}; k) - A(\mathbf{w})| \leq d/(2\sqrt{k})$. For the second statement, since $\mathbb{E}(Y_{k,i}) = kw_i, i = 1, \dots, d$, we have

$$B_A(\mathbf{w}; k) - A(\mathbf{w}) = \mathbb{E}\{A(\mathbf{Y}_k/k) - A(\mathbf{w}) - (\mathbf{Y}_k/k - \mathbf{w})^\top \nabla A(\mathbf{w})\}.$$

By the smoothness assumption on A , it follows that

$$|B_A(\mathbf{w}; k) - A(\mathbf{w})| \leq c \sum_{i=1}^d \mathbb{E}\{(Y_{k,i}/k - w_i)^2\} = c \sum_{i=1}^d w_i(1 - w_i)/k \leq cd/(4k),$$

for some positive constant c . □

Proof of Proposition 4.4

On one hand we have that if $B_A(\mathbf{w}; k) \geq \max(w_1, \dots, w_d)$, then $D_{\mathbf{v}_i - \mathbf{v}_j} B_A(\mathbf{v}_j; k) \geq -1$. Indeed, $\max(w_1, \dots, w_d)$ is the intersection of the planes $z_0 = 1 - w_1 - w_2 - \dots - w_{d-1}$, $z_1 = w_1, \dots, z_{d-1} = w_{d-1}$, then by the assumption

$$B_A(\mathbf{v}_j; k) \geq z_j, \quad j = 0, 1, \dots, d-1.$$

The directional derivatives of B_A calculate for $\mathbf{v}_j, j = 0, 1, \dots, d-1$, are equal to

$$D_{\mathbf{v}_i - \mathbf{v}_j} B_A(\mathbf{v}_j; k) = \begin{cases} D_{\mathbf{v}_i - \mathbf{v}_0} B_A(\mathbf{v}_0; k) & \text{if } i \neq 0 = j \\ -D_{\mathbf{v}_j - \mathbf{v}_0} B_A(\mathbf{v}_j; k) & \text{if } i = 0 \neq j \\ D_{\mathbf{v}_i - \mathbf{v}_0} B_A(\mathbf{v}_j; k) - D_{\mathbf{v}_j - \mathbf{v}_0} B_A(\mathbf{v}_j; k) & \text{if } i \neq 0 \neq j, i \neq j \end{cases} \quad (\text{A.6})$$

which are bounded from below by -1 . Then, considering the directional derivatives on both sides of the above inequality we obtain

$$D_{\mathbf{v}_i - \mathbf{v}_j} B_A(\mathbf{v}_j; k) \geq -1, \quad \forall i, j = 0, 1, \dots, d-1, i \neq j,$$

and hence the result.

On the other hand if $D_{\mathbf{v}_i - \mathbf{v}_j} B_A(\mathbf{v}_j; k) \geq -1, j = 0, \dots, d-1$ then $B_A(\mathbf{w}; k) \geq \max(w_1, \dots, w_d)$. Since B_A lies above the tangent plane

$$B_A(\mathbf{w}; k) \geq B_A(\mathbf{w}'; k) + (\mathbf{w}' - \mathbf{w})^\top \nabla B_A(\mathbf{w}'; k), \quad \forall \mathbf{w}, \mathbf{w}' \in \mathcal{S}_d. \quad (\text{A.7})$$

by the convexity assumption, then evaluating this inequality for $\mathbf{w}' = \mathbf{v}_j$ for $j \in \{0, 1, \dots, d-1\}$ we obtain the desired result $B_A(\mathbf{w}; k) \geq w_j$ for all $\mathbf{w} \in \mathcal{S}_d$. Indeed, considering (A.7) at $\mathbf{w}' = \mathbf{v}_0$ we find, for $\mathbf{w} \in \mathcal{S}_d$,

$$B_A(\mathbf{w}; k) \geq 1 + \mathbf{w}^\top \nabla B_A(\mathbf{v}_0; k) = 1 + \sum_{i=1}^{d-1} w_i D_{\mathbf{v}_i - \mathbf{v}_0} B_A(\mathbf{v}_0; k) \geq 1 + \sum_{i=1}^{d-1} w_i (-1) = w_d$$

where $w_d = 1 - w_1 - \dots - w_{d-1}$, as required. Furthermore, considering (A.7) at $\mathbf{w}' = \mathbf{v}_j$ for $j \in \{1, \dots, d-1\}$ we find for $\mathbf{w} \in \mathcal{S}_d$,

$$\begin{aligned} B_A(\mathbf{w}; k) &\geq 1 + (\mathbf{w} - \mathbf{v}_j)^\top \nabla B_A(\mathbf{v}_j; k) \\ &= 1 + (w_j - 1) D_{\mathbf{v}_j - \mathbf{v}_0} B_A(\mathbf{v}_j; k) + \sum_{\substack{i=1 \\ i \neq j}}^{d-1} w_i D_{\mathbf{v}_i - \mathbf{v}_0} B_A(\mathbf{v}_j; k) \\ &\geq 1 + (w_j - 1) D_{\mathbf{v}_j - \mathbf{v}_0} B_A(\mathbf{v}_j; k) + \sum_{\substack{i=1 \\ i \neq j}}^{d-1} w_i \left(D_{\mathbf{v}_j - \mathbf{v}_0} B_A(\mathbf{v}_j; k) - 1 \right) \\ &= 1 + (w_j - 1) D_{\mathbf{v}_j - \mathbf{v}_0} B_A(\mathbf{v}_j; k) + (1 - w_j - w_d) \left(D_{\mathbf{v}_j - \mathbf{v}_0} B_A(\mathbf{v}_j; k) - 1 \right) \\ &= w_j + w_d \left(1 - D_{\mathbf{v}_j - \mathbf{v}_0} B_A(\mathbf{v}_j; k) \right) \geq w_j \end{aligned}$$

given that $D_{\mathbf{v}_i - \mathbf{v}_0} B_A(\mathbf{v}_j; k) \geq D_{\mathbf{v}_j - \mathbf{v}_0} B_A(\mathbf{v}_j; k) - 1$ and $1 - D_{\mathbf{v}_j - \mathbf{v}_0} B_A(\mathbf{v}_j; k) \geq 0$. \square

Proof of Theorem 4.1

The proof proceeds by expressing the statistics and empirical \mathbf{w} -madogram $\nu_n(\mathbf{w})$ and $\hat{\nu}_n(\mathbf{w})$ in terms of the empirical distribution and copula and then exploiting known results on the latter. For $i = 1, \dots, d$ and $j = 1, \dots, n$, let

$$\begin{aligned} \mathbf{U}_j &= (U_{j,1}, \dots, U_{j,d}), \quad U_{j,i} = F_i(X_{j,i}), \\ \hat{\mathbf{U}}_j &= (\hat{U}_{j,1}, \dots, \hat{U}_{j,d}), \quad \hat{U}_{j,i} = F_{n,i}(X_{j,i}) = \frac{1}{n} \sum_{m=1}^n \mathbb{I}(X_{m,i} \leq X_{j,i}). \end{aligned}$$

Recall $\nu_{\mathbf{w}}$ in (A.1). The statistics and empirical \mathbf{w} -madogram are equal to

$$\nu_n(\mathbf{w}) = \frac{1}{n} \sum_{m=1}^n \nu_{\mathbf{w}}(\mathbf{U}_m) = \int_{[0,1]^d} \nu_{\mathbf{w}}(\mathbf{u}) dC_n(\mathbf{u}), \quad \hat{\nu}_n(\mathbf{w}) = \int_{[0,1]^d} \nu_{\mathbf{w}}(\mathbf{u}) d\hat{C}_n(\mathbf{u}),$$

respectively, where C_n and \hat{C}_n are the empirical distribution and copula:

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{m=1}^n \mathbb{I}(\mathbf{U}_m \leq \mathbf{u}), \quad \hat{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{m=1}^n \mathbb{I}(\hat{\mathbf{U}}_m \leq \mathbf{u}), \quad \mathbf{u} \in [0, 1]^d,$$

(component-wise inequalities). By Lemma A.1 we obtain

$$\nu_n(\mathbf{w}) = \frac{1}{d} \sum_{i=1}^d \int_0^1 C_n(1, \dots, 1, x^{w_i}, 1, \dots, 1) dx - \int_0^1 C_n(x^{w_1}, \dots, x^{w_d}) dx, \quad (\text{A.8})$$

and a similar expression is attained for $\widehat{\nu}_n(\mathbf{w})$ but with C_n replaced by \widehat{C}_n . Comparing the latter equation with (A.5) yields

$$\|\nu_n - \nu\|_\infty \leq 2\|C_n - C\|_\infty.$$

Standard empirical process arguments yield uniform strong consistency of the empirical copula (Deheuvels 1991). We come to a similar inequality for $\widehat{\nu}_n$. Uniform strong consistency of A_n and \widehat{A}_n follows.

Next, consider the empirical processes

$$\mathcal{C}_n = \sqrt{n}(C_n - C), \quad \mathbb{C}_n = \sqrt{n}(\widehat{C}_n - C).$$

Combining Equations (A.5) and (A.8) we obtain

$$\sqrt{n}(\nu_n(\mathbf{w}) - \nu(\mathbf{w})) = \frac{1}{d} \sum_{i=1}^d \int_0^1 \mathcal{C}_n(1, \dots, 1, x^{w_i}, 1, \dots, 1) dx - \int_0^1 \mathcal{C}_n(x^{w_1}, \dots, x^{w_d}) dx$$

and clearly a similar expression is obtained for $\sqrt{n}(\widehat{\nu}_n(\mathbf{w}) - \nu(\mathbf{w}))$ but replacing \mathcal{C}_n with \mathbb{C}_n . Now, two related results: in the space $\ell^\infty([0, 1]^d)$ equipped with the supremum norm, $\mathcal{C}_n \rightsquigarrow \mathcal{C}$, as $n \rightarrow \infty$, where \mathcal{C} is a C-Brownian bridge, and if Condition 4.1 holds, then $\mathbb{C}_n \rightsquigarrow \mathbb{C}$, as $n \rightarrow \infty$, where \mathbb{C} is the Gaussian process defined in (4.18). The map

$$\phi : \ell^\infty([0, 1]^d) \rightarrow \ell^\infty(\mathcal{S}_d) : f \mapsto \phi(f)$$

defined by

$$(\phi(f))(\mathbf{w}) = \frac{1}{d} \sum_{i=1}^d \int_0^1 f(1, \dots, 1, x^{w_i}, 1, \dots, 1) dx - \int_0^1 f(x^{w_1}, \dots, x^{w_d}) dx$$

is linear and bounded, and therefore continuous. The continuous mapping theorem then implies

$$\sqrt{n}(\nu_n - \nu) = \phi(\mathcal{C}_n) \rightsquigarrow \phi(\mathcal{C}), \quad \sqrt{n}(\widehat{\nu}_n - \nu) = \phi(\mathbb{C}_n) \rightsquigarrow \phi(\mathbb{C}), \quad n \rightarrow \infty,$$

in $\ell^\infty(\mathcal{S}_d)$. The Gaussian process \mathbb{C} satisfies

$$\mathbb{P}\{\forall i = 1, \dots, d : \forall u \in [0, 1] : \mathbb{C}(1, \dots, 1, u, 1, \dots, 1) = 0\} = 1.$$

This property follows from the continuity of its sample paths and by the form of the covariance function (4.17). We find, for $\mathbf{w} \in \mathcal{S}_d$,

$$(\phi(\mathbb{C}))(\mathbf{w}) = - \int_0^1 \mathbb{C}(x^{w_1}, \dots, x^{w_d}) dx.$$

Finally, apply the functional delta method (van der Vaart 2000, Ch. 20) to arrive at the conclusion. □

Remark 6.1. For $\mathbf{w} \in \mathcal{S}_d$, the Bernstein polynomials' basis of order k can be rewritten as a Bernstein polynomials' basis of order $k + 1$,

$$\begin{aligned} b_{\alpha}(\mathbf{w}; k) &\equiv b_{h_1, \dots, h_{d-1}}(\mathbf{w}; k) = \binom{k}{h_1, \dots, h_d} \prod_{j=1}^d w^{h_j} \\ &= \frac{h_d + 1}{k + 1} b_{h_1, \dots, h_{d-1}}(w; k + 1) + \sum_{j=1}^{d-1} \frac{h_j + 1}{k + 1} b_{h_1, \dots, h_j + 1, \dots, h_{d-1}}(w; k + 1). \end{aligned}$$

Then, the multivariate Bernstein polynomial $B(\mathbf{w}; k)$ can be rewritten as

$$\begin{aligned} B(\mathbf{w}; k) &= \sum_{h_1=0}^k \sum_{h_2=0}^{k-h_1} \cdots \sum_{h_{d-1}=0}^{k-h_1-\cdots-h_{d-2}} \beta_{h_1, \dots, h_{d-1}} b_{h_1, \dots, h_{d-1}}(\mathbf{w}; k) \\ &= \sum_{h_1=0}^k \sum_{h_2=0}^{k-h_1} \cdots \sum_{h_{d-1}=0}^{k-h_1-\cdots-h_{d-2}} \beta_{h_1, \dots, h_{d-1}} \left\{ \frac{h_d + 1}{k + 1} b_{h_1, \dots, h_{d-1}}(\mathbf{w}; k + 1) \right. \\ &\quad \left. + \sum_{j=1}^{d-1} \frac{h_j + 1}{k + 1} b_{h_1, \dots, h_j + 1, \dots, h_{d-1}}(\mathbf{w}; k + 1) \right\} \\ &= \beta_{0, \dots, 0} b_{0, \dots, 0}(\mathbf{w}; k + 1) \\ &\quad + \sum_{h_1=1}^k \left(\beta_{h_1, 0, \dots, 0} \frac{k - h_1 + 1}{k + 1} + \beta_{h_1-1, 0, \dots, 0} \frac{h_1}{k + 1} \right) b_{h_1, 0, \dots, 0}(\mathbf{w}; k + 1) \\ &\quad + \sum_{h_2=1}^k \left(\beta_{0, h_2, 0, \dots, 0} \frac{k - h_2 + 1}{k + 1} + \beta_{0, h_2-1, 0, \dots, 0} \frac{h_2}{k + 1} \right) b_{0, h_2, 0, \dots, 0}(\mathbf{w}; k + 1) + \cdots \\ &\quad + \sum_{h_{d-1}=1}^k \left(\beta_{0, \dots, 0, h_{d-1}} \frac{k - h_{d-1} + 1}{k + 1} + \beta_{0, \dots, 0, h_{d-1}-1} \frac{h_{d-1}}{k + 1} \right) b_{0, \dots, 0, h_{d-1}}(\mathbf{w}; k + 1) \\ &\quad + \sum_{h_1=1}^k \sum_{h_2=1}^{k-h_1+1} \left(\beta_{h_1, h_2, 0, \dots, 0} \frac{k - h_1 - h_2 + 1}{k + 1} + \beta_{h_1-1, h_2, 0, \dots, 0} \frac{h_1}{k + 1} + \right. \\ &\quad \left. + \beta_{h_1, h_2-1, 0, \dots, 0} \frac{h_2}{k + 1} \right) b_{h_1, h_2, 0, \dots, 0}(\mathbf{w}; k + 1) + \cdots \end{aligned}$$

$$\begin{aligned}
& + \sum_{h_{d-2}=1}^k \sum_{h_{d-1}=1}^{k-h_{d-2}+1} \left(\beta_{0,\dots,0,h_{d-2},h_{d-1}} \frac{k-h_{d-2}-h_{d-1}+1}{k+1} + \beta_{0,\dots,0,h_{d-2}-1,h_{d-1}} \frac{h_{d-2}}{k+1} \right. \\
& + \left. \beta_{0,\dots,0,h_{d-2},h_{d-1}-1} \frac{h_{d-1}}{k+1} \right) b_{0,\dots,0,h_{d-2},h_{d-1}}(\mathbf{w}; k+1) \\
& \vdots \\
& + \sum_{h_1=1}^k \sum_{h_2=1}^{k-h_1} \cdots \sum_{h_{d-1}=1}^{k-h_1-\dots-h_{d-2}+1} \left(\beta_{h_1,\dots,h_{d-1}} \frac{h_d+1}{k+1} + \beta_{h_1-1,\dots,h_{d-1}} \frac{h_1+1}{k+1} + \cdots \right. \\
& + \left. \beta_{h_1,\dots,h_{d-1}-1} \frac{h_{d-1}+1}{k+1} \right) b_{h_1,\dots,h_{d-1}}(\mathbf{w}; k) \\
& + \beta_{k,0,\dots,0} b_{k+1,0,\dots,0}(\mathbf{w}; k+1) + \beta_{0,k,0,\dots,0} b_{0,k+1,0,\dots,0}(\mathbf{w}; k+1) + \cdots \\
& + \beta_{0,\dots,0,k} b_{0,\dots,0,k+1}(\mathbf{w}; k+1).
\end{aligned} \tag{A.9}$$

Renaming the polynomial's coefficients as

$$\begin{aligned}
\tilde{\beta}_{0,\dots,0} &= \beta_{0,\dots,0} \\
\tilde{\beta}_{h_1,0,\dots,0} &= \beta_{h_1,0,\dots,0} \frac{k-h_1+1}{k+1} + \beta_{h_1-1,0,\dots,0} \frac{h_1}{k+1} \quad h_1 = 1, \dots, k \\
\tilde{\beta}_{0,h_2,0,\dots,0} &= \beta_{0,h_2,0,\dots,0} \frac{k-h_2+1}{k+1} + \beta_{0,h_2-1,0,\dots,0} \frac{h_2}{k+1} \quad h_2 = 1, \dots, k \\
&\vdots \\
\tilde{\beta}_{0,\dots,0,h_{d-1}} &= \beta_{0,\dots,0,h_{d-1}} \frac{k-h_{d-1}+1}{k+1} + \beta_{0,\dots,0,h_{d-1}-1} \frac{h_{d-1}}{k+1} \quad h_{d-1} = 1, \dots, k \\
\tilde{\beta}_{h_1,h_2,0,\dots,0} &= \beta_{h_1,h_2,0,\dots,0} \frac{k-h_1-h_2+1}{k+1} + \beta_{h_1-1,h_2,0,\dots,0} \frac{h_1}{k+1} + \beta_{h_1,h_2-1,0,\dots,0} \frac{h_2}{k+1} \\
&\quad h_1 = 1, \dots, k, \quad h_2 = 1, \dots, k-h_1+1 \\
&\vdots \\
\tilde{\beta}_{0,\dots,0,h_{d-2},h_{d-1}} &= \beta_{0,\dots,0,h_{d-2},h_{d-1}} \frac{k-h_{d-2}-h_{d-1}+1}{k+1} + \beta_{0,\dots,0,h_{d-2}-1,h_{d-1}} \frac{h_{d-2}}{k+1} + \\
&\quad + \beta_{0,\dots,0,h_{d-2},h_{d-1}-1} \frac{h_{d-1}}{k+1} \quad h_{d-2} = 1, \dots, k, \quad h_{d-1} = 1, \dots, k-h_{d-2}+1 \\
&\vdots \\
\tilde{\beta}_{h_1,\dots,h_{d-1}} &= \beta_{h_1,\dots,h_{d-1}} \frac{h_d+1}{k+1} + \beta_{h_1-1,\dots,h_{d-1}} \frac{h_1+1}{k+1} + \cdots + \beta_{h_1,\dots,h_{d-1}-1} \frac{h_{d-1}+1}{k+1} \\
&\quad h_1 = 1, \dots, k, \quad \dots, \quad h_{d-1} = 1, \dots, k-h_1+\dots+h_{d-2}+1 \\
\tilde{\beta}_{k+1,0,\dots,0} &= \beta_{k,0,\dots,0} \\
\tilde{\beta}_{0,k+1,0,\dots,0} &= \beta_{0,k,0,\dots,0} \\
&\vdots \\
\tilde{\beta}_{0,\dots,0,k} &= \beta_{0,\dots,0,k},
\end{aligned}$$

we obtain

$$B(\mathbf{w}; k) = \sum_{h_1=0}^{k+1} \sum_{h_2=0}^{k+1-h_1} \cdots \sum_{h_{d-1}=0}^{k+1-h_1-\cdots-h_{d-2}} \tilde{\beta}_{h_1, \dots, h_{d-1}} b_{h_1, \dots, h_{d-1}}(\mathbf{w}; k+1) = \tilde{B}(\mathbf{w}; k+1)$$

Therefore, $B(\mathbf{w}; k) \in \mathcal{A}_{k+1}$ and since $B(\cdot; k)$ is a polynomial in \mathcal{A}_k , then we also have that $\mathcal{A}_k \subseteq \mathcal{A}_{k+1}$.

For example, with $d = 2$:

$$\begin{aligned} b_j(w; k) &= \binom{k}{j} w^j (1-w)^{k-j} \\ &= \frac{k-j+1}{k+1} b_j(w; k+1) + \frac{j+1}{k+1} b_{j+1}(w; k+1), \quad w \in [0, 1]. \end{aligned}$$

Then, the Bernstein polynomial becomes

$$\begin{aligned} B(w; k) &= \sum_{j=0}^k \beta_{h-1} b_{h-1}(w; k) \\ &= \sum_{j=0}^k \beta_{h-1} \left\{ \frac{k-j+1}{k+1} b_{h-1}(w; k+1) + \frac{j+1}{k+1} b_{j+1}(w; k+1) \right\} \\ &= \sum_{j=0}^k \beta_{h-1} \frac{k-j+1}{k+1} b_{h-1}(w; k+1) + \sum_{j=0}^k \beta_{h-1} \frac{j+1}{k+1} b_{j+1}(w; k+1) \\ &= \sum_{j=0}^k \beta_{h-1} \frac{k-j+1}{k+1} b_{h-1}(w; k+1) + \sum_{j=1}^k \beta_{j-1} \frac{j}{k+1} b_j(w; k+1) \\ &+ \beta_k b_{k+1}(w; k+1) \\ &= \beta_0 b_0(w; k+1) + \sum_{j=1}^k \left(\beta_j \frac{k-j+1}{k+1} + \beta_{j-1} \frac{j}{k+1} \right) b_j(w; k+1) \\ &+ \beta_k b_{k+1}(w; k+1). \end{aligned}$$

Renaming the coefficients of the basis we obtain

$$B(w; k) = \sum_{j=0}^{k+1} \tilde{\beta}_{h-1} b_j(w; k+1) = \tilde{B}(w; k+1)$$

where

$$\begin{aligned} \tilde{\beta}_0 &= \beta_0 \\ \tilde{\beta}_j &= \beta_j \frac{k-j+1}{k+1} + \beta_{j-1} \frac{j}{k+1} \quad j = 1, \dots, k \\ \tilde{\beta}_{k+1} &= \beta_k. \end{aligned}$$

□

Proof of Proposition 4.5

Firstly, the polynomials in \mathcal{A}_k are nested (e.g., Wang and Ghosh, 2012; Farin, 1986). By the degree-raising property we have

$$B(\mathbf{w}; k) = \sum_{\alpha \in \Gamma_k} \beta_{\alpha} b_{\alpha}(\mathbf{w}; k) = \sum_{\alpha \in \Gamma_{k+1}} \tilde{\beta}_{\alpha} b_{\alpha}(\mathbf{w}; k+1) = \tilde{B}(\mathbf{w}; k+1)$$

where

$$\tilde{\beta}_{\alpha} = \sum_{h=1}^d \frac{\alpha_h}{k+1} \beta_{\alpha - \mathbf{v}_{h-1}}. \quad (\text{A.10})$$

We need to show that the coefficients $\tilde{\beta}_{\alpha}$ satisfy the constraints R1)-R2)-R3). For the case R1) we need to check that

$$\Delta_{i,0}^2 \tilde{\beta}_{\alpha} - \sum_{j \neq i} |\Delta_{i,0} \Delta_{j,0} \tilde{\beta}_{\alpha}| \geq 0, \quad \forall \alpha \in \Gamma_{(k+1)-2}, \quad i = 1, \dots, d-1.$$

This can be rewritten as

$$\Delta_{i,0}^2 \tilde{\beta}_{\alpha} - \sum_{j \neq i} (-1)^{I_{s,t}} \Delta_{i,0} \Delta_{j,0} \tilde{\beta}_{\alpha} \geq 0,$$

where $I_{s,t}$ is the set of all the possible combinations with repetition of the set $\{1, 2\}$ in sequences of $d-2$ terms, $s = 1, \dots, d-2$ and $t = 1, \dots, 2^{d-2}$. Using the relation in (A.10) we have

$$\tilde{\beta}_{\alpha} = \sum_{h=1}^d \frac{\alpha_h}{k+1} \beta_{\alpha - \mathbf{v}_{h-1}} = \sum_{h=1}^{d-1} \frac{\alpha_h}{k+1} \beta_{\alpha - \mathbf{e}_h} + \frac{\alpha_d}{k+1} \beta_{\alpha}.$$

Then, we obtain

$$\begin{aligned} \Delta_{i,0}^2 \tilde{\beta}_{\alpha} - \sum_{j \neq i} (-1)^{I_{s,t}} \Delta_{i,0} \Delta_{j,0} \tilde{\beta}_{\alpha} &= \Delta_{i,0}^2 \left\{ \sum_{h=1}^{d-1} \frac{\alpha_h}{k+1} \beta_{\alpha - \mathbf{e}_h} + \frac{\alpha_d}{k+1} \beta_{\alpha} \right\} \\ &- \sum_{j \neq i} (-1)^{I_{s,t}} \Delta_{i,0} \Delta_{j,0} \left\{ \sum_{h=1}^{d-1} \frac{\alpha_h}{k+1} \beta_{\alpha - \mathbf{e}_h} + \frac{\alpha_d}{k+1} \beta_{\alpha} \right\} \\ &= \sum_{h=1}^{d-1} \frac{\alpha_h}{k+1} \left\{ \Delta_{i,0}^2 \beta_{\alpha - \mathbf{e}_h} - \sum_{j \neq i} (-1)^{I_{s,t}} \Delta_{i,0} \Delta_{j,0} \beta_{\alpha - \mathbf{e}_h} \right\} \\ &+ \sum_{h=1}^{d-1} \frac{\alpha_d - 1}{k+1} \left\{ \Delta_{i,0}^2 \beta_{\alpha} - \sum_{j \neq i} (-1)^{I_{s,t}} \Delta_{i,0} \Delta_{j,0} \beta_{\alpha} \right\} \geq 0, \end{aligned}$$

and hence the result. For the case R2), using (A.10), it is immediate to verify for the set $\{\tilde{\beta}_{\alpha}, \alpha \in \Gamma_{k+1} : \alpha = \mathbf{0} \text{ or } \alpha = (k+1) \mathbf{e}_i, \forall i = 1, \dots, d-1\}$ that $\tilde{\beta}_{\alpha} = \beta_{\alpha} = 1$. Finally,

for the case R3) we need to check that $1 - 1/(k + 1) < \tilde{\beta}_\alpha$, where $\{\tilde{\beta}_\alpha, \alpha \in \Gamma_{k+1} : \alpha = \mathbf{e}_i \text{ or } \alpha = k\mathbf{e}_i \text{ or } \alpha_l = k\mathbf{e}_i + \mathbf{e}_j, \forall j \neq i = 1, \dots, d-1\}$. By definition we have

$$\tilde{\beta}_{\mathbf{e}_i} = \frac{k}{k+1}\beta_{\mathbf{e}_i} + \frac{1}{k+1}, \quad \tilde{\beta}_{k\mathbf{e}_i} = \frac{k}{k+1}\beta_{(k-1)\mathbf{e}_i} + \frac{1}{k+1}, \quad \tilde{\beta}_{k\mathbf{e}_i + \mathbf{e}_j} = \frac{k}{k+1}\beta_{(k-1)\mathbf{e}_i + \mathbf{e}_j} + \frac{1}{k+1}.$$

Substituting $\tilde{\beta}_\alpha$, with $\alpha = \mathbf{e}_i, \alpha = k\mathbf{e}_i, \alpha = k\mathbf{e}_i + \mathbf{e}_j$, in the previous inequality we obtain

$$\begin{aligned} \tilde{\beta}_\alpha &\geq 1 - \frac{1}{k+1} \\ \frac{1}{k+1} \{1 + k\beta_{\alpha - \mathbf{v}_{i-1}}\} &\geq \frac{1}{k+1} \left\{1 + k \left(1 - \frac{1}{k}\right)\right\} \\ \beta_{\alpha - \mathbf{v}_{i-1}} &\geq 1 - \frac{1}{k} \end{aligned}$$

for $i = 1, \dots, d-1$, and hence the result. Thus the first statement is proven.

Secondly, let A be a Pickands dependence function and consider the Bernstein polynomial

$$A_k(\mathbf{w}) = \sum_{\alpha \in \Gamma_k} A(\alpha/k) b_\alpha(\mathbf{w}; k),$$

that is, $A_k = B_A(\cdot; k)$ as in (4.9). Constraint R1) holds by assumption (4.19). Since $\max(w_1, \dots, w_d) \leq A(\mathbf{w}) \leq 1$ for all $\mathbf{w} \in \mathcal{S}_d$, the constraints in R2) and R3) are satisfied too. Finally, we have uniform convergence $A_k \rightarrow A$ by Proposition 4.3. \square

Proof of Proposition 6.1

The angular probability density can be obtained either by (6.9) and (6.14). As a consequence we have that

$$k \sum_{j=0}^{k-1} (\eta_{j+1} - \eta_j) b_j(w; k-1) = \frac{r(r-1)}{2} \sum_{i=0}^{r-2} (\beta_{i+2} - 2\beta_{i+1} + \beta_i) b_i(w; r-2), \quad w \in [0, 1].$$

Setting $r = k + 1$, then the right-hand side of the above equality can rephrase the as a sum for $j = 0, \dots, k-1$ and as a consequence each of the following identity

$$\eta_{j+1} - \eta_j = \frac{k+1}{2} (\beta_{j+2} - 2\beta_{j+1} + \beta_j), \quad j = 0, \dots, k-1. \quad (\text{A.11})$$

hold.

We start with claim i). In order to derive the beta coefficients we start considering (6.8) and assuming that (R4)-(R6) are verified. Applying (A.11) recursively and resolving

with respect to β_{j+2} we obtain,

$$\begin{aligned}
\beta_2 &= \frac{2}{k+1} \left(\eta_1 + (k+1)\beta_1 - \frac{k+1}{2}\beta_0 \right) \\
\beta_3 &= \frac{2}{k+1} \left(\eta_1 + \eta_2 + \frac{3(k+1)}{2}\beta_1 - (k+1)\beta_0 \right) \\
\beta_4 &= \frac{2}{k+1} \left(\eta_1 + \eta_2 + \eta_3 + 2(k+1)\beta_1 - \frac{3(k+1)}{2}\beta_0 \right) \\
\beta_5 &= \frac{2}{k+1} \left(\eta_1 + \eta_2 + \eta_3 + \eta_4 + \frac{5(k+1)}{2}\beta_1 - 2(k+1)\beta_0 \right) \\
&\vdots \\
\beta_{j+2} &= \frac{2}{k+1} \left(\sum_{i=0}^{j+1} \eta_i + \frac{(j+2)(k+1)}{2}\beta_1 - \frac{(j+1)(k+1)}{2}\beta_0 \right). \tag{A.12}
\end{aligned}$$

where we used the fact that $\eta_0 = 0$. So that the Pickands function is well defined it is required that $\beta_0 = \beta_{k+1} = 1$. In addition, we know that

$$p_0 = \frac{1 + A'(0)}{2} = \frac{1 + (k+1)(\beta_1 - \beta_0)}{2} \tag{A.13}$$

and

$$p_1 = \frac{1 - A'(1)}{2} = \frac{1 - (k+1)(\beta_{k+1} - \beta_k)}{2} \tag{A.14}$$

from which we obtain

$$\beta_1 = \frac{2p_0 + k}{k+1}, \quad \beta_k = \frac{2p_1 + k}{k+1}. \tag{A.15}$$

Substituting β_0 and β_1 in (A.12) and setting $j = t - 2$ the formula (6.18) is obtained.

In order to derive the eta coefficients we start considering (6.13) and assuming that (R1)-(R3) are verified. Applying (A.11) and resolving with respect to η_{j+1} we obtain,

$$\begin{aligned}
\eta_1 &= \eta_0 + \frac{k+1}{2}(\beta_2 - 2\beta_1 + \beta_0) \\
\eta_2 &= \eta_0 + \frac{k+1}{2}(\beta_3 - \beta_2 - \beta_1 + \beta_0) \\
\eta_3 &= \eta_0 + \frac{k+1}{2}(\beta_4 - \beta_3 - \beta_1 + \beta_0) \\
\eta_4 &= \eta_0 + \frac{k+1}{2}(\beta_5 - \beta_4 - \beta_1 + \beta_0) \\
&\vdots \\
\eta_{j+1} &= \eta_0 + \frac{k+1}{2}(\beta_{j+2} - \beta_{j+1} - \beta_1 + \beta_0). \tag{A.16}
\end{aligned}$$

To make sure that (A.16) has an unique solution it is required that $\eta_0 = 0$. In addition, substituting β_0 and β_1 in (A.16) and setting $t = j + 1$ the formula (6.19) is obtained.

Now, we move to claim ii). Consider (6.18). When $t = 0$, under the convention that $\eta_{-1} = 0$ we obtain $\beta_0 = 1$. When $t = k + 1$ we obtain $\beta_{k+1} = 2(\eta_1 + \dots + \eta_k + (k +$

$1)p_0)/(k+1)$ and then using (R6) this becomes $\beta_{k+1} = 1$. Thus (R1) is verified. When $t = 1$ we obtain the following result

$$\beta_1 = \frac{2p_0 + k}{k+1} \geq 1 - \frac{1}{k+1},$$

and when $t = k$, after few algebraic manipulation we obtain

$$\beta_k = \frac{2p_1 + k}{k+1} \geq 1 - \frac{1}{k+1}.$$

Thus (R2) is verified. Finally, from (6.18) with few algebraic manipulation it can be checked that

$$\beta_{t+2} - 2\beta_{t+1} + \beta_t = \frac{2}{k+1}(\eta_{t+1} - \eta_t), \quad t = 0, \dots, k-1.$$

Since $\eta_{j+1} \geq \eta_j$ for all $j = 0, \dots, k-1$ by (R1), then also (R6) is verified.

Consider (6.19). Now, from (6.18) it can be checked that

$$\eta_{t+1} - \eta_t = \frac{k+1}{2}(\beta_{t+2} - 2\beta_{t+1} + \beta_t), \quad t = 0, \dots, k-1.$$

The right-hand side is non-negative by (R3) and so (R4) is verified. When $t = 0$, using $\beta_0 = 1$ and β_1 in (A.15) we obtain

$$\begin{aligned} \eta_0 &= \frac{k+1}{2} \left(\beta_1 - \beta_0 + \frac{1-2p_0}{k+1} \right) \\ &= \frac{k+1}{2} \left(\frac{2p_0+k}{k+1} - 1 + \frac{1-2p_0}{k+1} \right) = 0. \end{aligned}$$

When $t = k$, using $\beta_{k+1} = 1$ and β_k in (A.15) we obtain

$$\begin{aligned} \eta_k &= \frac{k+1}{2} \left(\beta_{k+1} - \beta_k + \frac{1-2p_0}{k+1} \right) \\ &= \frac{k+1}{2} \left(1 - \frac{2p_1+k}{k+1} + \frac{1-2p_0}{k+1} \right) = 1 - p_0 - p_1. \end{aligned}$$

Therefore, (R5) is verified. Finally, applying (6.19) repeatedly and with few algebraic manipulations we obtain

$$\sum_{t=0}^k \eta_t = \frac{k+1}{2} \left(\beta_{k+1} - \beta_0 + \frac{(k+1)(1-2p_0)}{k+1} \right) = \frac{(k+1)}{2} (1-2p_0),$$

where we used the fact that $\beta_{k+1} = \beta_0 = 1$. Hence also (R6) is verified. \square

Tesi di dottorato "Non-Parametric Inference and Simulation in Extreme Value Theory"
di MARCON GIULIA

discussa presso Università Commerciale Luigi Bocconi-Milano nell'anno 2016

La tesi è tutelata dalla normativa sul diritto d'autore (Legge 22 aprile 1941, n.633 e successive integrazioni e modifiche).

Sono comunque fatti salvi i diritti dell'università Commerciale Luigi Bocconi di riproduzione per scopi di ricerca e didattici, con citazione della fonte.

Bibliography

- Antoniano-Villalobos, I. and S. G. Walker (2013). Bayesian nonparametric inference for the power likelihood. *Journal of Computational and Graphical Statistics* 22(4), 801–813.
- Apputhurai, P. and A. Stephenson (2011). Accounting for uncertainty in extremal dependence modeling using bayesian model averaging techniques. *Journal of Statistical Planning and Inference* 141(5), 1800–1807.
- Bassett Jr, G. W. (1997). Nonparametric bounds for the probability of future prices based on option values. *Lecture Notes-Monograph Series*, 287–300.
- Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons Ltd., Chichester.
- Beranger, B., G. Marcon, and S. A. Padoan (2015). Extremaldep: Extremal dependence modeling, r package version 0.1-2/r76.
- Beranger, B. and S. A. Padoan (2015). Extreme dependence models. In D. Dey and J. Yan (Eds.), *Extreme Value Modeling and Risk Analysis: Methods and Applications*. Chapman and Hall/CRC.
- Berghaus, B., A. Bücher, and H. Dette (2013). Minimum distance estimators of the pickands dependence function and related tests of multivariate extreme-value dependence. *Journal de la Société Française de Statistique* 154(1), 116–137.
- Bernard, E., P. Naveau, M. Vrac, and O. Mestre (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate*.
- Boldi, M. O. and A. C. Davison (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society, Series B* 69(2), 217–229.
- Bücher, A., H. Dette, and S. Volgushev (2011). New estimators of the pickands dependence function and a test for extreme-value dependence. *The Annals of Statistics* 39(4), 1963–2006.
- Capéraà, P., A.-L. Fougères, and C. Genest (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* 84, 567–577.

- Carnicer, J. and J. Peña (1993). Shape preserving representations and optimality of the Bernstein basis. *Advances in Computational Mathematics* 1(2), 173–196.
- Chang, I. S., C. A. Hsiung, Y. J. Wu, and C. C. Yang (2005). Bayesian survival analysis using bernstein polynomials. *Scandinavian Journal of Statistics* 32, 447–466.
- Coles, S. G. (2001). *An Introduction to Statistical Modelling of Extreme Values*. Springer, London.
- Coles, S. G. and J. A. Tawn (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B (Methodological)* 53(2), pp. 377–392.
- Coles, S. G. and J. A. Tawn (1994). Statistical methods for multivariate extremes: An application to structural design. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 43(1), pp. 1–48.
- Coles, S. G. and J. A. Tawn (1996). Modelling extremes of the areal rainfall process. *J. Roy. Statist. Soc. B* 58, 329–347.
- Cooley, D., R. A. Davis, and P. Naveau (2010). The pairwise beta distribution: a flexible parametric multivariate model for extremes. *J. Multivariate Anal.* 101(9), 2103–2117.
- Davison, A. C. and M. M. Gholamrezaee (2012). Geostatistics of extremes. *Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences* 468, 581–608.
- de Haan, L. and A. Ferreira (2006). *Extreme Value Theory An Introduction*. Springer.
- de Haan, L. and S. I. Resnick (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 40, 317–337.
- Deheuvels, P. (1991). On the limiting behavior of the pickands estimator for bivariate extreme-value distributions. *Statistics and Probability Letters* 12, 429–439.
- Ditzian, Z. (1986). Inverse theorems for multidimensional Bernstein operators. *Pacific Journal of Mathematics* 121(2), 293–319.
- Engelke, S., A. Malinowski, Z. Kabluchko, and M. Schlather (2015, July). Estimation of huesler-reiss distributions and brown-resnick processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77(1), 239–265.
- Falk, M., J. Hüsler, and R. D. Reiss (2010). *Laws of Small Numbers: Extremes and Rare Events* (Third ed.). Birkhäuser/Springer Basel AG, Basel.
- Farin, G. (1986). Triangular bernstein-bézier patches. *Computer Aided Geometric Design* 3(2), 83–127.
- Fils-Villetard, A., A. Guillou, and J. Segers (2008). Projection estimators of pickands dependence functions. *The Canadian Journal of Statistics* 36(3), 369–382.

- Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* 24, 180–190.
- Fonseca, C., L. Ferreira, H. Ferreira, and A. Martins (2013). Generalized madogram and pairwise dependence of maxima over two regions of a random field. *Submitted*.
- Genest, C. and J. Segers (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics* 37(5B), 2990–3022.
- Ghoudi, K., A. Khoudraji, and L.-P. Rivest (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 187–197.
- Godsill, S. J. (2001). On the relationship between markov chain monte carlo methods for model uncertainty. *Journal of Computational and Graphical Statistics* 10(2), 230–248.
- Goldfarb, D. and A. Idnani (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical programming* 27(1), 1–33.
- Gudendorf, G. and J. Segers (2011). Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *Journal of Multivariate Analysis* 102, 37–47.
- Gudendorf, G. and J. Segers (2012). Nonparametric estimation of multivariate extreme-value copula. *Journal of Statistical Planning and Inference* 142, 3073–3085.
- Gumbel, E. J. (1958). *Statistics of extremes*. Columbia University Press, New York.
- Gumbel, E. J. (1960). Distributions des valeurs extremes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris* 9(171–173).
- Hall, P. and N. Tajvidi (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli* 6(5), 835–844.
- Hüsler, J. and R.-D. Reiss (1989). Maxima of normal random vectors: between independence and complete dependence. *Statist. Probab. Lett.* 7(4), 283–286.
- Kotz, S. and S. Nadarajah (2000). *Extreme value distributions theory and applications*. Imperial College Press.
- Lai, M.-J. (1993). Some sufficient conditions for convexity of multivariate Bernstein-Bézier polynomials and box spline surfaces. *Studia Scientiarum Mathematicarum Hungarica* 28(3), 363–374.
- Li, F. J. (2011). Interpolation and convergence of Bernstein-Bézier coefficients. *Acta Mathematica Sinica, English Series* 27(9), 1769–1782.
- Lin, L. and D. B. Dunson (2014). Bayesian monotone regression using gaussian process projection. *Biometrika*, ast063.

- Lorentz, G. G. (1986). *Bernstein Polynomials* (Second ed.). Chelsea Publishing Company, New York.
- Marcon, G., S. A. Padoan, P. Naveau, and P. Muliere (2015). Nonparametric estimation of the pickands dependence function using bernstein polynomials. *Journal of Statistical Planning and Inference*, *Under revision*.
- Naveau, P., A. Guillou, D. Cooley, and J. Diebolt (2009). Modelling pairwise dependence of maxima in space. *Biometrika* *96*(1), 1–17.
- Nikoloulopoulos, A. K., H. Joe, and H. Li (2009). Extreme value properties of multivariate t copulas. *Extremes* *12*(2), 129–148.
- Padoan, S. A. (2011). Multivariate extreme models based on underlying skew- t and skew-normal distributions. *Journal of Multivariate Analysis* *102*(5), 977–991.
- Padoan, S. A. (2013). *Encyclopedia of Environmetrics*, Chapter Extreme Value Analysis. Number 2. John Wiley & Sons, Ltd., Chichester.
- Padoan, S. A., M. Ribatet, and S. A. Sisson (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association, Theory & Methods* *105*, 263–277.
- Petrone, S. (1999). Random Bernstein polynomials. *Scandinavian Journal of Statistics* *26*, 373–393.
- Petrone, S. (2004). On the role of mixtures in bayesian nonparametrics. *Atti della XLII Riunione Scientifica della Societa Italiana di Statistica* *1*, 257–268.
- Pickands, III, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, Volume 49, pp. 859–878, 894–902. With a discussion.
- Resnick, S. I. (2007). *Extreme values, regular variation, and point processes*. Springer.
- Ressel, P. (2013). Homogeneous distributions—and a spectral representation of classical mean values and stable tail dependence functions. *Journal of Multivariate Analysis* *117*, 246–256.
- Sabourin, A. and P. Naveau (2014). Bayesian dirichlet mixture model for multivariate extremes: A re-parametrization. *Computational Statistics and Data Analysis* *71*, 542–567.
- Sabourin, A., P. Naveau, and A.-L. Fougères (2013). Bayesian model averaging for multivariate extremes. *Extremes* *16*(3), 325–350.
- Sauer, T. (1991). Multivariate Bernstein polynomials and convexity. *Computer Aided Geometric Design* *8*(6), 465–478.

- Segers, J. (2012). Asymptotic of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* 18(3), 764–782.
- Smith, R. L., J. A. Tawn, and H. K. Yuen (1990). Statistics of multivariate extremes. *Int. Statist. Rev.* 58, 47–58.
- Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika* 77(2), 245–253.
- van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge University Press.
- Wang, J. and S. Ghosh (2012). Shape restricted nonparametric regression with bernstein polynomials. *Computational Statistics and Data Analysis* 56(9), 2729–2741.
- Ye, Y. and E. Tse (1989). An extension of karmarkar’s projective algorithm for convex quadratic programming. *Mathematical Programming* 44(1-3), 157–179.
- Zhang, D., M. T. Wells, and L. Peng (2008). Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *Journal of Multivariate Analysis* 99(4), 577–588.