



# Kent Academic Repository

**Ortiz Barranos, Antonio A. (2012) *Bayesian mixture models in extreme value theory with an application to investent portfolio analysis*. Doctor of Philosophy (PhD) thesis, University of Kent.**

## Downloaded from

<https://kar.kent.ac.uk/86506/> The University of Kent's Academic Repository KAR

## The version of record is available from

<https://doi.org/10.22024/UniKent/01.02.86506>

## This document version

UNSPECIFIED

## DOI for this version

## Licence for this version

CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

## Additional information

This thesis has been digitised by EThOS, the British Library digitisation service, for purposes of preservation and dissemination. It was uploaded to KAR on 09 February 2021 in order to hold its content and record within University of Kent systems. It is available Open Access using a Creative Commons Attribution, Non-commercial, No Derivatives (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) licence so that the thesis and its author, can benefit from opportunities for increased readership and citation. This was done in line with University of Kent policies (<https://www.kent.ac.uk/is/strategy/docs/Kent%20Open%20Access%20policy.pdf>). If y...

## Versions of research works

### Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

### Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal*, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

## Enquiries

If you have questions about this document contact [ResearchSupport@kent.ac.uk](mailto:ResearchSupport@kent.ac.uk). Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

BAYESIAN MIXTURE MODELS IN  
EXTREME VALUE THEORY WITH AN  
APPLICATION TO INVESTMENT  
PORTFOLIO ANALYSIS

A THESIS SUBMITTED TO  
THE UNIVERSITY OF KENT AT CANTERBURY  
IN THE SUBJECT OF STATISTICS  
FOR THE DEGREE  
OF DOCTOR OF PHILOSOPHY

by

© Antonio A. Ortiz Barranon

December 13, 2012

A la memoria de mi abuelo Alberto “Robespierre”

Barrañón.

# Acknowledgements

El primer párrafo de los agradecimientos no puede referirse a otra cosa que a la extrema gratitud que tengo hacia mis padres, que sin su apoyo, hubiera sido imposible haber realizado esta aventura llamada doctorado. ¡Gracias infinitas!

I would like to thank Prof. Stephen Walker for all the support he gave me during these four years. It was a pleasure working with him.

On the other hand, I would like to thank Manuel Mendoza, who guided me to the path to the PhD. He basically made me realise this was a good idea, and there was no point where I thought he was wrong, since I do not regret any day I spent here.

At least, but no last, I would like to thank all my family and my friends who supported me either from the other side of the ocean (Atlantic, for those who might think it is the Pacific) or from the charming (dark and windy, though) town called Canterbury. Special thanks go to Juan Carlos, Isadora, and Vasso, who helped me directly with some points of the project with no objections.

# Contents

<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>x</b>
<b>1 Introduction</b>	<b>2</b>
<b>2 Extreme Value Theory</b>	<b>9</b>
2.1 Univariate Framework . . . . .	10
2.1.1 Maximum Approach and Extreme Distributions . . . . .	11
2.1.1.1 Maximum Domain of Attraction . . . . .	16
2.1.2 The General Extreme Value Distribution . . . . .	19
2.1.3 Peaks Over the Threshold Approach . . . . .	23
2.1.3.1 Threshold Selection and Inference . . . . .	27
2.2 Multivariate Framework . . . . .	31
2.2.1 Multivariate Maximum Asymptotics . . . . .	32
2.2.2 Characterization of the Multivariate EVD . . . . .	35
2.2.2.1 A Necessary Condition for MDA . . . . .	37

2.2.3	Bivariate Framework . . . . .	39
2.2.3.1	Bivariate Maximum Model . . . . .	40
2.2.3.2	Bivariate POT Approach . . . . .	43
2.2.3.3	Bivariate Extreme parameters . . . . .	46
2.2.4	Summary . . . . .	56
<b>3</b>	<b>Bayesian Theory</b>	<b>58</b>
3.1	The Bayesian Idea . . . . .	59
3.1.1	Prior Distribution . . . . .	60
3.1.2	Posterior Distribution . . . . .	61
3.2	Markov Chain Monte Carlo Methods . . . . .	63
3.2.1	Rejection Sampling . . . . .	63
3.2.2	Gibbs Sampler . . . . .	65
3.2.3	MCMC Metropolis-Hastings . . . . .	66
3.2.4	Metropolis-Hastings within Gibbs sampler . . . . .	67
3.3	Reversible Jump MCMC . . . . .	68
3.4	Bayesian Methods in EVT . . . . .	70
3.5	Conclusions . . . . .	73
<b>4</b>	<b>Interim Bivariate Models</b>	<b>74</b>
4.1	Convolution Models . . . . .	77
4.1.1	Convolution Model 1. Unit Exponential Models . . . . .	78
4.1.2	Convolution Model 2. General Exponential . . . . .	80

4.1.3	Convolution Model 3. GPD . . . . .	82
4.2	Product Models . . . . .	87
4.2.1	General GPD Product Model . . . . .	87
4.2.2	Exponential-Gamma Model 1 . . . . .	89
4.2.2.1	Marginal Behavior . . . . .	90
4.2.2.2	Dependence Structure and Joint Distribution . . . . .	92
4.2.2.3	Inference Exclusive to $\theta$ . Model 1 . . . . .	93
4.2.2.4	Summary of Model 1 . . . . .	97
<b>5</b>	<b>The Bivariate Model</b>	<b>99</b>
5.1	Exponential-Gamma Model 2 . . . . .	100
5.2	Marginal Distributions: Model 2 . . . . .	101
5.3	Joint Distribution and Dependence . . . . .	104
5.4	Model 2 and the Literature . . . . .	106
5.5	Data Characterization for Model 2 . . . . .	111
5.6	Inference for Model 2 . . . . .	114
5.6.1	MLE Approach . . . . .	114
5.6.2	Bayesian Estimation . . . . .	117
5.6.3	General Simulation Case . . . . .	121
<b>6</b>	<b>The Multivariate Model</b>	<b>125</b>
6.1	Multivariate Dependence . . . . .	126
6.2	Trivariate Model . . . . .	130

6.2.1	Trivariate Model Motivation . . . . .	130
6.2.2	Trivariate structure . . . . .	138
6.2.3	Trivariate Inference . . . . .	145
6.2.3.1	General Details of Reversible Jump . . . . .	145
6.2.3.2	Data Characterization for Trivariate Model . . . . .	149
6.2.3.3	Reversible Jump MCMC . . . . .	156
6.2.4	Trivariate Simulation . . . . .	157
6.2.4.1	Simulations for Model Selection . . . . .	158
6.2.4.2	Simulations for Model Averaging . . . . .	160
6.3	Multivariate Model . . . . .	165
6.4	Our Multivariate Model and Literature . . . . .	171
6.5	Discussion of Multivariate Framework . . . . .	174
<b>7</b>	<b>Illustrations</b>	<b>176</b>
7.1	Data Description . . . . .	176
7.2	Tail Dependence Determination Analysis . . . . .	178
7.2.1	Bivariate Dependence Determination . . . . .	178
7.2.2	Trivariate Dependence Determination . . . . .	180
7.3	Portfolio Loss Analysis . . . . .	182
7.4	Conclusions . . . . .	190
<b>8</b>	<b>Conclusions</b>	<b>193</b>
8.1	Final Discussion . . . . .	193

8.2 Future Work . . . . .	197
<b>A Regular Variation</b>	<b>198</b>
<b>B Proofs</b>	<b>200</b>
B.1 Proof of Theorem 4.1.1 . . . . .	200
B.2 Proof of Theorem 4.1.2 . . . . .	203
B.3 Proof of Theorem 4.2.1 . . . . .	208
B.4 Conditions of Theorem 4.2.2 . . . . .	209
B.5 Proof of Theorem 4.2.3 . . . . .	211
B.6 Proof of Theorem 5.3.1 . . . . .	213
B.7 Proof of Theorem 5.3.2 . . . . .	214
B.8 Proof of Theorem 6.2.1 . . . . .	215
B.9 Proof of Theorem 6.2.2 . . . . .	216
B.10 Proof of Theorem 6.3.1 . . . . .	218
<b>C Tables</b>	<b>222</b>
C.1 Tables . . . . .	222
<b>D Figures</b>	<b>235</b>

# List of Tables

C.1	MLE in Bivariate Model 1: estimating $\theta$ , $n=500$ (standard error of replications is given in brackets). . . . .	223
C.2	MLE estimates in Bivariate Model 2. Standard errors are given in brackets. Average computing time: 10.9 and 2.9 seconds for $N = 500$ and $N = 70$ , respectively. . . . .	224
C.3	MLE in Bivariate Model 2 in low tail dependence with large sample size. . . . .	225
C.4	Trivariate independent simulations and Reversible Jump estimations. . . . .	225
C.5	Bayesian estimations for Model 2 simulations. Standard errors are given in brackets. Average computing time: 76.8 and 21.5 seconds for $N = 500$ and $N = 70$ , respectively. . . . .	226
C.6	Bayesian estimations for data simulations. Standard errors are given in brackets. Average computing time: 11.9 seconds per replication. . . . .	227

C.7 Trivariate $M_5$ simulations and R. J. estimations: small and medium size. . . . .	228
C.8 Trivariate $M_5$ simulations and R. J. estimations: large size. . .	229
C.9 Trivariate $M_2$ simulations and R. J. estimations: large size. . .	229
C.10 Trivariate $M_2$ simulations and R. J. estimations: small and medium size. . . . .	230
C.11 Bayesian estimations for pair Indices and 95% Credibility Intervals for $\theta$ , and other extreme estimations (average computing time: 21.8s). . . . .	231
C.12 Bayesian Model Selection for trivariate Indices (average computing time: 22.2s). . . . .	232
C.13 Bayesian Trivariate estimates of triple tail dependence models.	233
C.14 Optimum allocations of different pairs of Global Indices. . . .	234
C.15 Predictive mean daily excess loss over 90% loss threshold of original pairs of Global Indices. . . . .	234

# List of Figures

D.1	Bivariate regions: $R_{01}$ , $R_{10}$ , and $R_{11}$ . . . . .	235
D.2	$\chi$ in the simple exponential case. . . . .	236
D.3	Examples of Model 1 Marginals. . . . .	236
D.4	Examples of Model 2 Marginals. . . . .	237
D.5	MCMC sample means for $\theta$ . . . . .	238
D.6	Bayesian predictions for data simulations. Predictions vs data in log-log scale. . . . .	239
D.7	Bayesian predictions for $\chi(u)$ vs empirical $\chi(u)$ . . . . .	240
D.8	Sampling processes of different weights for simulations of the trivariate model. . . . .	241
D.9	Simulated data and predictive sampling for trivariate simula- tions: medium triple tail dependence. . . . .	241
D.10	Simulated and predictive $\chi(u)$ and $\chi_3(u)$ for trivariate simula- tions: triple medium tail dependence. . . . .	242
D.11	Simulated data and predictive sampling for trivariate simula- tions: triple high tail dependence . . . . .	242

---

D.12 Simulated and predictive $\chi(u)$ and $\chi_3(u)$ for trivariate simulations: triple high tail dependence. . . . .	243
D.13 Simulated data and predictive sampling related graphics for trivariate simulations: exact independence . . . . .	244
D.14 Simulated data and predictive sampling related graphics for trivariate simulations: medium double tail dependence ( $X_3$ exact independent). . . . .	245
D.15 France CAC 40. . . . .	246
D.16 France CAC 40 filtered and unfiltered series. . . . .	246
D.17 Threshold selection methods. On the left side, the Mean Excess. On the right side, estimations of $k$ and its 95% confidence intervals. . . . .	247
D.18 Scatterplot of excesses over the 0.95 quantile of the filtered pairs: a) France-UK, b) France-Germany, and c) UK-US. . . . .	248
D.19 $\chi(u)$ of filtered pairs: a) France-UK, b) France-Germany, and c) UK-US. . . . .	248
D.20 Portfolio analysis tail region $RL$ . . . . .	248
D.21 Portfolio optimum allocations plots for different Global Indices. Computing time: 184.5 minutes per pair. . . . .	249
D.22 Portfolio Loss histograms. Left hand: historical, right hand: predictive. . . . .	250

# Chapter 1

## Introduction

In general terms, extreme events can be interpreted as: catastrophes, disasters, crisis, and crashes. For the purposes of this thesis, an extreme event can be defined as an event that has a small chance of occurrence, but has a high impact on the phenomenon in study.

The number of fields where there is an interest in the study of extremes is quite extensive: hydrology, seismology, climatology, economics, insurance, finance, epidemiology, medicine, and even in sports science.

Let us take the example of flooding. A common problem is to propose the height of a river barrier, such that it is feasible to build and it is unlikely to be breached. A useful way to approach this issue is to build the barrier up to a level, which is expected to be breached once every hundred years, for example. But to know what height this is, one needs to study extreme data. Similar examples or motivations can be posed for each of the fields listed.

In the thesis, we focus on financial extremes, keeping in mind the following questions: what is an extreme?, how likely is it to occur?, what would be the impact of its occurrence?, and what can be done with this information?

The case of the univariate framework has been developed quite extensively in the last 50 years. The aim of this work is to model multivariate behaviors of extremes. Multivariate extremes is a relatively new area and as such, it is still in development. Beside, the theory states that, unlike the univariate framework, there is not a unique parametric family of multivariate extreme distributions. An important part of a multivariate model is the dependence structure. The selection of the parametric form will determine the dependence structure, and in most of the cases, it is a rigid structure, in the sense that it hardly embraces different types of dependence among the variables.

The variables we are modelling are semi-heavy and heavy tailed variables, generically denoted by  $X$ . This means, firstly, that the variables do not have an upper limit (or lower limit, if the study is the left tail). The second consequence is that the variables we are modelling are such that:  $E(X^r) = \infty$ , for some  $r < \infty$ . This means our model does not include light-tailed variables. We consider this assumption as reasonable as a first attempt of modelling extremes.

We introduce a model that applies exclusively to the tails. We do not assume any model for any other region and it is not the concern of the thesis.

We propose a mixture of individual models which have the following features:

- They belong to the same parametric family.
- They embrace different dependence structure from each other.
- They have marginal coherence.
- They have a simple parametric form.

It is important to notice that we are assuming that the only form of asymptotic independence of each individual model is the exact independence. In other words, two variables can either be asymptotically dependent or exact independent.

Let  $\mathbf{X} = (X_1, \dots, X_d)$  be the random vector in study. Let  $C$  be the class of non-empty subsets of the set  $B = \{1, \dots, d\}$ . Denote  $C = (C_1, \dots, C_{d'})$ , where each  $C_j$  is an index variable over the set  $B$ , i.e.,  $C_j$  denotes a collection of the elements in  $\mathbf{X}$ . Note that  $\bigcup_{j=1}^{d'} C_j = B$ . For example, if  $C_j = (1, 2)$ , then  $\mathbf{X}_{C_j} = (X_1, X_2)$ ; as well,  $\mathbf{X}_B = \mathbf{X}$ .

We define a latent variable  $AD(C)$  such that:

$\mathbf{X} | AD(C_j)$  means the elements in  $\mathbf{X}_{C_j}$  are asymptotically dependent and the elements in  $\mathbf{X}_{B \setminus C_j}$  are exact independent, for  $j = 1, \dots, d'$ .

A natural question is to determine which of the variables  $AD(C_j)$ , for  $C_j \in C$ , are plausible for describing a given set of data. For example, if  $d = 3$ , interest would be on determining the feasibility of the different possible asymptotic dependencies among  $X_1, X_2$ , and  $X_3$ .<sup>1</sup>

---

<sup>1</sup>With this definition, it is clear that if  $C_j$ , a collection of a single element, is plausible,

The mixture model we propose is expressed, in general terms, as:

$$\sum_{j=1}^{d'} w_j f_j(\mathbf{X} | AD(C_j)),$$

where  $0 \leq w_j \leq 1$  and  $\sum w_j = 1$ , and  $f_j$  is the individual model associated to the asymptotic dependence structure given by the latent variable  $AD(C_j)$ .

The individual models are parametric models defined with a special construction of dependent variables, such that the asymptotic dependence relies only on one interpretable parameter. This parameter, let it be  $\theta$ , does not have any consequential influence on the univariate marginals. As well,  $\theta$  is common to all the individual models. Although, in principle, this fact seems restrictive, it must be noticed that the weights of the mixture model play as well an important role in the general dependence structure. Therefore, the relevant information regarding asymptotic dependence of individual model  $f_j$  is given by the pair  $(w_j, \theta)$ .

In general terms, the mixture models can be seen as model selection or model averaging interpretations. We study the application of both interpretations. However, we found that the model averaging approach respects the flexibility of the mixture model. Therefore we need to stress that this representation is not constructed so that only one model is chosen. In the same sense, we are not assuming that only one model is correct.

---

then all the collections with a single parameter are plausible. In fact, according to our model, all of them are simplified in a single model, where every variable is exact independent to each other.

For example, consider, once more, the case where  $d = 3$ . Suppose the asymptotic dependence of the pair  $(X_1, X_2)$  is stronger than in the pair  $(X_1, X_3)$ . It is clear that this behavior of data cannot be explained with the model selection approach, since the only asymptotic dependence information is contained in  $\theta$ , which is common for both pairs. However, the model averaging approach can tackle this problem. Suppose  $C_1 = (1, 2)$ , so that  $w_1$  is the weight of the model where  $X_1$  and  $X_2$  are asymptotically dependent. In a similar way, assume  $w_2$  is the weight of the model where  $X_1$  and  $X_3$  are asymptotically dependent, whereas  $w_3$  is the weight of the model where the three variables are asymptotically dependent. Therefore, the asymptotic dependence of the pair  $(X_1, X_2)$  is given by the information in  $(w_1, w_3, \theta)$ , whereas for the pair  $(X_1, X_3)$ , the information is in  $(w_2, w_3, \theta)$ . Hence, by no means, the mixture model, seen as model averaging, can be seen as rigid for having a common  $\theta$ .

In terms of inference problems in extremes, we study two cases. The first case is the main purpose of this model, which is the prediction, via the model averaging interpretation.

The second type of problem is the asymptotic dependence determination. In a broad sense, the task was to select the most likely individual model, i.e., a model selection problem. The reason to study this problem is to make a comparison with standard estimations in literature, since the interpretation of a dependence structure in a mixture model cannot be compared straightforward to a single model dependence structure.

It is our belief that the importance of the mixture model is not the determination of a unique model, but the prediction produced by spreading the uncertainty via a flexible model averaging.

Finally, the mixture models inference has been extensively developed inside the Bayesian paradigm, through the Markov Chain Monte Carlo and reversible jump methods. Therefore, this will constitute the inference basis of the thesis. The conjunction of these methodologies gives us the advantage of estimating the parameters of the individual models, the weights of the mixture model and the predictions in only one stage.

The overview of the thesis is as follows:

In Chapter 2, we present the fundamentals of the Extreme Value Theory, both the univariate and the multivariate cases.

We present some basics of the Bayesian paradigm in Chapter 3.

The following two chapters deal with the construction of a family of univariate and bivariate distributions which can be the basis of the mixture model.

In Chapter 4, we introduce the first bivariate models we constructed before getting to the final bivariate model. We study their properties and limitations.

The Chapter 5 consists of our final bivariate model. We present its probabilistic and statistical features. We do some simulation cases to show the inference procedures feasibility.

We present the generalization of the bivariate model in Chapter 6, which is the mixture model. We begin with the explanation of the trivariate model,

before stepping into the general multivariate model. As in the bivariate case, simulation cases are given for the trivariate case.

In Chapter 7, we proceed with the financial applications. It is divided mainly in two parts: a dependence structure study and a portfolio study case.

Finally, in Chapter 8, we present our conclusions.

# Chapter 2

## Extreme Value Theory

This section is dedicated to the foundations of Extreme Value Theory (EVT). We begin by explaining what an extreme event is, and then following on how it can be measured. In general, there are two methods to treat extremes: studying the behavior of the largest of the observations, or the observations exceeding a given threshold. These methods are, respectively, the Maximum and the Peaks Over the Threshold (POT) methods. Although the methodologies are different in principle, there is a strong relationship between them.

In the first part of this section (Section 2.1), we introduce the most important features of the univariate EVT. We present the connections and differences between both the Maximum and POT methods. The typical concept in this section is the asymptotic behavior of variables: how does a variable behave and what are its properties when it takes large values. The main goal is to develop asymptotic models for extremes and to describe their properties.

In Section 2.2 we consider the multivariate case. The easiest way to do this is to begin with the bivariate framework. Naturally, the multivariate EVT can be explained as well by the Maximum or the POT methodologies. As in the univariate case, the aim is to develop asymptotic models, but this time for a set of variables. One of the interesting issues in this section is to find the connections between multivariate and univariate cases: which properties are inherited or remain and which ones differ. We anticipate that there is an important difference in both frameworks, due mainly to the need of modeling the dependence among variables (as in any multivariate case). Therefore, in this section, special attention has to be kept on the dependence structure of the asymptotic models.

## 2.1 Univariate Framework

An extreme value can be explained as a very large or a very small value. What is a magnitude such that it can be called a large or small value depends on the nature of the data. What in a specific case could be called a large or small value, in a different context, can be a common value. And this is the essence of the extremes: they are not common, they are rare. Therefore, we need to know the nature of the data so that we can spot the difference between something common and something rare. Hence, in probability terms, one of the extremes features is that they have a low probability of occurrence.

However, in EVT, it does not suffice to be rare to define an extreme. As

we mentioned above, an extreme value is a large or small magnitude value. What we mean with this is that its presence must have an impact in order to be worthwhile to study. This impact is measured as the large or small magnitude. However, we focus on large extremes for the sake of explanation (later, it will become clear why this is the case).

Therefore, we now can define an extreme as an event of low probability of occurrence which has a large magnitude.

The next issue is to propose how to measure the presence of extremes. One way, given a sample, is to look at the largest value. This procedure is the basis of the maximum methodology, which is the beginning of our explanation.

### 2.1.1 Maximum Approach and Extreme Distributions

Assume that the data come from a specific non-degenerate probability distribution  $F$ . Define the right endpoint of  $F$  as follows:

**Definition 2.1.1.**  $X_0$  is the right endpoint of the probability distribution  $F$  of the random variable  $X$ , if and only if:

$$X_0 = \sup \{x \in \mathbb{R} : F(x) < 1\}.$$

Since we are not dealing with light tails, we have that  $X_0 = \infty$ . However, for explanation purposes of this chapter, we keep the notation of  $X_0$ .

The aim is to denote the large values as those which are in the neighborhood of the right endpoint of a distribution. Given a sample, a reasonable way to approach to the right endpoint is to study the sample maximum. The next

question is to study the properties of this statistic, so that we can establish if it is the only information needed.

Let  $X_1, X_2, \dots$  be a sequence of independent and identical distributed (i. i. d.) real variables with probability distribution  $F$ . Define  $M_n$  as the maximum of a sequence of size  $n$ . Then, its distribution can be expressed as follows:

$$\begin{aligned} P(M_n \leq x) &= P(X_1 \leq x, \dots, X_n \leq x) \\ &= F^n(x). \end{aligned} \tag{2.1}$$

The reason why we can focus on the maximum in our explanation is because the minimum framework can be implied by the following relation:

$$\text{Max}(\mathbf{X}) = -\text{Min}(-\mathbf{X}),$$

for  $\mathbf{X} = (X_1, \dots, X_n)$ .

Once we have the distribution of the maximum, we need to study its properties. The first property is that  $M_n$  converges almost surely (a.s.) to the right endpoint as we state in the next property (see Resnick (1987)).

**Property 2.1.1.** *Maximum a.s. convergence*

$$M_n \xrightarrow{\text{a.s.}} X_0.$$

We need to standardize  $M_n$  in a way such that we get a non-degenerate function. One concept that will help us to standardize the maximum is *max-stability* (see the introduction of this concept in de Haan (1984)), which we define now.

**Definition 2.1.2.** *Max-Stability*

A r.v.  $X$  is said to be max-stable if there exist constants  $a_n > 0$  and  $b_n \in \mathbb{R}$ , such that for  $X_1, \dots, X_n$ , independent copies of  $X$

$$M_n = \max(X_1, \dots, X_n) \stackrel{d}{=} a_n X + b_n; \quad (2.2)$$

for any integer  $n \geq 2$ .

The expression  $Y \stackrel{d}{=} Z$  stands for “ $Y$  has the same distribution as  $Z$ ”. In other words, if  $X$  is max-stable then, there exists a standardization  $M_n^*$  for the maximum

$$M_n^* = \frac{M_n - b_n}{a_n},$$

such that  $M_n^*$  and  $X$  have the same distribution. We will refer to such  $a_n$  and  $b_n$  as the normalizing constants.

Because of the ease of computation, the max-stability is an ideal property for a r.v. However, just a few distribution functions have this property. Nevertheless, as we will state further, the max-stable distributions are the only possible families of non-degenerate limit distributions for the maximum.

We present two cases as illustration of the max-stability and the lack of it.

**Example 2.1.1.** *Max-stability case*

Let  $F(x) = \exp(-x^{-\alpha})$ , for  $x, \alpha > 0$ . The question is to determine whether it is max-stable or not.

We need to find  $a_n$  and  $b_n$ , such that

$$F^n(a_n x + b_n) = F(x),$$

*i.e.*

$$\exp(-n(a_n x + b_n)^{-\alpha}) = \exp(-x^{-\alpha}),$$

which holds if we pick  $a_n = n^{1/\alpha}$  and  $b_n = 0$ .

**Example 2.1.2.** *Uniform. A non max-stability case*

Recall that the uniform distribution is  $F(x) = x$ , for  $x \in (0, 1)$ . As in the previous example, we are aiming to find if the max-stability holds for this model.

Thus, we need to find  $a_n$  and  $b_n$  such that

$$(a_n x + b_n)^n = x, \text{ for all } x \in (0, 1), n \in \mathbb{N}.$$

However, the left hand side of the equation is a polynomial of degree  $n$ , whereas the right hand side is of degree 1. Therefore, there is a solution only when  $n = 1$ . Therefore, we conclude that the Uniform distribution is not a max-stable distribution.

We present the standardized maximum characterization theorem, which constitutes the main theorem for the EVT, from Fisher and Tippett (1928) (refer to Resnick (1987) for the complete proof).

**Theorem 2.1.3.** *Fisher-Tippett theorem*

Let  $X$  be a non-degenerate r.v. with distribution function  $F$  and suppose there exists constants  $a_n > 0$  and  $b_n \in \mathbb{R}$ , such that

$$F^n(a_n x + b_n) \rightarrow G(x) \tag{2.3}$$

or in other words,

$$P\left(\frac{M_n - b_n}{a_n}\right) \rightarrow G,^1$$

as  $n \rightarrow \infty$ , for a non-degenerate distribution  $G$ . Then,  $G$  must be of one of the following types:

1. *Extreme Fréchet type*

$$G(x) = \Phi(x) = \begin{cases} 0, & x < 0, \\ \exp\{-(x)^{-\alpha}\}, & x > 0, \alpha > 0 \end{cases}$$

2. *Extreme Weibull type*

$$G(x) = \Psi(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & x \leq 0, \alpha > 0 \\ 1, & x > 0. \end{cases}$$

3. *Extreme Gumbel type*

$$G(x) = \Lambda(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R}.$$

We will refer to the three types as the extreme value distributions.

There are some important things to notice. The first one is that it is possible to transform among the three types of distributions. If  $X$  is an Extreme Fréchet variable, then  $\log(X^\alpha)$  has an Extreme Gumbel distribution. Similarly, if  $X$  is an Extreme Weibull variable then  $-\log(-X^\alpha)$  has an Extreme Gumbel distribution.

---

<sup>1</sup>This limit stands for convergence in distribution.

The second one is that a variable with the extreme Weibull distribution is different from a Weibull distributed variable. In fact, if  $X$  has a extreme Weibull distribution, then  $Y = -X$  is a Weibull distributed variable.

The last thing to note is that the extreme value distributions are the only max-stable distributions. On one side, it is easy to see that all the extreme distributions are max-stable. On the other side, if a distribution is max-stable, then the expression in (2.2) and the condition (2.3) are equivalent. Then if a distribution is max-stable, it must be of the type of one of the three extreme distributions.

The next question is: how does this result help with the inference of extremes? The answer relies on the concept of Maximum Domain of Attraction (MDA).

### 2.1.1.1 Maximum Domain of Attraction

We present the general ideas of the MDA. A wide extension of this topic can be found in the works of Resnick (1987), Kotz and Nadarajah (2002), and Embrechts *et al.* (1997).

The general problem in extreme values is the following: let  $X$  be a r.v. with a distribution function  $F$ , and let  $\mathbf{X} = (X_1, \dots, X_n)$  be a replica of  $X$ . Denote the survival function of  $F$  as:

$$\bar{F}(x) = P(X > x) = 1 - F(x).$$

The aim is to determine the distribution of the maximum  $M_n$ . Consider

Theorem 2.1.3. What follows is to find conditions for  $F$ ,  $a_n$ , and  $b_n$  in order to determine which extreme distribution is the corresponding maximum limit. These conditions constitute the concept of Maximum Domain of Attraction, which we define as follows.

**Definition 2.1.3.** *Maximum Domain of Attraction*

*A non-degenerate distribution  $F$  is said to belong to the Maximum Domain of Attraction of an Extreme distribution  $G$ , i.e.  $F \in MDA(G)$ , if there exist  $a_n > 0$  and  $b_n$  such that*

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x),$$

*or equivalently*

$$\lim_{n \rightarrow \infty} n\bar{F}(a_n x + b_n) = -\ln G(x).$$

Different types of Extreme distributions lead to different types of tail heaviness. Therefore, a way to characterize the MDA of an Extreme distribution is through a necessary rate of decay for the original distribution. We will only describe the main features of each of the domains and the distributions that belong to each of them. Refer to Resnick (1987), Kotz and Nadarajah (2002), and Embrechts *et al.* (1997) to see the conditions for  $F$  and the normalizing constants.

**Maximum Domain of Attraction of  $\Phi$ .**

The MDA of  $\Phi$  consists of regularly varying distributions, which are described in Appendix A. We can derive the domain of attraction if we study

the tail of  $\Phi_\alpha$ :

$$\bar{\Phi}_\alpha(x) = 1 - \exp(-x^{-\alpha}) = x^{-\alpha} - \frac{x^{-2\alpha}}{2!} + \dots \sim x^{-\alpha},$$

as  $x \rightarrow \infty$ . Therefore,  $\bar{\Phi}_\alpha \in RV_{-\alpha}$ , i.e.  $\bar{\Phi}_\alpha$  behaves as a power function for large values of  $x$  and the distributions that have it as their limit distribution of the maximum should behave similarly.

In general, the distributions that belong to this MDA are fat-tailed. This means that some moments might not exist. In fact, if we define the density as  $f(x) = -\frac{d}{dx} \bar{\Phi}(x)$ , then the  $m$ th-momentum of  $X$ , given by:

$$E[X^m] = \int_0^\infty x^m f(x) dx,$$

is finite iff  $m < \alpha$ .

The most important models that belong to this domain of attraction are the Pareto, the Burr, and the Cauchy distributions.

#### Maximum Domain of Attraction of $\Psi$ .

Since the Extreme Weibull distribution is related to the Fréchet with a change of sign and a reciprocal, then it involves as well power functions on the tail, but for finite right endpoint distributions. Two examples of distributions of this domain are the Uniform and the Beta distributions.

#### Maximum Domain of Attraction of the Gumbel Distribution, $\Lambda$ .

The Extreme Gumbel involves lighter tailed variables, since:

$$1 - \Lambda(x) = 1 - \exp\{-\exp\{-x\}\} = \exp(-x) + o(e^{-x}),$$

as  $x \rightarrow \infty$ . Therefore, it will characterize exponential tails. In contrast with the previous MDA, distributions with infinite or finite right endpoints can belong to the extreme Gumbel distribution.

In the Fréchet case, we found that some of the moments were not finite. In the Gumbel case, all the moments exist, as the following property notes (see Resnick (1987)).

**Property 2.1.2.** *Moments of Gumbel*

*Let  $X$  be a r.v. with distribution  $F$ , such that  $F \in MDA(\Lambda)$ . Then*

$$E[(X_+)^m] < \infty, \text{ for } m > 0,$$

*and  $X_+ = \max(0, X)$ .*

Some of the models that belong to this domain are the exponential, the normal, the Weibull, and the gamma distributions.

## 2.1.2 The General Extreme Value Distribution

We now present a model that includes all the extreme models detailed in the previous section. This general model facilitates inference, since it includes the three extreme distributions in one model: the generalized extreme value distribution of Jenkinson (1955).

**Definition 2.1.4.** *The Generalized Extreme Value Distribution*

*Define  $H$ , the generalized extreme value distribution (GEVD), as:*

$$H(x) = \exp \left\{ - \left( 1 + k \left( \frac{x - \mu}{\sigma} \right) \right)^{-1/k} \right\}, \quad (2.4)$$

for  $(1 + k(\frac{x-\mu}{\sigma})) > 0$ ,  $\mu, k \in \mathbb{R}$ , and  $\sigma > 0$ .

The three extreme distributions are included on this model, since:

- If  $k > 0$ , then  $H_k = \Phi$ ,  $\alpha = 1/k$ , and  $x > \mu - \frac{\sigma}{k}$ .
- If  $k < 0$ , then  $H_k = \Psi$ ,  $\alpha = 1/k$ , and  $x < \mu - \frac{\sigma}{k}$ .
- If  $k = 0$ , then  $H_k = \Lambda$ , since  $\lim_{k \rightarrow \infty} (1 + k(\frac{x-\mu}{\sigma}))^{-1/k} = e^{-(\frac{x-\mu}{\sigma})}$ , and  $x \in \mathbb{R}$ .

Therefore, with this representation, inference is just a matter of estimating the parameter set and the MDA will be given by the estimate of  $k$ .

There are two classical ways to make inference of this model. The first one divides the data into blocks of the same length. In each individual block, the maximum is registered. Of course, the smaller the length of the blocks, the more data will be obtained. However, the asymptotics behind (2.4) will not be plausible. On the other hand, if the length is chosen too large, the amount of data would be too scarce. The common procedure is to fix the length on a practical basis. For example, in environmental analysis, it is common to take the annual maximum in order to avoid any seasonality. In the financial case, a common choice is to take the daily maximum. In any of these examples, it is reasonable to suppose the maxima would be equally distributed. Thus, it is assumed the data follow the model in (2.4) exactly.

The next step is to use maximum likelihood estimation (MLE). This methodology has some drawbacks for estimating, since the support of the variable

depends on the parameter space. Smith (1985) showed the following features of the MLE:

- When  $k > -0.5$ , the information matrix is finite and standard asymptotic properties for MLE hold.
- When  $-1 < k < -0.5$ , the MLE exist but are not asymptotically normal.
- when  $k < -1$ , the MLE may not exist.

In practice, the issue of the choice of the norming constants,  $a_n$  and  $b_n$ , is solved as follows (see Coles (2001)):

Suppose

$$P\left(\frac{M_n - b_n}{a_n} < x\right) \approx H(x),$$

for large  $n$  and where  $H$  is a member of the GEV family.

Then,

$$P(M_n < x) \approx H\left(\frac{x - b_n}{a_n}\right) = H^*(x),$$

where  $H^*$  is a different member of the same GEV family (meaning, the same parameter  $k$ ). Hence, a GEVD can be fitted approximately to the untransformed maxima, without affecting the extreme value type. In practice, it is irrelevant to estimate the scale and location parameters of  $H^*$ , instead of those of  $H$ . In this way, it is not necessary to estimate the norming constants.

The second option is to fit the model assuming that the data belong to the MDA of  $H_k$ . It is clear that it is a completely different approach. Recall that

belonging to the MDA of  $H_k$  is equivalent to

$$\lim_{n \rightarrow \infty} n\bar{F}(a_n x + b_n) = -\ln H(x).$$

Then, the convergence is more related to the tail of the distribution, rather than the maximum.

Let  $X_1, X_2, \dots, X_n$  be i.i.d. observations with a distribution  $F$ , such that  $F \in MDA(H_k)$ . Define the order statistics as  $X_{1,n} \geq X_{2,n} \geq \dots \geq X_{n,n}$ . Then, the usual way to proceed is to take the largest observations, say the  $m$  largest values. The value of  $m$  must grow as  $n$  does, but it should remain small in comparison with  $n$ , since it is the tail behavior that is needed. In general terms,  $m(n) \rightarrow \infty$  and  $m(n)/n \rightarrow 0$ , as  $n \rightarrow \infty$ . We will present two typical estimators when dealing with this approach: the Pickands and the Hill estimators.

The first estimator is introduced in Pickands (1975).

**Definition 2.1.5.** *Pickands Estimator*

*We define the Pickands estimator of  $k$  as:*

$$\hat{k}_{m,n}^P = -\ln(2) \times \ln \frac{X_{m,n} - X_{2m,n}}{X_{2m,n} - X_{4m,n}}.$$

The consistency of this estimator depends on the choice of  $m$  (see Pickands (1975) for details), but with the conditions on  $m$  mentioned earlier, it is at least consistent in probability.

The next estimator will deal only with the  $k \geq 0$  case, and it is called the Hill's estimator (see Embrechts *et al.* (1997)). It uses the upper  $m$  statistics,

where  $m$  can be defined as in the Pickand's estimator.

**Definition 2.1.6.** *Hill's Estimator*

We define the Hill's estimator of  $k$  as:

$$\hat{k}_{m,n}^H = \left( m^{-1} \sum_{j=1}^m \ln X_{j,n} - \ln X_{m,n} \right)^{-1},$$

where  $m/n \rightarrow 0$ .

The different ways to derive this estimator can be found in Embrechts *et al.* (1997). They use the fact that the case  $k \geq 0$  consists of distributions whose tails are regularly varying functions.

Pickands and Hill's estimators give a different view on how to treat data: extreme relevant data does not necessarily consist of the block maxima, but on the observations that exceed a large value. This procedure is directly related to the excesses over the threshold approach. We will elaborate on this question in the next section.

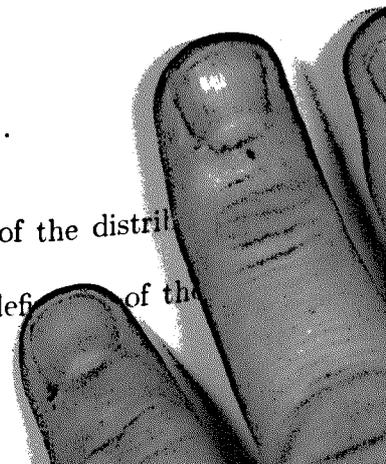
### 2.1.3 Peaks Over the Threshold Approach

Assume that the variable  $X$  belongs to the MDA for some  $H$ , as in (2.4).

Then, recall that the following condition holds:

$$\lim_{n \rightarrow \infty} n \bar{F}(a_n x + b_n) = \left( 1 + k \left( \frac{x - \mu}{\sigma} \right) \right)^{-1/k}.$$

We can see from this expression a hint of how the tail of the distribution should look like. Keeping this in mind, we make the definition of the



Generalized Pareto distribution. Balkema and de Haan (1974) derived it as a limit of Residual Life models, whereas in parallel, Pickands (1975) derived it working with order statistics. As well, there is the early work of Hill (1975) where a characterization of the tail of any distribution is given, and the final distribution proposed is a Pareto distribution.

**Definition 2.1.7.** *Generalized Pareto Distribution*

We define the distribution  $G_{k,\beta}$  for  $\beta > 0$  as the Generalized Pareto Distribution (GPD) by

$$G_{k,\beta}(x) = 1 - \left(1 + \frac{kx}{\beta}\right)^{-1/k}, \quad (2.5)$$

with  $x \geq 0$ , when  $k \geq 0$ , and  $0 \leq x \leq -\beta/k$ , when  $k < 0$ .

Of course, if we consider  $\lim_{k \rightarrow 0} G_{k,\beta}$ , we get the expression:  $G_\beta(x) = 1 - e^{-x/\beta}$ .

In order to extend on the importance of the GPD, we need to define the excess distribution (see Embrechts *et al.* (1997)).

**Definition 2.1.8.** *Excess Distribution*

Define the excess distribution  $F_u$  of a r.v.  $X$  over a value  $u$  by

$$F_u(x) = P(X - u \leq x | X > u).$$

We define the value  $u$  as a threshold. Therefore, a threshold  $u$  is the value where a distribution  $F_u$  can be defined for all the values above  $u$ .

The next theorem will present the relation between the excess distribution

and the GPD. Moreover, it will present a connection as well with the GEVD (see Embrechts *et al.* (1997)).

**Theorem 2.1.4. GPD-GEVD Relation**

Let  $H_{k,\sigma,\mu}$  denote a GEVD with shape parameter  $k$ , and let  $X$  be a r.v. with a distribution function  $F$  and a right endpoint  $X_0$ . Then  $F \in MDA(H_k)$  iff there exists a positive measure function  $a(\cdot)$  such that for  $1 + \xi x > 0$ :

$$\lim_{u \rightarrow X_0} P \left( \frac{X - u}{a(u)} > x, | X > u \right) = G_{\xi,1}(x). \quad (2.6)$$

Therefore, the limit distribution of a properly scaled excess over a high threshold is the GPD, with the same shape parameter as that of the domain of attraction of  $F$ .

It is useful to introduce the scale effect of  $a(u)$  into the GPD and thus, we get that  $\beta = a(u)$  for the expression (2.5). Hence, when a large enough threshold is chosen, the GPD is a good limit distribution for the excesses over this threshold, i.e.

$$P(X - u > x | X > u) \approx \overline{G}_{k,\beta}(x), \text{ for } x > 0. \quad (2.7)$$

It is worthwhile to remark on a couple of things. First of all, regardless of the approach of using the information (whether it is via the maximum or the excesses over a threshold), the main parameter to estimate is a shape parameter common to both approaches. This should not surprise us, since we saw in the Pickands' and Hill's estimators that when estimating  $k$  for the maximum in the MDA case, it was necessary to include some information from

the tails. In fact, both estimators involved the excesses over a fixed threshold (e.g., the Hill estimator used the  $m$ th order statistic).

The second thing to notice is that the scale parameter will depend on the threshold selected. The selection of the threshold will play an important role in inference. Notice what happens when we change a threshold level. Suppose we have a GPD variable  $X$ , say  $G_{k,\beta}$ , such that  $X > 0$ , i.e. the initial threshold is 0. Then, when we adjust the threshold level to  $u > 0$ , we get:

$$\begin{aligned} P(X > x + u | X > u) &= \frac{\left(1 + \frac{k}{\beta}(x + u)\right)^{-1/k}}{\left(1 + \frac{k}{\beta}u\right)^{-1/k}} \\ &= \left(1 + k\frac{x}{\beta + ku}\right)^{-1/k} \\ &= \bar{G}_{k,\beta+ku}(x). \end{aligned} \tag{2.8}$$

As would be expected, a change in threshold would only affect the scale parameter. Thus, the GPD is closed to changes in threshold levels. However, this property is not so simple when making inference. The higher the threshold is chosen, the better the asymptotics of  $G_{k,\beta}$  in Theorem 2.1.4 will hold. In this way, the higher the threshold, the better is the approximation in (2.7). Nevertheless, the amount of data included is less. Therefore, the variance of the estimates will be larger. We will extend on threshold selection in the next section.

For a  $x > u$ , the survival distribution can be split as follows:

$$\bar{F}(x) = P(X > u)P(X - u > x - u | X > u) = \bar{F}(u)\bar{F}_u(x - u). \tag{2.9}$$

For the second factor, define  $y = x - u$  as the excess. Then, with a proper

selection of  $u$  and by (2.7),  $\bar{F}_u(y) \approx \bar{G}_{k,\beta}(y)$ .

There are two ways to proceed for the first factor: choose a parametric or a non-parametric model, e.g. the empirical distribution.

Suppose  $u$  is given for a sample of size  $n$ . Let  $x_1^*, x_2^*, \dots, x_m^*$  be the observations that exceed  $u$  and  $x_{(1)}, x_{(2)}, \dots, x_{(n-m)}$ , the rest of them. As well, define  $\Theta$  as the whole parameter space. Assume that the empirical distribution is selected to model the observations that fail to exceed  $u$  and denote it as  $\hat{F}$ .

Then, the likelihood function is represented as

$$L(\Theta) = \left\{ \prod_{j=1}^{n-m} \hat{F}(x_{(j)}) \right\} \times \hat{F}(u)^m \prod_{i=1}^m \frac{1}{\beta} \left( 1 + \frac{k}{\beta}(x_i^* - u) \right)^{-1/k-1}. \quad (2.10)$$

The most common way to estimate the parameters is by MLE. As in the maximum case, the regularity conditions fail to hold when  $k < -0.5$ . Refer to Smith (1985) for details.

### 2.1.3.1 Threshold Selection and Inference

As we stated, the relevance of the selection of a threshold consists on a trade off between the variance and the bias of the excesses modeling. The task is to get the smallest threshold for which the GPD fits reasonably well the excesses.

The first option to treat the threshold (and most common) is to fix it at a high level. This level could even be imposed to the analysis beforehand. For example, consider the case of studying the losses over a million pounds in an insurance company, where this threshold is fixed just because a million pounds is a high value, rather than having a deeper statistical meaning.

We present two other options where the value of the threshold is estimated. The first one is based on the relation between the threshold and the mean excess function, and it can be found in Coles (2001) and Embrechts *et al.* (1997).

Define the mean excess function as follows.

**Definition 2.1.9.** *Mean Excess Function*

Let  $X$  be a r.v. with distribution function  $F$  and right endpoint  $X_0$ . Then, for a fixed  $u < X_0$ ,

$$e(u) = E[X - u | X > u]$$

is the mean excess function.

In the case of the GPD, the mean excess function is computed as follows:

$$e(u) = \frac{1}{\bar{F}(u)} \int_u^{X_0} \bar{F}(x) dx = \frac{\beta + ku}{1 - k},$$

for  $k < 1$ ,  $u < X_0$ , and  $\beta + ku > 0$ .

This shows a linear relation between  $e(u)$  and  $u$ , for  $\xi < 1$ .

Suppose there is a sample  $X_1, \dots, X_n$ , and for which  $X_1^*, \dots, X_m^*$  are the excesses over a threshold  $u$ . The mean excess function can be approximated by the empirical estimation:

$$e_n(u) = \frac{1}{m} \sum_{i=1}^m (X_i^* - u).$$

Then, the first threshold selection method consists on choosing a threshold  $u^*$  for which  $e(u)$  is approximately linear, for all  $u^* < u < X_0$ .

Although this method assumes that  $k < 1$ , it is widely used to fix a threshold for any type of tail behavior.

We now present the last threshold selection method. Suppose the GPD fits some data above a threshold. Then, recall from (2.8) that a change in the threshold only affects the scale parameter. In other words, the value of  $k$  does not change. In this way, the method consists on choosing a value  $u^*$ , for which the estimates  $\hat{k}_u$  are considerably stable, for any  $u^* < u < X_0$ .<sup>2</sup>

The last two methods are graphical based. It is a matter of visually sensing where a linear or a stable behavior is achieved. The difference between the two methods is that the first one is an estimation previous to inference. Whereas for the second method, estimates must be carried out for a range of different thresholds.

We present an application of this method on real data on Figure D.17. On the real data analysis of Section 7.1, we aimed to select a threshold as a quantile of the marginal data. In fact, we opted to choose the same quantile for all the variables. Following the criteria of the threshold selection methods we described, the figures suggest a linearity on ME above a threshold of a quantile of 0.96. For the estimates of  $k$ , we present as well its 95% confidence intervals. In some of the cases, when the threshold was above the quantile 0.98, the estimates were below -0.5. Therefore, the confidence intervals could not be computed, since the regularity conditions do not hold. However, around

---

<sup>2</sup>The parameter  $\beta_u$  has to be estimated as well.

the selected threshold (quantile 0.96), the estimates were just above 0. Then, the regularity conditions hold, and the estimates suggest some stability.

The topic of goodness of fit in extremes has been explored extensively. The first natural methods consist on graphics, for example:  $Q - Q$  and  $\log Q - Q$  plots, and for the POT approach, the mean excess function graphic (see, for example, Beirlant et al. (2006) or Coles (2001)). However, in the POT approach, if the threshold is chosen based on the mean excess function plot, as we explained, it is not possible to use it as a goodness-of-fit tool.

In the case of goodness-of-fit parametric methods, a wide treated case is testing  $k = 0$  vs  $k \neq 0$  (see (2.4) and (2.5)). The works of Bardsley (1977), Hosking (1984), and Otten and van Montfort (1978) focus on the maximum approach, i.e. the goodness-of-fit of the GEVD, for both the  $k = 0$  vs  $k \neq 0$ , and the one sided hypothesis tests (either  $k > 0$  or  $k < 0$ ), and Stephens (1977) focuses only on testing  $k = 0$  vs  $k \neq 0$ .

For the POT approach, Davison and Smith (1990) treat the case  $k = 0$  vs  $k \neq 0$  in the GPD, whereas Beirlant et al. (2006) and Choulakian and Stephens (2001) test the GPD fit for MLE's. This goodness-of fit literature works with modification of tests as: Kolmogorov-Smirnov, Anderson and Darling, Jackson and Cramér-von Mises.

We end the univariate section with a summary of the extreme values fundamentals. We found that there are two approaches to deal with the extremes: the maximum and the excesses over a threshold approaches. Whichever ap-

proach is chosen, the model is linked with the behavior of the variable on its tail. This tail behavior is characterized by one single shape parameter. Both approaches are so related, that both models share this shape parameter. We found that, although theoretically, the maximum approach does estimate the tail behavior via the shape parameter, some information can be discarded from the sample. This is why we prefer the POT framework when making inference.

## 2.2 Multivariate Framework

During the last twenty years or so, significant attention has been given to inference methods for multivariate extremes, following original work of Smith *et al.* (1990) and Tawn (1988). Such a framework has not been easy to set up and most of the literature has focused on the bivariate case. Any multivariate extreme model has two aspects: the marginal model and the dependence structure. The marginal models, as we explained in the previous section, are typically members, or belong to the domain of attraction, of the univariate extreme value family of distributions.

The main focus on multivariate study is the dependence structure on the extremes. Since any attempt to model this structure should be able to embrace light and heavy tails (meaning that some moments might not exist), traditional dependence measures of covariance such as the Pearson correlation coefficient or Spearman's  $\rho$  are not useful.

In the multivariate case, the first attempts to create a dependence structure

on the distribution of the maximum are given in Gumbel (1958), Tiago de Oliveira (1958), Sibuya (1960), and the point process representations of de Haan and Resnick (1977). We will focus on a version of de Haan and Resnick (1977), where it can be seen that the multivariate extreme models do not have a unique representation (see as well Resnick (1987), Kotz and Nadarajah (2002), Coles and Tawn (1991), and Ledford and Tawn (1996)).

The aim of this section is to set the foundations of the representations of the multivariate extreme models and their properties.

We will denote vectors as  $\mathbf{X}$ , for r.v., and  $\mathbf{x}$  for fixed values. We will keep the subindex for the sequences indices and for order statistics. Thus, to denote the elements of a vector, we will use the superindex, e.g.  $\mathbf{x} = (x^{(1)}, \dots, x^{(d)})$ . And to order vectors:  $\mathbf{x} > \mathbf{y}$  means  $x^{(j)} > y^{(j)}$ , for  $j = 1, \dots, d$ .

### 2.2.1 Multivariate Maximum Asymptotics

The first framework we will explain can be found in de Haan and Resnick (1977), Resnick (1987), and Kotz and Nadarajah (2002).

The aim is to model the behavior of a random vector  $\mathbf{X} = (X^{(1)}, \dots, X^{(d)}) \in \mathbb{R}^d$ . Assume  $\mathbf{X}$  has a joint distribution  $F$ , which means that for a fixed vector  $\mathbf{x} = (x^{(1)}, \dots, x^{(d)})$ :

$$F(\mathbf{x}) = P(X^{(1)} \leq x^{(1)}, \dots, X^{(d)} \leq x^{(d)}).$$

In multivariate analysis, there is no unique definition of the maximum, since there are different ways to order vectors. The common practice is to define it

as the vector of the marginal maximum (therefore, the sample maximum are not necessarily real observed data).

Suppose  $\{(X^{(i,1)}, \dots, X^{(i,d)}), i = 1, \dots, n\}$  are i. i. d.  $d$ -variate vectors with distribution  $F$ . Define  $\mathbf{M}_n$ , the sample maximum as

$$\mathbf{M}_n = (M_n^{(1)}, M_n^{(2)}, \dots, M_n^{(d)}),$$

where  $M_n^{(j)} = \max(X^{(1,j)}, \dots, X^{(n,j)})$ , for  $j = 1, \dots, d$ .

As in the univariate case, we seek to find a limit distribution for  $\mathbf{M}_n$  and its properties. The distribution of the maximum is computed as follows:

$$\begin{aligned} P(\mathbf{M}_n \leq \mathbf{x}) &= P(M_n^{(1)} \leq x^{(1)}, \dots, M_n^{(d)} \leq x^{(d)}) \\ &= \prod_{i=1}^n P(X^{(i,1)} \leq x^{(1)}, \dots, X^{(i,d)} \leq x^{(d)}) \\ &= (F(\mathbf{x}))^n \\ &= F^n(\mathbf{x}). \end{aligned}$$

The class of the max-infinitely divisible (max-id) distributions is that where  $F^n$  is a distribution for  $n \geq 0$ . We use the following theorem to define this class (see Resnick (1987)):

**Theorem 2.2.1.** *Max-Infinitely Divisible Distributions*

*A distribution  $F$  is max-id if and only if  $F^{1/t}$  is a distribution  $\forall t > 0$ .*

Regarding the distribution of the maximum, if the joint probability of the maximum is a max-id distribution, then it follows that the joint distribution of the original variables is a distribution as well.

The max-id concept will play the role that max-stability did in the univariate case, see (2.2) in Section 2.1: It will characterize the extreme multivariate distribution.

As in Section 2.1, see (2.3), in the univariate case, seeking for a non-degenerate limit distribution of the standardized maximum is analogous to seeking constants  $a_n^{(1)} > 0, \dots, a_n^{(d)} > 0$  and  $b_n^{(1)}, \dots, b_n^{(d)}$ , such that for  $M_n$  and  $\mathbf{X}_n$  defined as before:

$$\begin{aligned} P \left( \frac{M_n^{(1)} - b_n^{(1)}}{a_n^{(1)}} \leq x^{(1)}, \dots, \frac{M_n^{(d)} - b_n^{(d)}}{a_n^{(d)}} \leq x^{(d)} \right) \\ = F^n \left( a_n^{(1)} x^{(1)} + b_n^{(1)}, \dots, a_n^{(d)} x^{(d)} + b_n^{(d)} \right) \longrightarrow G(\mathbf{x}), \text{ as } n \rightarrow \infty, \end{aligned} \quad (2.11)$$

where  $\mathbf{x} = (x^{(1)}, \dots, x^{(d)})$ . We will call  $G$  the class of multivariate EV distributions.

If we let  $x^{(m)} = \infty$  in (2.11), for all  $m \neq j$ , for some  $j \in \{1, \dots, d\}$ , then we will get the  $j$ -th marginal,  $F_j$ . Therefore, the limit of the marginal distribution is expressed as follows:

$$F_j^n \left( a_n^{(j)} x^{(j)} + b_n^{(j)} \right) \longrightarrow G_j(x^{(j)}), \text{ as } n \rightarrow \infty, \quad (2.12)$$

for  $j = 1, \dots, d$ . Hence, the marginals of  $G$  must be a non-degenerate EV distribution.

Two special extreme (in terms of dependence structure) cases arise on  $G$ : the asymptotic exact independence and complete dependence cases.

**Definition 2.2.1.** *Asymptotic Exact Independence and Complete Dependence*

*The EV distribution  $G$  follows an asymptotic exact independence if*

$$G(\mathbf{x}) = G_1(x^{(1)}) \cdots G_d(x^{(d)}). \quad (2.13)$$

*The EV distribution  $G$  follows asymptotic complete dependence if*

$$G(\mathbf{x}) = \min \{G_1(x^{(1)}), \dots, G_d(x^{(d)})\}.$$

We will discuss more dependence structures in the following sections.

## 2.2.2 Characterization of the Multivariate EVD

In this section we will find a characterization of the multivariate EV distribution (MEVD)  $G$ .

Recall the max-stability concept from Section 2.1, see (2.2). In a similar way, we can define it for the multivariate case.

**Definition 2.2.2.** *Multivariate Max-stability*

*A distribution  $G(\mathbf{x})$  is max-stable for  $\mathbf{x} \in \mathbb{R}^d$ , if there exist functions  $\alpha^{(j)}(t) > 0$  and  $\beta^{(j)}(t)$  for  $j = 1, \dots, d$  and all  $t > 0$ , such that for any  $\mathbf{x}$ :*

$$G^t(\mathbf{x}) = G(\alpha^{(1)}(t)x^{(1)} + \beta^{(1)}(t), \dots, \alpha^{(d)}(t)x^{(d)} + \beta^{(d)}(t)).$$

Therefore,  $G^{1/n}$  is a distribution for all  $n > 0$ , then all the max-stable distributions are max-id. As in the univariate case, the class of max-stable distributions have an important use as the next theorem shows (cf. Resnick (1987)).

**Theorem 2.2.2.** *The class of multivariate extreme value distributions is the class of the max-stable distributions with non-degenerate marginals.*

This theorem gives us the first characterization of the MEVD: these are max-stable distributions.

In order to go further in the analysis, it is useful to separate the marginal effects from the joint distribution via standardization. In the next definition, we present the inverse function.

**Definition 2.2.3.** *Inverse Function*

*Let  $F$  be a non decreasing function on  $\mathbb{R}$ . Then, define the inverse (or more precisely, the left continuous inverse) as:*

$$F^{\leftarrow}(x) = \inf \{y : F(y) \geq x\}.$$

The most common way to standardize the variables is through the marginal transformation

$$U_j(X^{(j)}) = -\frac{1}{\log F_j(X^{(j)})},$$

so that the marginal distributions are unit Fréchet with distribution function:

$$F_j(x^{(j)}) = \exp \left\{ -\frac{1}{x^{(j)}} \right\},$$

for  $j = 1, \dots, d$ . This transformation does not change the multivariate extreme structure of the joint distribution, as the following theorem shows (cf. Resnick (1987) for the proof).

**Theorem 2.2.3.** *Fréchet Transformations*

Suppose  $G$  is a MEVD with continuous marginals  $G_j$ . Define the transformations

$$U_j = -\frac{1}{\log G_j(X^{(j)})}, \text{ for } j = 1, \dots, d,$$

and for  $\mathbf{x} \in \mathbb{R}^d$

$$G^*(\mathbf{x}) = G(U_1^{\leftarrow}(x^{(1)}), \dots, U_d^{\leftarrow}(x^{(d)})), \quad (2.14)$$

then,  $G^*$  has unit Fréchet marginals and  $G$  is a MEVD iff  $G^*$  is MEVD.

We now present the characterization of the domain of attraction of the MEVD.

### 2.2.2.1 A Necessary Condition for MDA

The first condition is elaborated from a Poisson Process point of view and it can be found in Kotz and Nadarajah (2002).

Let  $\mathbf{X}^{(i)} = (X^{(i,1)}, \dots, X^{(i,d)})$ , for  $i = 1, \dots, \infty$ , be random vectors with distribution function  $F \in MDA(G)$  and marginals  $F_j$ . Define as well  $U_j(X^{(i,j)}) = -1/\log F_j(X^{(i,j)})$ , for  $j = 1, \dots, d$ . Let,  $S_d$  be the  $(d-1)$ -dimensional unit simplex:

$$S_d = \left\{ (w_1, \dots, w_{d-1}) : \sum_1^{d-1} w_k \leq 1, w_k \geq 0, k = 1, \dots, d-1 \right\},$$

and define the transformation vector:

$$T(y^{(1)}, \dots, y^{(d)}) = \left( \sum_1^d y^{(k)}, y^{(1)}/\sum_1^d y^{(k)}, \dots, y^{(d)}/\sum_1^{d-1} y^{(k)} \right).$$

Then

$$P_n = \{(U_1(X^{(i,1)})/n, \dots, U_d(X^{(i,d)})/n), i = 1, \dots, n\} \rightarrow P, \text{ as } n \rightarrow \infty,$$

where  $P$  is a non-homogeneous Poisson Process on  $\mathbb{R}_+^d$  with an intensity measure  $\mu$  such that

$$\mu(T^{\leftarrow}(dr, d\mathbf{w})) = r^{-2} dr H(d\mathbf{w}),$$

for  $r > 0$  and  $\mathbf{w} \in S_d$ , where  $H$  is a non-negative measure on  $S_d$  such that

$$H(S_d) = d \text{ and } \int_{S_d} w_k H(d\mathbf{w}) = 1, \text{ for } k = 1, \dots, d-1.$$

Then, the transformed MEVD  $G^*$ , as in (2.14), can be written as

$$G^*(y^{(1)}, \dots, y^{(d)}) = \exp \{-V(y^{(1)}, \dots, y^{(d)})\}, \quad (2.15)$$

where  $V$  is called the exponent measure function, such that:

$$\begin{aligned} V(y^{(1)}, \dots, y^{(d)}) &= \mu(\{[0, y^{(1)}] \times \dots \times [0, y^{(d)}]\}^c) \\ &= \int_{S_d} \max\left(w_1/y^{(1)}, w_2/y^{(2)}, \dots, \left(1 - \sum_{k=1}^{d-1} w_k\right)/y^{(d)}\right) H(d\mathbf{w}). \end{aligned}$$

It is clear that there is not a unique representation of the limit MEVD.

As it can be seen, this representation works with pseudo-polar coordinates via  $T$ . However, it is  $H$  which determines the dependence structure of the model. The strength of the dependence will rely on where  $H$  puts the mass.

Therefore, there are as many limit distributions  $G^*$  as  $H$  functions can be defined. There are two ways to define  $H$ : with a parametric or with a non-parametric model. As examples of parametric models, we have the Logistic, Bilogistic, and Gaussian models. Different parametric models lead to different

type of dependence (see Tawn (1988), Tawn (1990), and Smith *et al.* (1990)), and as we will show, some models can embrace different types of dependence. Therefore, it is a difficult task to construct a parametric family of models which can be flexible enough to fit the complexity in the multivariate extremes framework. For the purposes of the thesis, we are interested in the complexity of the dependence among the variables. The aim in this respect, therefore, is the construction of a flexible model for which a high number of types of dependence is achieved. We will address again this issue in the bivariate section. For more models and their properties, refer to Joe (1994), Tawn (1994), and Kotz and Nadarajah (2002).

We now focus on the bivariate case. The bivariate case is the most studied and it is where we need to point out further concepts and theories.

### 2.2.3 Bivariate Framework

In this section we will present the main results of the bivariate extremes framework. To begin with, we give the bivariate equivalent characterization of Section 2.2.2.1 (refer to Coles (2001)). We then give a second representation of the maximum, the representation of Pickands (1981). Refer to Kotz and Nadarajah (2002) for more characterizations.

### 2.2.3.1 Bivariate Maximum Model

Let  $X$  and  $Y$  be independent random variables, unit Fréchet distributed (i.e.  $F_X(x) = e^{-1/x}$ , and equivalently for  $Y$ ). Let  $X_i$  and  $Y_i$  be copies of  $X$  and  $Y$ , respectively for  $i = 1, \dots, n$ . Define:

$$\mathbf{M}_n^* = (n^{-1}\max(X_1, \dots, X_n), n^{-1}\max(Y_1, \dots, Y_n)). \quad (2.16)$$

The form of the marginal distributions of  $\mathbf{M}_n^*$  are expressed as:

$$P(n^{-1}\max(X_1, \dots, X_n) \leq x) = e^{-1/x}, \quad x > 0, \text{ for all } n \geq 1,$$

(and similarly for the second marginal) since the Fréchet is a max-stable distribution.

We present the equivalent version of Theorem 2.1.3 (see Coles (2001)).

**Theorem 2.2.4.** *Define  $\mathbf{M}_n^*$  as in (2.16) for  $(X_i, Y_i)$ , independent vectors with unit Fréchet marginals, for  $i = 1, \dots, n$ . Then, if*

$$P(\mathbf{M}_n^* \leq \mathbf{x}) \longrightarrow G(\mathbf{x}), \text{ as } n \rightarrow \infty,$$

for  $\mathbf{x} = (x, y)$ ,  $x > 0, y > 0$ , and where  $G$  is a non-degenerate distribution.

Then,  $G$  has the following representation:

$$G(x, y) = \exp \{-V(x, y)\}; \quad (2.17)$$

where

$$V(x, y) = 2 \int_0^1 \max(w/x, (1-w)/y) dH(w), \quad (2.18)$$

for  $H$ , a distribution satisfying:

$$\int_0^1 w dH(w) = 1/2. \quad (2.19)$$

For example, define:

$$dH(w) = 0.5 \mathbf{1}_{[w=1]} + 0.5 \mathbf{1}_{[w=0]}.$$

Then,  $G$  is expressed as:

$$G(x, y) = \exp \left\{ - (x^{-1} + y^{-1}) \right\}.$$

Therefore,  $X$  and  $Y$  are asymptotically exact independent.

Similarly, define:

$$dH(w) = \mathbf{1}_{[w=0.5]}.$$

Then,  $G$  is has the following form:

$$\begin{aligned} G(x, y) &= \exp \left\{ -\max(x^{-1}, y^{-1}) \right\} \\ &= \min \left\{ \exp(-x^{-1}), \exp(-y^{-1}) \right\}. \end{aligned}$$

Therefore,  $X$  and  $Y$  are asymptotically complete dependent.

A feature to notice about  $V$  is that it is a homogeneous function of order  $-1$ , which means:

$$V(cx, cy) = c^{-1}V(x, y), \quad (2.20)$$

for a constant  $c > 0$ .

Another equivalent expression for  $G$  is the representation of Pickands (1981):

$$G(x, y) = \exp \left\{ - \left( \frac{1}{x} + \frac{1}{y} \right) A \left( \frac{x}{x+y} \right) \right\};$$

where:

$$A(w) = \int_0^1 \max \{w(1-q), (1-w)q\} dH(q),$$

and its main features are:

- $\max(w, 1-w) \leq A(w) \leq 1$  for  $0 \leq w \leq 1$ ;
- $A$  is convex.

Similarly to the previous examples, if we set  $A(w) = 1, \forall w \in [0, 1]$ . Then  $X$  and  $Y$  are asymptotically exact independent, whereas when  $A(w) = \max(w, 1-w)$ , then  $X$  and  $Y$  are asymptotically complete dependent.

Since these conditions and (2.19) are satisfied for an infinite number of functions (meaning infinite possible forms for  $V$  or  $A$ ), there is no unique representation of the distribution of the maximum in the bivariate case.

As we stated in the multivariate representation, parametric models can be assumed for  $V$  or  $A$ , and therefore, different measures lead to different type of dependence, and as we will show, some models can embrace different types of dependence.

Similarly to the univariate maximum approach, the selection of the block for the maximum is an issue to consider. As well, some information might be wasted when considering this approach. Therefore, we present the bivariate version of Section 2.1.3.

### 2.2.3.2 Bivariate POT Approach

In this section, we present the bivariate version of the POT approach. The generalization to the multivariate case is straightforward from (2.22).

Recall from (2.9) that a univariate tail can be split as:

$$\bar{F}_X(x) = P(X > u)P(X - u > x - u | X > u) = \bar{F}_X(u)\bar{F}_{X,u}(x - u),$$

where  $\bar{F}_{X,u}$ , can be approximated by a GPD for a large  $u$ , i.e.

$$P(X > x) \approx P(X > u) \left(1 + \frac{k}{\beta}(x - u)\right)^{-1/k}, \text{ for } x > u.$$

A simplistic procedure is to introduce the effect of  $P(X > u)$  in one parameter (see Coles (2001) for estimations and implications), as follows:

$$P(X \leq x) = 1 - \varsigma_u \left(1 + \frac{k}{\beta}(x - u)\right)^{-1/k}, \text{ for } x > u. \quad (2.21)$$

Suppose  $X$  and  $Y$  have a joint distribution  $F$  and marginal distributions as the expression in (2.21), for  $X > u_x$  and  $Y > u_y$ . Then, the variables

$$X^* = - \left( \log \left( 1 - \varsigma_x \left( 1 + \frac{k_x}{\beta_x} (X - u_x) \right)^{-1/k_x} \right) \right)^{-1}, \text{ and}$$

$$Y^* = - \left( \log \left( 1 - \varsigma_y \left( 1 + \frac{k_y}{\beta_y} (Y - u_y) \right)^{-1/k_y} \right) \right)^{-1},$$

have marginal unit Fréchet distributions for  $X > u_x$  and  $Y > u_y$ , and the joint distribution  $F^*$  of the pair  $(X^*, Y^*)$  can be approximated, for large  $n$ , as (see

Coles (2001)):

$$\begin{aligned} F^*(x^*, y^*) &= (F^n(x^*, y^*))^{1/n} \\ &\approx \exp\{-V(n^{-1}x^*, n^{-1}y^*)\}^{1/n} \\ &= \exp\{-V(x^*, y^*)\}. \end{aligned}$$

The second line holds from Theorem 2.2.4, whereas the third line, from (2.20).

This means that for large  $u_x$  and  $u_y$ :

$$F(x, y) = F^*(x^*, y^*) \approx \exp\{-V(x^*, y^*)\}, \text{ for } x > u_x, y > u_y. \quad (2.22)$$

The next question is how we can fit model (2.22) to extreme data. The answer is formulated by censoring the observations as stated in Smith (1994) and Ledford and Tawn (1996). We divide  $\mathbb{R}^2$  into four different regions as follows:

$$\{R_{ij}; i = \mathbf{1}_{[X > u_X]}, j = \mathbf{1}_{[Y > u_Y]}\}, \quad (2.23)$$

as Figure D.1 shows. Then, model (2.22) applies directly to region  $R_{11}$ . However, data falling on other regions also contribute to the information of the tails. For example, an observation in  $R_{10}$  means only  $x$  is extreme. Hence, the contribution to the likelihood function of this observation is given by:

$$P(X = x, Y < u_Y) = \frac{\delta F}{\delta x} \Big|_{(x, u_Y)}.$$

Therefore, a  $Y$  is censored, i.e. it is only known to be below  $u_Y$ .

The censoring process is explained as follows:

Let  $\mathbf{x} = \{(x_1, y_1), \dots, (x_n, y_n)\}$  be a replica of size  $n$  of the pair  $(X, Y)$  with joint distribution function  $F$  with parameter  $\omega$ , for which (2.22) holds for high thresholds  $u_X$  and  $u_Y$  of  $X$  and  $Y$ , respectively. Then, the likelihood function for the extreme data is given by:

$$L(\omega | \mathbf{x}) = \prod_{i=1}^n g(\omega | (x_i, y_i)),$$

where

$$g(\omega | (x_i, y_i)) = \begin{cases} \frac{\delta^2 F}{\delta x \delta y} \Big|_{(x,y)} & \text{if } (x, y) \in R_{11} \\ \frac{\delta F}{\delta x} \Big|_{(x, u_y)} & \text{if } (x, y) \in R_{10} \\ \frac{\delta F}{\delta y} \Big|_{(u_x, y)} & \text{if } (x, y) \in R_{01} \\ F(u_x, u_y) & \text{if } (x, y) \in R_{00}, \end{cases} \quad (2.24)$$

where  $F(x, y)$  is as in (2.22).

It is important to notice that, although this model incorporates the information in the regions  $R_{10}$ ,  $R_{01}$ , and  $R_{00}$ , the support is on the region  $R_{11}$ . This fact will play an important role when we introduce our model, since we will present two versions of how to treat data in our model: a modification of the censoring of Smith (1994) and a modeling of the three regions  $R_{10}$ ,  $R_{01}$ , and  $R_{11}$  (see Section 5.5).

As in the bivariate maximum approach, the joint distribution in the tails depends on the definition of the measure  $V$ . Therefore, the representation is not unique and the usual way to proceed is to define a parametric model for  $V$ . The choice of  $V$  will determine the dependence structure of the model. In

order to study the joint tail behavior implied by a form of  $V$  and to compare it with other families, some extreme parameters must be introduced.

### 2.2.3.3 Bivariate Extreme parameters

The first parameter to define was introduced in Sibuya (1960). It measures the asymptotic dependence between two variables when considering the tail. This parameter will play an important role in the following chapters, since it will determine the dependence structure of our bivariate model.

**Definition 2.2.4.** *Tail Dependence parameter  $\chi$*

*Let  $X$  and  $Y$  have the same marginal distribution. Then define the following asymptotic dependence measure:*

$$\chi = \lim_{u \rightarrow \infty} P(X > u | Y > u). \quad (2.25)$$

*The following cases are defined from this parameter:*

- $\chi = 1$ :  $X$  and  $Y$  are complete tail dependent.
- $\chi = c > 0$ :  $X$  and  $Y$  are asymptotically or tail dependent.
- $\chi = 0$ :  $X$  and  $Y$  are asymptotically or tail independent.

Define, as well, as in Coles (2001), the following parameter for  $0 < r < 1$ :

$$\begin{aligned} \chi(r) &= 2 - \frac{\log P(F_X(X) < r, F_Y(Y) < r)}{\log P(F_X(X) < r)} \\ &= 2 - \frac{\log P(F_X(X) < r, F_Y(Y) < r)}{\log r}, \end{aligned} \quad (2.26)$$

for  $F_X$  and  $F_Y$ , the marginal distributions of  $X$  and  $Y$ . The relation with the parameter  $\chi$  is the following:

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

The parameter  $\chi(r)$  is a helpful empirical tool for inference. The value in (2.26) can be plotted for different values of  $r$ , and the graphic can help to have a guess when  $r$  approaches 1 where  $\chi$  is expected to be. In Chapter 6 we present some examples of this parameter.

In fact,  $\chi$  has the following relation with the exponent measure in (2.18):

$$\chi = 2 - V(1, 1).$$

Recall the Figure D.1 (in this case  $u_x = u_y$ ). A complete tail dependence ( $\chi = 1$ ) case would mean that all the observations for which the  $Y$  component is above  $u_y$  are expected to lie on  $R_{11}$ , for a large  $u_y$ .

Some problems can arise when  $\chi$  is close to zero in the classical inference approach, i.e. when there is near independence. For example, Coles and Tawn (1991) found some examples where the exponent measure  $V$  (recall (2.18)) places all the mass on the vertices of  $S_p$ , thus, the densities of  $H$  (recall (2.18)) present discontinuities. This problem is found in both approaches: the maximum and the POT. In fact, as Coles and Tawn affirm, in the POT approach, the joint distribution fails to factorise into the product of the marginals in the tail independence or near independence cases, and even more, the joint distribution gives as well a zero probability to the case of both variables exceeding

a high threshold.

On the other hand, Tawn (1988) found some examples where the independence case implies nonregular behavior in the classic estimation framework. However, Tawn was able to find in a couple of cases an asymptotic behavior of the MLE of the parameters. As a consequence of these difficulties in the classical approach, Tawn's tests of independence present some problems when constructing them. This framework can only test exact independence against the tail dependence.

Suppose  $X$  and  $Y$  are unit Fréchet distributed and nonnegative associated. Ledford and Tawn (1996) introduced a way to estimate smoothly the dependence structure between the asymptotic dependence and exact independence. Ledford and Tawn had the target to set a model between the following two cases:

$$P(X > r, Y > r) \sim \begin{cases} cr^{-1} & \text{for asymptotic dependence, and } 0 < c \leq 1. \\ r^{-2} & \text{for exact independence,} \end{cases} \quad (2.27)$$

for large  $r$ .

Thus, Ledford and Tawn introduced a parameter  $\eta$  (defined as the coefficient of tail dependence) that governs the power of  $r$  in between these extreme dependence cases, such that:

$$P(X > r, Y > r) \sim \mathcal{L}(r)r^{-1/\eta}, \text{ as } r \rightarrow \infty, \quad (2.28)$$

where  $0 < \eta \leq 1$ , and  $\mathcal{L}(r)$  is a slow varying function, meaning that:

$$\lim_{t \rightarrow \infty} \mathcal{L(tr)/\mathcal{L}(t) = 1,$$

for fixed  $r > 0$ .

We can relate  $\chi$  and  $\eta$  with the following expression:

$$P(X > r | Y > r) \sim \mathcal{L}(r)r^{1-1/\eta}, \text{ as } r \rightarrow \infty. \quad (2.29)$$

Hence, the extreme dependence cases are as follows, with the restriction that  $\mathcal{L}(r) \rightarrow 0$ , as  $r \rightarrow \infty$ :

- $\eta = 1/2$  and  $\mathcal{L}(r) = 1$ : exact independence ( $\chi = 0$ ).
- $1/2 < \eta < 1$ : asymptotic independence ( $\chi = 0$ ) and the variables are positively associated.
- $0 < \eta < 1/2$ : asymptotic independence ( $\chi = 0$ ) and the variables are negatively associated.
- $\eta = 1$  and  $\mathcal{L}(r) = 1$ : complete asymptotic dependence ( $\chi = 1$ ).
- $\eta = 1$  and  $\mathcal{L}(r) \rightarrow 0$ , as  $r \rightarrow \infty$ : asymptotic dependence ( $0 < \chi \leq 1$ ).

Consider a tail independence case where  $\eta \in (1/2, 1)$ , and suppose it is assumed that the parametric family to model the joint distribution only embraces the exact tail independence and tail dependence cases. Then, such a model will tend to underestimate the joint survival distribution over a large

$r$ . The introduction of the parameter  $\eta$  has the purpose to distinguish the strength of the independence.

Ledford and Tawn model the type of extremal dependence with the parameter  $\eta$  and the strength of this dependence, with  $\mathcal{L}(r)$ . Ledford and Tawn then give different parametric models for each case of  $\eta$ , i.e. for each type of dependence. In order to estimate  $\eta$ , they use the fact that for  $Z = \min(X, Y)$ :

$$P(Z > r) = P(X > r, Y > r) \sim \mathcal{L}(r)r^{-1/\eta}, \text{ as } r \rightarrow \infty,$$

i.e., they transform a bivariate problem into a univariate problem. Since  $P(Z > r)$  is a survival function, it can be approximated with a  $\text{GPD}(k^*, \sigma^*)$ . So,  $\eta = k^*$ , which makes clear that the restriction on the estimation must be set:  $k^* \in [1/2, 1]$ , for the positive associated variables. Ledford and Tawn give some tests for dependence using likelihood ratios. Once this is done, a parametric family will be fixed for the bivariate model (for example, the bivariate normal distribution when  $1/2 < \eta < 1$ ).

The parameter  $\eta$  suggests tail dependence when it is near 1. For the rest of the range, it gives a measure of the tail independence. For the same purpose, Coles *et al.* (1999) defined another measure of asymptotic dependence, denoted as  $\bar{\chi}$ .

Let  $X$  and  $Y$  be r.v. with Fréchet marginals. Then, define the following measure:

$$\bar{\chi}(r) = \frac{2 \log P(X > r)}{\log P(X > r, Y > r)} - 1,$$

where  $-1 \leq \bar{\chi}(r) \leq 1$ . Define as well:

$$\bar{\chi} = \lim_{r \rightarrow \infty} \bar{\chi}(r)$$

Then,  $\bar{\chi}$  is another measure of the strength of the tail independence, for which  $-1 \leq \bar{\chi} \leq 1$ , and:

- If  $\bar{\chi} = 1$ :  $X$  and  $Y$  are asymptotically (tail) dependent.
- If  $\bar{\chi} = 0$ :  $X$  and  $Y$  are exact independent.
- If  $\bar{\chi} \in (0, 1)$ :  $X$  and  $Y$  are asymptotically (tail) independent and the variables are positively associated.
- If  $\bar{\chi} \in [-1, 0)$ :  $X$  and  $Y$  are asymptotically (tail) independent and the variables are negatively associated.

Coles *et al.* (1999) give priority on the estimation of  $\bar{\chi}$ : only when  $\bar{\chi}$  is estimated significantly near 1, then it is worth to estimate  $\chi$ . Therefore, the test of tail independence against tail dependence is through a test of  $\bar{\chi} < 1$  against  $\bar{\chi} = 1$ , rather than  $\chi = 0$  against  $\chi > 0$ .

The estimation of  $\bar{\chi}$  is similar to  $\eta$ , since the following relation holds:

$$\bar{\chi} = 2\eta - 1.$$

Both  $\chi$  and  $\bar{\chi}$  determine all the possible asymptotic dependence types, as we summarize in the following remark.

**Remark 2.1.** *Asymptotic Dependence Types*

*X and Y are:*

- Complete tail dependent if  $\chi = 1$ .
- Tail dependent if  $0 < \chi < 1$  (and  $\bar{\chi} = 1$ ).
- Exact independent if  $\chi = 0$  and  $\bar{\chi} = 0$ .
- Tail independent if  $\chi = 0$  and  $-1 \leq \bar{\chi} < 1$ .

An interesting framework is that of Ramos and Ledford (2009). They mix the classic multivariate model with  $\eta$ , but with the parameter included in the model. Let  $X$  and  $Y$  have Fréchet marginals. Define  $(S, T) = (X/u, Y/u)$  for  $X$  and  $Y$  larger than a large  $u$ . Ramos and Ledford propose the following model:

$$P(S > s, T > t) = \nu^{-1} \left[ (\rho s)^{-1/\eta} + \left(\frac{t}{\rho}\right)^{-1/\eta} - \left( (\rho s)^{-1/\alpha} + \left(\frac{t}{\rho}\right)^{-1/\alpha} \right)^{\alpha/\eta} \right],$$

where  $(s, t) \in [1, \infty)^2$ ,  $\nu > 0$ ,  $\eta \in (0, 1]$ ,  $\alpha > 0$ , and  $\rho > 0$ .

The joint model recognizes the presence of  $\eta$ . Then, inference is carried out using MLE.

An applied work was developed by Poon *et al.* (2004). Here we present the main features of Poon *et al.* (2004). Consider two random variables,  $X$  and  $Y$ , with distribution functions  $F_X$  and  $F_Y$ . Apply the following transformations:

$$S = -\frac{1}{\log(F_X(X))}; \quad T = -\frac{1}{\log(F_Y(Y))}; \quad (2.30)$$

so that both  $S$  and  $T$  have marginal Fréchet distributions. Working with  $S$  and  $T$  is theoretically equivalent to working with  $X$  and  $Y$  for the joint survival,

since:

$$\begin{aligned} P(X > x, Y > y) &= P\left(S > -\frac{1}{\log(F_X(x))}, T > -\frac{1}{\log(F_Y(y))}\right) \\ &= \bar{F}_{S,T}(s, t), \end{aligned} \quad (2.31)$$

where  $s = -\frac{1}{\log(F_X(x))}$ , and  $t = -\frac{1}{\log(F_Y(y))}$ .

As in the case where the dependence parameter is  $\eta$ , the choice of the parametric model of  $F_{S,T}$  will depend on the estimated values of  $\chi$  and  $\bar{\chi}$ . In fact, only when  $\bar{\chi}$  is near 1, is the tail dependence  $\chi$  estimated.

The model that results from this framework is a quasi copula one.<sup>3</sup> For more details of copula based models, see Heffernan (2000) and Joe (1997). A problem with the copula model is that although in theory the transformations do not change the dependence structure, in practice the estimation of the marginals (which is generally done via empirical distributions) leads to inaccuracies. As well, there is no clear evidence on how much any inexactitude will affect the accuracy of the joint model.

A clear and concise discussion that considers the disadvantages of the copula model is given in Mikosch (2005). Among other issues, Mikosch discusses the problems of fixing the marginals to a standard distribution. This standard procedure in the extreme values framework is a well known difficulty when making inference. Mikosch addresses that, although fixing the marginals is relevant from a probabilistic point of view, when dealing with data, it becomes

---

<sup>3</sup>The main difference with the copula models is that the  $S$  and  $T$  transformations do not have a support on  $[0, 1]^2$ .

hard work, since the marginal distributions are unknown. Mikosch addresses as well, that in practice, the split of the marginals and the dependence structure when using copulas, implies a great uncertainty for just one sample. In brief, it is unclear how sensitive the copula estimations are to the fixing of the marginals.

Finally, we need to discuss the flexibility of parametric multivariate models, regarding their dependence structure. So far, the determination of the asymptotic dependence among variables has not been a clear procedure. When selecting a parametric model family either in the maximum or the POT methods, the asymptotic dependence is determined for all the subsets of variables. Some family of models have improved the flexibility of modeling extremes. For example, the introduction of the asymmetric logistic model by Tawn (1988) solved the problem of non-exchangeable variables. However, the model asymptotic dependence structure is determined only through one single parameter. In the bivariate case, this does not seem to be too restrictive, and some models, as that in Ramos and Ledford (2009), can even model different tail independence behaviors through an additional parameter. However, although most of the bivariate cases have straightforward generalizations, these turn out to be restrictive. For example, it is a hard task to model, with a single parametric model, different asymptotic dependence for different combinations of variables. The MEVD's have the property that pairwise asymptotic exact independence implies  $d$ -dimensional asymptotic exact independence. Therefore, pairwise

analysis can help to select a given parametric family of models. However, the help of the pairwise analysis is restricted, since the multivariate asymptotic dependence is uniquely determined by the pairwise dependence.

In order to overcome the lack of flexibility, some attempts have been given by Joe (1990) and Tawn (1990). We focus on the latter work. Tawn (1990) introduced the multivariate asymmetric model, following the works of Dagsvik (1988) and McFadden (1978). This model was planned to model different types of dependencies among the variables. The way Tawn constructed the individual models is a hierarchical model similar to the individual models of our multivariate model. Tawn (1990) modeled different combinations of subsets of the  $d$ -dimensional problem, as our model will do. However, we find three differences with our work. The first one is the definition of the latent variables in the hierarchical model. In our case, each variable of the multivariate model is explicitly defined by its subvariables. The second difference is that the model in Tawn (1990), although includes all the possible dependencies among the variables, it is not defined as a mixture model. Finally, Tawn (1990) worked with the maximum of *GPD* variables. The model leads to an overparametrization in most of the applications. Therefore, inference has to be combined with the understanding of the application, so that the model can be simplified. Even in the simplified version, the form of the density was not easy to handle. In fact, in the application, Tawn (1990) implemented a model selection framework, over a set of simplified versions of the multivariate

asymmetric model. The models were tested against each other using likelihood ratio tests.

We believe the Bayesian mixture models are a helpful framework to overcome the difficulties. The aim of the thesis is to overcome the difficulty of the lack of flexibility of the classical models regarding the asymptotic dependence. The way we construct the individual models and the closure of the high dimension distribution to marginalization, coupled with the Bayesian paradigm, the MCMC and the reversible jump methodologies, ease the estimation of both the marginal parameters and a wide tail dependence structure.

#### 2.2.4 Summary

In this chapter, we have presented the ideas on the EVT that we believe are sufficient to have a clear idea of the difficulties and challenges that this theory implies.

We have seen that there exist two approaches to study extremes: the Maximum and the POT. Both methodologies give similar results, theoretically. However, we saw that the maximum approach is vulnerable to dismiss some of the information on the extremes. Hence, we have adopted the POT as our general approach.

Regarding the inference, the univariate framework has been well developed during the last decades. However, it is on the multivariate framework where there is a lot to improve on. The main reason relies on the fact that there is not

---

a unique representation of the limit distribution for the vector maximum. This fact has the consequence of not having a direct way to set a tail dependence structure. We will present a multivariate framework based on a mixture model that will have individual models with an easy parametric representation, where the tail dependencies among any subset of the variables can be calculated, just via the parameters of our model.

There is a vast literature dealing on the bivariate case. Most of the literature has the copulas as the main approach to study the bivariate framework. As we have established, theoretically, the transformation of the variables in order to get standard marginal distributions does not change the theoretical dependence structure. However, it does affect when making inference. We present a way to construct parametric families, such that the parameters of a joint distribution determine directly the dependence structure.

The natural inference approach of dealing with mixture models is the Bayesian paradigm, which is what we treat in the next chapter.

# Chapter 3

## Bayesian Theory

In this chapter we will present a review of the Bayesian framework. Since it is a big area we will only focus on the ideas we will use in this work.

Regarding the theoretic foundations, some authors have attempted to explain the Bayesian paradigm by providing an axiomatic system for the rational behavior involved when dealing with statistic problems, e.g. refer to Bernardo and Smith (1995), Lindley (1985), and Robert (1994). This system places the statistician in a consumer environment of the type found in Samuelson (1948), Chapter V. In general terms, the system can be summarized in the following axioms (see, for example Bernardo and Smith (1995)):

- Comparability: the statistician can set preferences over consequences.
- Transitivity of preferences: there exists a rational choice of preferences.
- Consistency of preferences: additional information does not affect the

preferences.

- Precise quantification of uncertainty of events via probability functions

With this framework, the authors provide a rational decision making system, which they claim leads to the Bayesian paradigm.

On the other hand, functionally Bayesian statistics works through the construction of prior distributions on parameters of statistical models. These priors are updated to the posterior distribution once data have been observed. Hence, the posterior distribution constitutes the main function of Bayesian inference.

The outline of this chapter is as follows: In Section 3.1 we present the basic definitions and the general ideas of Bayesian inference. In Section 3.2 we present simulation and Markov Chain Monte Carlo (MCMC) methods. In Section 3.3 we present a particular tool we will use when simulating: the Reversible Jump MCMC. In Section 3.4 we make a review of the Bayesian statistics applied to the EVT, and lastly, in Section 3.5, we present our conclusions.

### 3.1 The Bayesian Idea

Suppose a parametric family is assumed as a model for data. In terms of probability functions, for the Bayesian approach, a joint model of the observations and the parameters is constructed.

A general representation of the joint model of both the data and the parameters is found in Schervish (1995), and is as follows:

Suppose a r.v.  $X$ , given the parameter  $\Theta$ , has a continuous distribution function  $f_{X|\Theta}(x|\theta)$ . Define the distribution of the parameter as  $f_{\Theta}(\theta)$ . Then, the joint distribution has the form:

$$P((X, \Theta) \in B) = \iint \mathbf{1}[(X, \Theta) \in B] f_{X|\Theta}(x|\theta) f_{\Theta}(\theta) dx d\theta. \quad (3.1)$$

From this representation, we can identify the main functions that are involved.

### 3.1.1 Prior Distribution

The distribution of the parameter  $f_{\Theta}(\theta)$  is defined as the prior distribution. This function describes the uncertainty of the parameters prior to any information provided by the data. Thus, in principle, even with no sampling, some information can be established for the parameters, such as the inclusion of information coming from personal and/or experts opinions. This information would be included exclusively in the parameters of  $f_{\Theta}$ , which are called hyperparameters.

For example, suppose the r.v.  $X$  is such that  $X|\theta \sim N(\theta, 1)$ . Assume that it is believed that  $\theta$  is positive, so that an exponential prior can be set, which means:

$$f_{\Theta}(\theta) = \gamma e^{-\gamma\theta},$$

for  $\theta > 0$ , and where  $\gamma > 0$  is the hyperparameter. Hence, the beliefs are

included exclusively in  $f_{\Theta}(\theta)$ .

Another option is the choice of assuming no information is given prior to data. Such a prior is called non-informative and often is improper. This is a widely used approach in practice. For more on non-informative priors, we refer the reader to Bernardo and Smith (1995), Box and Tiao (1973), and Gelman *et al.* (2003).

### 3.1.2 Posterior Distribution

In (3.1), the prior distribution is complemented by the conditional model  $f(x|\theta)$ .<sup>1</sup> Both functions form the joint model for the pair  $(X, \Theta)$ , since  $f(x, \theta) = f(x|\theta)f(\theta)$ . Once we have observed the data  $x$ , we set  $X = x$  and the function given the sample is called the posterior distribution, and is as follows:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta}. \quad (3.2)$$

This expression is known as the Bayes' rule and constitutes the updating rule of the Bayesian framework.

The denominator is not a function of  $\theta$ . Therefore, the posterior distribution is determined by  $f(x|\theta)f(\theta)$  up to a constant of proportionality. Thus, it is common to express it via:

$$f(\theta|x) \propto f(x|\theta)f(\theta). \quad (3.3)$$

---

<sup>1</sup>For the sake of explanation, we will drop the subindices on the distributions.

The posterior distribution is the function from where inference about the parameters can be made. For example, an estimator  $\hat{\theta}$  is commonly selected as the posterior mean or median.

Suppose  $\theta$  is a single parameter. An interval estimator of level  $p$  for  $\theta$  in the posterior distribution is called a credibility interval and is of the form  $(\hat{\theta}_1, \hat{\theta}_2)$ , where:

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2 | x) = p.$$

In many cases, the posterior is intractable for analytic results, such as the computation of the moments (provided they exist) or even the probability of a set. Hence, it is necessary to rely on sampling methods, such as the MCMC methods, which will be described later.

In general terms, the sampling overcomes the intractability of the posterior via the approximation of it by the empirical distribution of a sample. This means that, if  $\theta$  is the random variable of interest and we sample  $\theta_1, \dots, \theta_N$  from the posterior  $f(\theta | x)$  and if the interest relies on the computation of the mean of  $\theta$ , then:

$$E(\theta | x) \approx \frac{1}{N} \sum_{i=1}^N \theta_i.$$

More generally,

$$\int g(\theta) f(\theta | x) d\theta \approx \frac{1}{N} \sum_{i=1}^N g(\theta_i).$$

In the next section, we will present some of the most important simulation methods found in the Bayesian literature.

## 3.2 Markov Chain Monte Carlo Methods

For a detailed explanation and applications of MCMC methodology, we refer the reader to Casella and Robert (1999), Gelman *et al.* (2003), Robert (1994), Smith and Roberts (1993), and Tierney (1994).

Suppose we want to draw a r.v.  $Z$  with distribution function  $F$  and density  $f$ , which in this case is the posterior distribution and we define it as the target function.

The simplest method to sample from is called direct sampling. Suppose  $F$  has an inverse  $F^{\leftarrow}$ , then a single draw is given by:

$$z = F^{\leftarrow}(u)$$

where  $u$  is a standard uniform r.v. Thus, the sample, say  $z_1, \dots, z_n$ , consists of i. i. d. draws.

However, in practice, the form of  $F^{\leftarrow}$  will rarely be analytically tractable and its form is generally unknown. Specially, in the case when  $Z$  is a vector.

### 3.2.1 Rejection Sampling

The Rejection Sampling method works when it is not possible to sample from the target function  $f$ , but there exists a function  $g(z)$  such that:

- It is possible to draw samples from it.
- There exists a constant  $M > 0$ , such that  $f(z)/g(z) \leq M$  for all  $z$ .

The  $g$  function need not integrate to 1, but because of the second condition, it needs to have a finite integral. The  $g$  function gives a reference for the behavior of  $f$ . The algorithm goes as follows:

1. Generate a proposal  $z$  from  $g$  and  $u$  from a unit uniform distribution.
2. If  $f(z)/(Mg(z)) \leq u$ , accept  $z$  as coming from  $f$ , otherwise, return to step 1.

As in the direct sampling method, the sample  $z_1, \dots, z_n$  consists of i. i. d. draws.

Unfortunately, when  $f$  is not tractable, it is difficult to find a suitable  $g$  function. Therefore, more sophisticated methods have to be found, such as MCMC.

When the direct or the rejection sampling are not useful, it is common to drop the possibility of drawing i. i. d. samples. The way to proceed then is to set up a Markov Chain  $Z_n$ , for which  $f$  is the stationarity density, which means:

$$Z_n \xrightarrow{d} Z \sim f.$$

Then, from the ergodic theorem, or more specifically, the Birkhoff theorem—refer to Birkhoff (1942) and Wiener (1939):

$$\frac{1}{N} \sum_{n=1}^N h(Z_n) \rightarrow \int h(Z)f(Z)dZ \text{ as } N \rightarrow \infty,$$

with probability one, for a real-valued measurable function  $h$ . Thus, the mean of the Markov Chain  $Z_n$  is, in the limit, the same as the process of the direct or rejecting sampling. However, the variance from the Markov Chain is larger.

Hence, the natural question is how to set up a Markov Chain for which  $f$  is the stationarity density.

### 3.2.2 Gibbs Sampler

The first Markov Chain (MC) sampling we are describing is the Gibbs sampler—refer to Smith and Roberts (1993), Casella and Robert (1999) and Gelman *et al.* (2003).

Suppose the variable  $Z$  is a vector that can be divided into  $d$  subvectors, so that  $Z = (Z_1, \dots, Z_d)$ . The aim is to draw a MC which, for notational convenience, we now represent as  $Z^{(1)}, \dots, Z^{(N)}$ .

Define  $Z_{-j}$  as the vector with all the elements in  $Z$ , except those of  $Z_j$ , i.e.:

$$Z_{-j} = (Z_1, \dots, Z_{j-1}, Z_{j+1}, \dots, Z_d). \quad (3.4)$$

Then, the Gibbs sampler works when the conditional distributions  $f(Z_j | Z_{-j})$  are known and it is possible to draw, directly or by rejection, samples from them, for  $j = 1, \dots, d$ . Hence, the algorithm proceeds as follows at iteration  $t$ :

1. Sample  $Z_1^{(t)}$  from  $f(Z_1 | Z_{-1}^{(t-1)})$ .
2. Similarly, sample  $Z_j^{(t)}$  from  $f(Z_j | Z_1^{(t)}, \dots, Z_{j-1}^{(t)}, Z_{j+1}^{(t-1)}, \dots, Z_d^{(t-1)})$ .
3. Proceed until  $t = N$ .

However, frequently, it is not possible to sample from the conditional posterior distributions. Therefore, it necessary to modify this method for those

cases.

### 3.2.3 MCMC Metropolis-Hastings

The MCMC methods are used to deal with a target distribution that is not possible to sample, directly or by rejection. They constitute a clever and versatile modification of the rejection sampling. In this case, the  $g$  function is substituted by a transition or proposal distribution, which depends only on the previous draw, and which we define as  $T(z_t | z_{t-1})$ . When the transition distribution is not symmetric, i.e.  $T(z_t | z_{t-1}) \neq T(z_{t-1} | z_t)$ , the method is called Metropolis-Hastings.

As the  $g$  function in rejection sampling,  $T$  has to be a function that is possible to sample from. However, the second condition of rejection sampling is substituted by a rejection criteria that involves the previous state.

Let  $f$  be the target density of  $Z$ . We define the algorithm at iteration  $t$  as follows:

1. Draw  $z^*$  from  $T(Z | z_{t-1})$ .
2. Define  $r$ , the acceptance ratio as:

$$r = \frac{f(z^*)}{f(z_{t-1})} \div \frac{T(z^* | z_{t-1})}{T(z_{t-1} | z^*)}, \quad (3.5)$$

and, consequently, define:

$$z_t = \begin{cases} z^* & \text{with probability } \min(r, 1) \\ z_{t-1} & \text{otherwise.} \end{cases} \quad (3.6)$$

Then, the sample  $(Z_1, \dots, Z_N)$  is a M.C. that has  $f$  as its stationarity density. For the proof of this convergence, please refer to Gelman *et al.* (2003) and Tierney (1994).

The speed of the convergence will depend on each case. For models where information of the behavior near the mode of the distribution is required, the convergence speed will be quicker than cases where the very high or low quantiles are required. Since the starting point is hardly a representative value, some of its following values will share the same feature. This fact leads to the burn-in concept. This is the amount of values that are dropped from the sampling. For example, in some our applications, we drop the first 1,000 observations, out of 10,000.

### 3.2.4 Metropolis-Hastings within Gibbs sampler

We used the Metropolis-Hastings (M-H) within Gibbs Sampler method, which is a slight modification of the Gibbs sampler. Instead of drawing via direct or rejection sampling, a MCMC M-H is used.

Define  $Z = (Z_1, \dots, Z_d)$ , in a convenient order, and  $Z_{-j}$  as in (3.4). Then, the algorithm proceeds as follows:

1. Pick starting values  $(Z_1^{(0)}, \dots, Z_d^{(0)})$ .
2. At iteration  $t$ , sample  $Z_j^{(t)}$  via MCMC M-H from

$$f(Z_j | Z_1^{(t)}, \dots, Z_{j-1}^{(t)}, Z_{j+1}^{(t-1)}, \dots, Z_d^{(t-1)}),$$

for  $j = 1, \dots, d$ .

3. Proceed until  $t = N$ .

### 3.3 Reversible Jump MCMC

In many Bayesian problems involving model uncertainty, MCMC algorithms are required to “jump” from one model to another. In this case, the most common procedure is known as the Reversible Jump MCMC of Green (1995).

Suppose we need to select a parametric model for a r.v.  $X$ , out of  $K$  models:  $M_1, \dots, M_K$ , each one having a parameter vector  $\theta_k$  of dimension  $d_k$ , for  $k = 1, \dots, K$ . Define the individual models as  $f(x | \theta_k, M_k)$ , and for Model  $k$  define the posterior as  $f(\theta_k | M_k)$ , and the prior of  $\theta_k$ , as  $\pi_k$ .

The general idea of Reversible Jump MCMC is to sample from a joint target distribution for the pair  $(k, \theta_k)$ , say  $f(k, \theta_k)$ . Define  $y = (k, \theta_k)$ , so that the target distribution is  $f(y)$ . The jump at step  $t$  will only depend on the  $(t - 1)$ th step. Thus, a Markov chain is constructed, whose limit distribution must match with the target distribution. This limit is achieved if the Markov Chain follows the detailed balance requirement, which we will explain. Once the joint sample is achieved, the selection of a model will be based on the empirical frequencies of  $k$ , so that, for example, the selected model should coincide with the empirical mode of  $k$ . We now describe in some extension the algorithm:

Select an initial model  $M_0$  with initial parameter vector  $\theta_0$ . The first step is to propose a jump to a model  $M_{k^*}$ . In a general notation, denote the probability of jumping from model  $k$  to  $k^*$  as  $J_{k,k^*}$ . If  $M_{k^*} = M_k$ , then the next step is a single MCMC simulation, thus a new  $\theta_k$  must be sampled. Otherwise, it is necessary to compare the models in order to accept or reject  $M_{k^*}$ . To do so, we need to address a couple of issues.

The first one is that the jump from  $M_k$  to  $M_{k^*}$  involves as well a jump from  $\theta_k$  to  $\theta_{k^*}$ , which might imply a change of dimension. Therefore,  $f(z|\theta_k, M_k)$  and  $f(z|\theta_{k^*}, M_{k^*})$  are not comparable yet. In order to deal with this change of dimension, an auxiliary r.v.  $u$ , with distribution  $J(u|k, k^*, \theta_k)$ , must be drawn. This variable will help to fill the gap between the dimensions  $d_k$  and  $d_{k^*}$ . The way to do so is by defining a deterministic function of  $u$ , say  $g_{k,k^*}$ , such that  $(\theta_{k^*}, u^*) = g_{k,k^*}(\theta_k, u)$ . Thus,  $d_k + \dim(u) = d_{k^*} + \dim(u^*)$ . Finally, the last factor to consider is the Jacobian of the transformation  $g_{k,k^*}$ .

Suppose  $y \in A$  and  $y^* \in B$ . Then, it is necessary that the probability of accepting the move from  $y$  to  $y^*$ , say  $\alpha(y, y^*)$ , must hold the detailed balance requirement. This is expressed as:

$$\int_R f(y) \beta_{k,k^*}(u|\theta_k) \alpha(y, y^*) dy du = \int_R f(y^*) \beta_{k^*,k}(u^*|\theta_{k^*}) \alpha(y^*, y) dy^* du^*;$$

where  $R = (y, y^*) \in A \times B$ , and  $\beta_{k,k^*}(u|\theta_k) = J_{k,k^*} J(u|k, k^*, \theta_k)$ , and similarly for  $\beta_{k^*,k}(u^*|\theta_{k^*})$ .

Assume  $g_{k,k^*}$  is differentiable, then a valid choice for  $\alpha$  is:

$$\alpha(y, y^*) = \min(1, r).$$

where  $r$  is the acceptance ratio of model  $k$ , and is expressed as:

$$r = \frac{f(z^* | \theta_{k^*}, M_{k^*})p(\theta_{k^*} | M_{k^*})\pi_{k^*}}{f(z | \theta_k, M_k)p(\theta_k | M_k)\pi_k} \times \frac{J_{k^*,k}J(u^* | k^*, k, \theta_{k^*})}{J_{k,k^*}J(u | k, k^*, \theta_k)} \times \left| \frac{\nabla g_{k,k^*}(\theta_k, u)}{\nabla(\theta_k, u)} \right| \quad (3.7)$$

Therefore, the rule equivalent to (3.6) would be:

$$(k+1, \theta_{k+1}) = \begin{cases} (k^*, \theta_{k^*}) & \text{with probability } \min(r, 1) \\ (k, \theta_k) & \text{otherwise.} \end{cases}$$

### 3.4 Bayesian Methods in EVT

In this section we will expose a brief summary of the literature that combines the Bayesian framework and the EVT. So far, the Bayesian framework has not been a great influence on the EVT. The few works there are, are applied to a univariate environment. The reason for this is that the multivariate EVT is still in an early development stage.

However, the path to the Bayesian framework is clear with the current multivariate theory: for example, given the representation in (2.15), a parametric model must be selected for  $H$ , so that prior distributions can be defined to the parameters of  $H$ . Therefore, a posterior distribution can be assigned to the MEVD. The same would work for the parameters of any copula, such as

those used in Poon *et al.* (2004).

A good review of the Bayesian framework in the EVT is given in Coles and Powell (1996). Coles and Powell affirm that the Bayesian applications were almost limited to give some priors for specific cases of the GEV. Pickands (1997) proposes a prior distribution for the  $GPD_{k,\beta}$  as:

$$f(k, \beta) \propto \frac{1}{\beta} \mathbf{1}_{[\beta > 0]}.$$

Although this prior has the drawback that it depends on the threshold  $u$  (recall that the scale parameter of the GPD in the POT depends on the choice of the threshold, see (2.8)), it constitutes the first attempt to work with a Bayesian framework with the GPD.

Some other works have gone beyond only specification of prior distributions. One of them is Coles and Tawn (2005): they give a complete Bayesian approach, using the point process of Smith (1989) applied to wave surges. Coles and Tawn compare it with classical estimators and deal with model issues such as seasonality.

In previous work, Coles and Tawn (1990) defined a spatial model for annual maximum sea levels. Ever since, environmental extreme statistics have relied widely on these models. One of these works is Cooley *et al.* (2007) where a Bayesian hierarchical model is applied, using the high threshold approach.

In Bottolo *et al.* (2003), a complete Bayesian model was introduced to model large insurance claims. Bottolo *et al.* (2003) used a model where a Reversible Jump MCMC was needed. This model adjusts as well to difficulties

such as seasonality.

Recently, Ferraz *et al.* (2011) proposed a combination of a mixture of gamma distributions for the data under a high threshold and a GPD to model the tails.

Guillote and Perron (2008) define a way to estimate the dependence function of Pickands' representation in (2.21) via non-parametric methods and a Bayesian approach for the marginals.

One more Bayesian topic to address is the work done on elicitation. In Coles and Tawn (1996), an application to rainfall data is introduced. This paper is innovative because it incorporates experts knowledge to the GEVD through the quantiles of the model. If we invert the GEVD, the  $(1-p)$  quantile is given by:

$$q_p = \frac{\beta}{k} [(-\log(1-p))^{-k} - 1].$$

Coles and Tawn selected three quantiles, say  $q_1 < q_2 < q_3$ , and in order to avoid dependence among them, they worked with the differences  $\tilde{q}_1 = q_1 - l$ ,  $\tilde{q}_2 = q_2 - q_1$ , and  $\tilde{q}_3 = q_3 - q_2$ . Therefore, it was assumed that the  $\tilde{q}_i$  are gamma distributed. It was possible to ask the experts two estimates (since the gamma has two parameters) of each  $\tilde{q}_i$ , in order to determine the hyperparameters of the gamma distributions. This elicitation resulted in an improvement of the estimates: the Bayesian interval constructed was half of the width of the corresponding classical confidence interval.

There are two other Bayesian works that we will discuss in Section 6.4,

which are Coles and Pauli (2002) and Apputhurai and Stephenson (2011), once we have introduced our multivariate model.

### 3.5 Conclusions

We presented in this chapter, the main ideas involved in the Bayesian framework. All of them will help us to build a framework to model our extreme data. Besides, we introduced the main sampling algorithms that will allow us to do the inference in our model.

So far, the EVT is practically a desert area in terms of Bayesian applications. This means a huge amount of exploration is expected in the forthcoming years. We expect that, since the essence of the extremes makes them scarce, so that any additional information to data can make a considerable difference when making inference. Besides, as we will show later, the Bayesian estimations avoid some difficulties that arise when working with MLE, in the classical framework.

# Chapter 4

## Interim Bivariate Models

The purpose of this chapter is to show the development of the framework for building the form of the final individual model of the mixture model. The best way to construct a multivariate model is to start with the simplest version: the bivariate model. Once we find the final model, we will generalize it to the trivariate or the general multivariate case. This final bivariate individual model will be given in Chapter 5.

We seek a model for which the following properties hold:

1. – Realistic marginal tail models.
2. – Simplicity of the joint and marginal distributions. (4.1)
3. – Embrace different types of asymptotic dependence.

We begin this chapter with some first attempts, which get rejected because of at least one of these properties. Then, we modify the features until we obtain the final model. The final model will be presented in the next chapter. For a

specific model we construct its joint distribution and we derive the tail dependence structure the way we mentioned in Section 2.2: via the standardization of the marginals. We work with Fréchet marginals, so that the corresponding transformations are as follows (recall (2.30)):

$$S = -\frac{1}{\log(F_X(X))}; \quad T = -\frac{1}{\log(F_Y(Y))}.$$

Thus, the joint distribution is expressed as follows (recall (2.31)):

$$\begin{aligned} P(X > x, Y > y) &= P\left(S > -\frac{1}{\log(F_X(x))}, T > -\frac{1}{\log(F_Y(y))}\right) \\ &= P(S > s, T > t). \end{aligned} \quad (4.2)$$

We believe the tail dependence parameter  $\chi$  is a proper measure of the tail dependence between variables. Recall from (2.25) that it can be expressed as:

$$\chi = \lim_{s \rightarrow \infty} P(T > s | S > s) = \lim_{s \rightarrow \infty} \frac{P(T > s, S > s)}{P(S > s)},$$

and for any marginal  $F_X$  (or similarly for  $F_Y$ ):

$$P(S > s) = P(X > F_X^{\leftarrow}(e^{-1/s})) = 1 - e^{-1/s} \sim s^{-1},$$

as  $s \rightarrow \infty$ .

We need to recall as well the definition of the tail coefficient from (2.28):

$$P(S > s, T > s) \sim \mathcal{L}(s)s^{-1/\eta}, \quad \text{as } s \rightarrow \infty, \quad (4.3)$$

where  $0 < \eta \leq 1$  and  $\mathcal{L}$  is a slow varying function (recall Appendix A for the definition and some examples). Therefore (and recalling (2.29)), we get that

$$P(S > s | T > s) \sim \mathcal{L}(s)s^{1-1/\eta}, \quad \text{as } s \rightarrow \infty.$$

The coefficient  $\eta$  will help us to identify the asymptotic form of the model.

In order to compute (4.2) we:

1. Compute the marginal distributions.
2. Compute the marginal inverse distributions or its asymptotic form.
3. Compute the joint survival function for the transformed variables  $S$  and  $T$ .
4. Look at the asymptotic form of  $P(S > s | T > s)$ .

We present some general bivariate models, which we build through the Common Variable Method or the Trivariate Reduction Method. Let  $X$  and  $Y$  be positive random variables to “join together”. We seek to split the marginal and the joint effects using different random variables, which we will onwards call (random) subvariables. Thus, those subvariables which are present in both  $X$  and  $Y$  will deal just with the joint structure and the subvariables exclusive to  $X$  or  $Y$ , will deal with the marginal behavior. This method has been applied by Ahmed (1961) and Walhin and Paris (2000) in convolution of Poisson variables, whereas Sarabia and Gomez (2008) present a review of the literature in this method for both discrete and continuous variables, with different types of associations among subvariables.

The present chapter has: Section 4.1 which presents the convolution models, and Section 4.2 explains the first product model.

## 4.1 Convolution Models

Our starting point is to represent each variable as the sum of two subvariables. This representation leads to the well known convolution distributions. The models have the following general representation:

$$\begin{aligned}X &= A + B, \\Y &= A + C,\end{aligned}\tag{4.4}$$

with all  $A$ ,  $B$ , and  $C$  positive and mutually independent.

This type of model is a simple way of creating dependence between variables. The magnitude of the subvariable  $A$  will settle the dependence structure: the larger the magnitude of  $A$ , the more related  $X$  and  $Y$  will be. As a consequence, when the second moment exists for each subvariable, we have that

$$\begin{aligned}\text{Cov}(X, Y) &= \text{Cov}(A + B, A + C) \\&= \text{Cov}(A, A) + \text{Cov}(A, C) + \text{Cov}(B, A) + \text{Cov}(B, C) \\&= \text{Cov}(A, A) = \text{Var}(A).\end{aligned}$$

Thus, the dependence between variables is determined only by the common subvariable, through its variance.

We start by looking at exponential models for the subvariables. Although we know these models do not have the first property in (4.1), since the assumption of having exponential marginals is very restrictive, we consider this case to get into the computations involving on the convolution framework.

### 4.1.1 Convolution Model 1. Unit Exponential Models

As a start, we consider the model:

$$\begin{aligned} X &= A + B, \\ Y &= A + C, \end{aligned} \tag{4.5}$$

with independent  $B, C \sim \text{Exp}(1)$  and  $A \sim \text{Exp}(\alpha)$ .

It is easy to see that as  $\alpha$  decreases, the tail dependence would increase, since  $E(A) = 1/\alpha$ .

As we mentioned previously, the first thing to look at is the marginal distributions. So we have that

$$\begin{aligned} F_X(x) &= P(B < x - A) \\ &= \int_0^x P(B < x - a | a) f_A(a) da \\ &= \int_0^x (1 - e^{-(x-a)}) \alpha e^{-\alpha a} da \\ &= 1 - e^{-\alpha x} - \alpha e^{-x} \int_0^x e^{-(\alpha-1)a} da \\ &= 1 - e^{-\alpha x} - \alpha e^{-x} \left( \frac{1}{\alpha-1} \right) (1 - e^{-(\alpha-1)x}) \\ &= 1 - e^{-\alpha x} - \left( \frac{\alpha}{\alpha-1} \right) e^{-x} + \left( \frac{\alpha}{\alpha-1} \right) e^{-\alpha x} \\ &= 1 + \left( \frac{1}{\alpha-1} \right) e^{-\alpha x} - \left( \frac{\alpha}{\alpha-1} \right) e^{-x}. \end{aligned}$$

Therefore, we can define the marginal as follows:

$$F_X(x) = 1 + \left( \frac{1}{\alpha-1} \right) e^{-\alpha x} - \left( \frac{\alpha}{\alpha-1} \right) e^{-x}. \tag{4.6}$$

For notation purposes, we define the  $G$  function as:

$$G_X(s) = F_X^{\leftarrow} (e^{-1/s}), \quad (4.7)$$

which, in this case, is the same as  $G_Y(s)$ , since both have the same distribution.

Thus, we can define  $G(s) = G_X(s) = G_Y(s)$ . So, we can compute the joint survival function of the transformed variables as follows:

$$\begin{aligned} P(S > s, T > s) &= P(-1/\log(F_X) > s, -1/\log(F_Y) > s) \\ &= P(F_X > \exp(-1/s), F_Y > \exp(-1/s)) \\ &= P(A + B > F_X^{\leftarrow}(\exp(-1/s)), A + C > F_Y^{\leftarrow}(\exp(-1/s))) \\ &= P(B > G_X(s) - A, C > G_Y(s) - A) \\ &= \int P(B > G(s) - a, C > G(s) - a | a) f_A(a) da \\ &= \int_0^{G(s)} e^{-2(G(s)-a)} \alpha e^{-\alpha a} da + \int_{G(s)}^{\infty} P(B > 0, C > 0 | a) f_A(a) da \\ &= \alpha e^{-2G(s)} \int_0^{G(s)} e^{-(\alpha-2)a} da + e^{-\alpha G(s)} \\ &= \alpha e^{-2G(s)} \left( \frac{1}{\alpha-2} \right) (1 - e^{-(\alpha-2)G(s)}) + e^{-\alpha G(s)} \\ &= \frac{\alpha}{\alpha-2} e^{-2G(s)} - \frac{\alpha}{\alpha-2} e^{-\alpha G(s)} + e^{-\alpha G(s)} \\ &= \frac{\alpha}{\alpha-2} e^{-2G(s)} - \frac{2}{\alpha-2} e^{-\alpha G(s)}. \end{aligned}$$

Hence, the joint survival function is represented in two parts:

$$P(S > s, T > s) = \begin{cases} \frac{\alpha}{\alpha-2} e^{-2G(s)} - \frac{2}{\alpha-2} e^{-\alpha G(s)} & \alpha \neq 2 \\ (2G(s) + 1)e^{-2G(s)} & \alpha = 2. \end{cases} \quad (4.8)$$

We present the dependence structure of this model in the following theorem.

**Theorem 4.1.1.** *Tail Dependence in Simple Exponential Model*

Let  $X$  and  $Y$  be defined as in (4.5), then the tail dependence is given by

$$\chi = \begin{cases} \frac{2(1-\alpha)}{2-\alpha}, & 0 < \alpha < 1 \\ 0, & \alpha \geq 1. \end{cases} \quad (4.9)$$

We present the proof in Appendix B.1. In Figure D.2 we plot the shape of the tail dependence given in (4.9). As it can be seen, when  $\alpha$  tends to zero,  $\chi$  tends to one. And when  $\alpha$  gets closer to one, the tail dependence tends to zero. Hence, this type of model only allocates asymptotic dependence when  $\alpha < 1$ , which is the interval where the expected value of  $A$  is larger than the corresponding  $B$  and  $C$ , which means that the subvariable  $A$  is large enough to give a common dependence to both variables in the tails. The next task is to represent the subvariables as general exponential variables. However, we are aware that these models do not have the first property in (4.1), i.e., they do not embrace a wide range of types of tails.

**4.1.2 Convolution Model 2. General Exponential**

We present the general exponential convolution model:

$$\begin{aligned} X &= A + B, \\ Y &= A + C, \end{aligned} \quad (4.10)$$

with independent  $B \sim \text{Exp}(\beta)$ ,  $C \sim \text{Exp}(\gamma)$ , and  $A \sim \text{Exp}(\alpha)$ . For the ease of explanation, and without losing generalization, we will suppose  $\beta > \gamma$ .

As in the simple case, we begin with the calculation of the marginal distribution.

$$\begin{aligned}
 F_X(x) &= P(B < x - A) \\
 &= \int_0^x P(B < x - a | a) f_A(a) da \\
 &= \int_0^x (1 - e^{-\beta(x-a)}) \alpha e^{-\alpha a} da \\
 &= 1 - e^{-\alpha x} - \alpha e^{-\beta x} \left( \frac{1}{\alpha - \beta} \right) (1 - e^{-(\alpha - \beta)x}) \\
 &= 1 - e^{-\alpha x} - \frac{\alpha}{\alpha - \beta} e^{-\beta x} + \frac{\alpha}{\alpha - \beta} e^{-\alpha x} \\
 &= 1 + \frac{\beta}{\alpha - \beta} e^{-\alpha x} - \frac{\alpha}{\alpha - \beta} e^{-\beta x}.
 \end{aligned}$$

Therefore, we can define both marginals as follows:

$$F_X(x) = \begin{cases} 1 + \frac{\beta}{\alpha - \beta} e^{-\alpha x} - \frac{\alpha}{\alpha - \beta} e^{-\beta x} & \alpha \neq \beta; \\ 1 - (\alpha x + 1) e^{-\alpha x} & \alpha = \beta. \end{cases} \quad (4.11)$$

$$F_Y(y) = \begin{cases} 1 + \frac{\gamma}{\alpha - \gamma} e^{-\alpha y} - \frac{\alpha}{\alpha - \gamma} e^{-\gamma y} & \alpha \neq \gamma; \\ 1 - (\alpha y + 1) e^{-\alpha y} & \alpha = \gamma. \end{cases}$$

Thus, we note the marginals are mixtures of exponential distributions.

In the next theorem, we introduce the dependence structure derived by this model.

**Theorem 4.1.2.** *Tail Dependence in Exponential Model*

*Define  $X$  and  $Y$  as in (4.10). Then, the tail dependence between  $X$  and  $Y$*

is given by:

$$\chi = \begin{cases} 1 - \frac{\beta\alpha}{(\gamma-\alpha)(\beta+\gamma-\alpha)} \left(1 - \frac{\alpha}{\gamma}\right)^{\gamma/\alpha} \left(1 - \frac{\alpha}{\beta}\right)^{1-\gamma/\alpha}, & \alpha < \gamma < \beta \\ 0, & \gamma < \beta < \alpha \text{ or } \gamma < \alpha < \beta. \end{cases} \quad (4.12)$$

We present the proof in Appendix B.2. This theorem states that there is tail dependence only when  $\alpha$  is the smallest parameter, which means, only when  $E(A) > \max\{E(B), E(C)\}$ . This makes sense as in the simple exponential case: the larger the magnitude of the subvariable  $A$ , the larger the dependence between the variables.

As we have mentioned previously, the exponential based models have the purpose of getting us to grips with the maths involved. The exponential distribution is a model with tails that are too light, so that it cannot be expected to fit a wide range of extreme data. This means that the exponential based models violate the first property in (4.1). The next step is to modify the distribution of the subvariables, so that heavier tails can be embraced. The GPD is the natural distribution to fit any type of tail, as we saw in Chapter 2. Hence, we propose in the next section a model that convolutes GP distributions.

### 4.1.3 Convolution Model 3. GPD

The GPD is the non-degenerate limit distribution of the tail of any distribution. Therefore, if the convolution model works for any distribution, then it is expected to work with the GPD. The tail characterization of this model will

not depend on the value of the scale parameter of the GPD. Therefore, we will set it as 1 for all the subvariables.

We define the convolution model as in the previous cases:

$$X = A + B,$$

$$Y = A + C,$$

with independent  $A \sim \text{GPD}(k_1, 1)$ ,  $B \sim \text{GPD}(k_2, 1)$ , and  $C \sim \text{GPD}(k_3, 1)$ , for  $k_1 > 0$ ,  $k_2 > 0$  and  $k_3 > 0$ .

Firstly, as in the exponential case, we derive the marginal distributions. So

$$\begin{aligned} F_X(x) &= \int_0^x P(B < x - a | a) f_A(a) da \\ &= \int_0^x [1 - (1 + k_2(x - a))^{-1/k_2}] (1 + k_1 a)^{-1/k_1 - 1} da \\ &= 1 - \int_0^x (1 + k_2(x - a))^{-1/k_2} (1 + k_1 a)^{-1/k_1 - 1} da. \end{aligned}$$

However, this integral does not have an analytical solution. Nevertheless, Barbie and McCormick (2005) introduced a way to compute convolutions of regularly varying distributions. Recall the definition of regular variation in (A.1), then

$$\lim_{t \rightarrow \infty} \frac{\bar{F}_A(tx)}{\bar{F}_A(t)} = \frac{(1 + k_1 tx)^{-1/k_1}}{(1 + k_1 t)^{-1/k_1}} = x^{-1/k_1},$$

for  $k_1 > 0$ . Therefore,  $\bar{F}_A \in RV_{-1/k_1}$ , and similarly  $\bar{F}_B \in RV_{-1/k_2}$ .

Define the convolution operation on two distributions  $F$  and  $G$ , for random variables  $W$  and  $Z$ , respectively, as:

$$F * G(w) = P(W + Z \leq w).$$

The main results of Barbie and McCormick need some requirements on the distributions to convolute. For example, Barbie and McCormick (2005) defined the following concept:

**Definition 4.1.1.** *Asymptotic Smoothness*

*A function  $f$  is asymptotically smooth with index  $-\alpha$  if*

$$\lim_{\delta \rightarrow 0} \limsup_{t \rightarrow \infty} \sup_{0 < |x| < \delta} \left| \frac{f(t(1-x)) - f(t)}{xf(t)} - \alpha \right| = 0.$$

This property, as Barbie and McCormick state, is held by those distributions functions  $F$  for which  $\bar{F}$  is regular varying and has an ultimately monotone density. Hence, the GPD is asymptotically smooth.

Barbie and McCormick defined as well the following:

**Definition 4.1.2.** *Right Tail Dominance*

*A distribution  $F$  is right tail dominant if*

$$\lim_{t \rightarrow \infty} \frac{F(-t\delta)}{F(t)} = 0,$$

for any  $\delta > 0$ , which is clearly held by the GPD.

Denote, as well, the truncated mean of the distribution  $F$  as:

$$\mu_F(t) = \int_0^t x dF(x), \text{ for } t \geq 0.$$

The result of Barbie and McCormick (2005) divides into two cases. The first one is related to light tails.

**Theorem 4.1.3.** *Convolution of Light Tails*

Let  $F$  and  $G$  be two distributions, such that  $\bar{F} \in RV_{-\alpha}$  and  $\bar{G} \in RV_{-\beta}$  with  $\alpha \wedge \beta \geq 1$ . Assume  $F$  and  $G$  are asymptotically smooth and right tail dominant. Then,

$$1 - F * G(t) = \bar{F}(t) + \bar{G}(t) + \frac{1}{t} (\alpha \bar{F}(t) \mu_G(t) + \beta \bar{G}(t) \mu_F(t)) (1 + o(t)),$$

as  $t \rightarrow \infty$ .

In the GPD case, this result applies when  $k_1 \vee k_2 < 1$ , which is the case when both means exist, and are the following:

$$\mu_{F_A}(t) = \frac{1}{1-k_1} \left( 1 - (1 + k_1 t)^{-\frac{1}{k_1} + 1} \right), \text{ and } \mu_{F_B}(t) = \frac{1}{1-k_2} \left( 1 - (1 + k_2 t)^{-\frac{1}{k_2} + 1} \right).$$

Therefore,

$$\begin{aligned} F_A * F_B(t) &\approx 1 - \bar{F}_A(t) - \bar{F}_B(t) + \frac{1}{k_1(1-k_2)} \bar{F}_A(t) \left( t^{-1} - t^{-1}(1 + k_2 t)^{-\frac{1}{k_2} + 1} \right) \\ &\quad + \frac{1}{k_2(1-k_1)} \bar{F}_B(t) \left( t^{-1} - t^{-1}(1 + k_1 t)^{-\frac{1}{k_1} + 1} \right) \\ &\approx 1 - \left( 1 - \frac{1}{1-k_1} \right) \bar{F}_A(t) - \left( 1 - \frac{1}{1-k_2} \right) \bar{F}_B(t), \end{aligned}$$

as  $t \rightarrow \infty$ , for  $k_1, k_2 < 1$ .

Therefore,

$$F_X(t) \approx 1 - \bar{F}_A(t) - \bar{F}_B(t), \text{ as } t \rightarrow \infty. \quad (4.13)$$

However, this means that the tail of the marginal is asymptotically equivalent to the heaviest of the tails of  $A$  or  $B$ . Thus, the joint model can only imply either complete tail dependence or exact tail independence. Therefore, this model does not have the third property in (4.1).

The next theorem of Barbie and McCormick (2005) corresponds to heavy tail distributions.

**Theorem 4.1.4.** *Convolution of Heavy Tail Distributions*

If  $F$  and  $G$  are asymptotically smooth distributions on  $\mathbb{R}^+$ , such that  $\bar{F} \in RV_{-\alpha}$  and  $\bar{G} \in RV_{-\beta}$ , with  $\alpha \vee \beta < 1$ , then

$$\lim_{t \rightarrow \infty} \frac{1 - F * G(t) - \bar{F}(t) - \bar{G}(t)}{\bar{F}(t)\bar{G}(t)} = I(\alpha, \beta) + I(\beta, \alpha) + 2^{\alpha+\beta} - 2^\alpha - 2^\beta,$$

with  $I(\alpha, \beta) = \int_0^{1/2} ((1-w)^{-\alpha} - 1)\beta w^{-\beta-1} dw$ , and similarly for  $I(\beta, \alpha)$ .

Therefore, for the GPD case, we have:

$$\begin{aligned} F_A * F_B(t) &\approx 1 - \bar{F}_A(t) - \bar{F}_B(t) \\ &\quad - \bar{F}_A(t)\bar{F}_B(t) [2^{1/k_1+1/k_2} - 2^\alpha - 2^\beta + I(1/k_1, 1/k_2) + I(1/k_2, 1/k_1)] \end{aligned}$$

as  $t \rightarrow \infty$ , for  $k_1, k_2 \geq 1$ , and with the  $I$  function defined as previously .

Hence,

$$F_X(t) \approx 1 - \bar{F}_A(t) - \bar{F}_B(t), \text{ as } t \rightarrow \infty.$$

Therefore, as in the previous case, the fourth condition in (4.1) does not hold.

In conclusion, the convolution models offer an easy expression of the general dependence between variables. As well, they constitute realistic and non-standard marginal tail models. However, they can only produce complete tail dependence or exact tail independence. Therefore, these models do not fulfill all the conditions in (4.1). In the next section, we present a modification of the convolution model to overcome the problems identified in this section.

## 4.2 Product Models

The convolution models failed to embrace different types of asymptotic dependence. They lacked as well a mathematically simple representation of the distributions. These two facts are related, since the mathematics involved in the convolution operation as defined in (4.4) only lead to the extreme tail dependence cases.

In this section, we present another trivariate reduction method that eases the computation of the distributions: the product model.

### 4.2.1 General GPD Product Model

The general representation of the product model is as follows:

$$\begin{aligned} X &= (B \times D), \\ Y &= (C \times D). \end{aligned} \tag{4.14}$$

This model can be viewed as a modification of the convolution model, since it is precisely a convolution in log scale, where:

$$\log X = \log B + \log D, \text{ and } \log Y = \log C + \log D.$$

For ease of explanation, we will take  $A = D^{-1}$ . As in any trivariate reduction model, of main importance is the choice of the subvariables distributions.

However, with the following model, we can get a GPD representation:

$$\begin{aligned} X &= (B/A), \\ Y &= (C/A), \end{aligned} \tag{4.15}$$

with independent  $B \sim \text{Exp}(\alpha_1)$ ,  $C \sim \text{Exp}(\alpha_2)$ ,  $A \sim \text{Ga}(\theta, \sigma)$ , for  $\{\theta, \alpha_1, \alpha_2, \sigma\} \in \mathbb{R}_+^4$ . Notice that we could have written, for example,  $B/A$  as  $B \times D$ , where  $D \sim \text{Inv-Gamma}(\theta, \sigma)$ , so that it had a proper product representation. However, we keep the first representation for ease of explanation.

Then, the marginal is as follows:

$$\begin{aligned} P(X \leq x) &= P(B \leq Ax) = \int_0^\infty P(B \leq ax|a) f_A(a) da \\ &= \frac{\sigma^\theta}{\Gamma(\theta)} \int_0^\infty [1 - e^{-\alpha_1 ax}] e^{-\sigma a} a^{\theta-1} da \\ &= 1 - \frac{\sigma^\theta}{\Gamma(\theta)} \int_0^\infty e^{-a(\sigma + \alpha_1 x)} a^{\theta-1} da \\ &= 1 - \left(1 + \frac{\alpha_1}{\sigma} x\right)^{-\theta}, \end{aligned}$$

and, similarly:

$$P(Y \leq y) = 1 - \left(1 + \frac{\alpha_2}{\sigma} y\right)^{-\theta}.$$

Thus, the marginals for  $X$  and  $Y$  are  $GPD\left(\frac{1}{\theta}, \frac{\alpha_1 \theta}{\sigma}\right)$  and  $GPD\left(\frac{1}{\theta}, \frac{\alpha_2 \theta}{\sigma}\right)$ , respectively.

On the other hand, the joint survival function is given by:

$$\begin{aligned} P(X > x, Y > y) &= \int_0^\infty P(B > ax, C > ay|a) f_A(a) da \\ &= \frac{\sigma^\theta}{\Gamma(\theta)} \int_0^\infty e^{-a(\sigma + \alpha_1 x + \alpha_2 y)} a^{\theta-1} da \\ &= 1 - \left(1 + \frac{\alpha_1 x + \alpha_2 y}{\sigma}\right)^{-\theta}, \end{aligned}$$

for  $x > 0, y > 0$ .

Consequently, this model leads to equivalent marginal tails, since both have the same shape parameter. Therefore, it is not a realistic model; it violates the first condition in (4.1). Hence, we need to modify the tail of each variable.

### 4.2.2 Exponential-Gamma Model 1

The model we present in this section modifies the shape of the distributions of the previous model by adding two shape parameters as follows:

$$\begin{aligned} X &= (B/A)^{\psi_1}; \\ Y &= (C/A)^{\psi_2}; \end{aligned} \tag{4.16}$$

with independent  $B \sim \text{Exp}(\alpha_1)$ ,  $C \sim \text{Exp}(\alpha_2)$ ,  $A \sim \text{Ga}(\theta, 1)$ , with

$$\{\theta, \alpha_1, \alpha_2, \psi_1, \psi_2\} \in \mathbb{R}_+^5.$$

We will denote this as Model 1.

Note that in the previous model, the parameter  $\sigma$  did not have an important role, since the parameters  $\alpha_1$  and  $\alpha_2$  can be taken as the scale parameters. This is the reason why we fixed  $\sigma = 1$ .

### 4.2.2.1 Marginal Behavior

The marginal distribution is given by:

$$\begin{aligned}
 P(X \leq x) &= P(B \leq Ax^{\frac{1}{\psi_1}}) = \int_0^\infty P\left(B \leq ax^{\frac{1}{\psi_1}} \mid a\right) f_A(a) da \\
 &= \frac{1}{\Gamma(\theta)} \int_0^\infty \left[1 - e^{-\alpha_1 ax^{\frac{1}{\psi_1}}}\right] e^{-a} a^{\theta-1} da \\
 &= 1 - \left(1 + \alpha_1 x^{\frac{1}{\psi_1}}\right)^{-\theta},
 \end{aligned} \tag{4.17}$$

and, similarly:

$$P(Y \leq y) = 1 - \left(1 + \alpha_2 y^{\frac{1}{\psi_2}}\right)^{-\theta}.$$

These lead to the following densities:

$$f_X(x) = \frac{\theta \alpha_1}{\psi_1} x^{\frac{1}{\psi_1}-1} \left(1 + \alpha_1 x^{\frac{1}{\psi_1}}\right)^{-\theta-1} \text{ and } f_Y(y) = \frac{\theta \alpha_2}{\psi_2} y^{\frac{1}{\psi_2}-1} \left(1 + \alpha_2 y^{\frac{1}{\psi_2}}\right)^{-\theta-1}. \tag{4.18}$$

Thus, it means that  $F_X$  and  $F_Y$  are Burr type XII (Burr-XII) distributions. We now describe the main properties of the Burr-XII distribution. Refer to Burr (1942), Fry (1993), Johnson *et al.* (1994), Rodriguez (1977), and Tadikamalla (1980) for more details.

For  $X$  as in (4.16), define, as in Burr (1942),  $Z = \alpha_1 X^{\frac{1}{\psi_1}}$ , then

$$f_Z(z) = \theta(1+z)^{-\theta-1},$$

$z > 0$ . Therefore, the  $r$ th-moment of  $X$  is given by:

$$E(X^r) = E\left(\alpha_1^{-r\psi_1} Z^{r\psi_1}\right) = \alpha_1^{-r\psi_1} \theta \text{Be}(r\psi_1 + 1, \theta - r\psi_1), \text{ for } r < \frac{\theta}{\psi_1},$$

where  $\text{Be}(w, v)$  is the Beta function. In particular,  $X$  will have a finite mean only when  $\psi_1 < \theta$ . Hence, the marginal distributions covers heavy tails.

Another feature of the Burr-XII distribution is the characterization of the shape of the density via the mode. The derivative of the density of  $X$  in (4.18) is expressed as:

$$\frac{d}{dx} f_X(x) = \frac{\theta \alpha_1}{\psi_1^2} x^{\frac{1}{\psi_1}-2} \left(1 + \alpha_1 x^{\frac{1}{\psi_1}}\right)^{-\theta-1} \left[ (1 - \psi_1) - (\theta + 1) \alpha_1 x^{\frac{1}{\psi_1}} \left(1 + \alpha_1 x^{\frac{1}{\psi_1}}\right)^{-1} \right].$$

Hence, the mode is given by:

$$\tilde{X} = \left( \frac{1 - \psi_1}{\alpha_1(\theta + \psi_1)} \right)^{\psi_1}, \text{ for } \psi_1 < 1,$$

and therefore:

- The marginal of  $X$  is unimodal if  $\alpha_1 \theta + \psi_1(1 + \alpha_1) < 1$ .
- The marginal of  $X$  is L-shaped if  $\alpha_1 \theta + \psi_1(1 + \alpha_1) \geq 1$ .

In Figure D.3, we present an example of each one of these cases.

Rodriguez (1977) gives a more detailed analysis of the Burr-XII distribution. One of the interesting results is the following: suppose  $W|w_1 \sim \text{Weibull}$ , with  $w_1$  a scale parameter, and that  $w_1 \sim \text{Gamma}$ . Then,  $W \sim \text{Burr XII}$ . Another source of Burr-XII distributions is Fry (1993), who treats both the univariate and multivariate cases of all types of Burr distributions. Nevertheless, these models differ considerably from the models we are introducing.

Compared to the GPD product model of Section 4.2.1, Model 1 leads to different marginal distributions.

### 4.2.2.2 Dependence Structure and Joint Distribution

In the next theorem, we present the tail dependence structure and its proof can be found in Appendix B.3.

#### Theorem 4.2.1. Tail Dependence of the Exponential-Gamma Model 1

Define  $X$  and  $Y$  as in (4.16). Then, the tail dependence between  $X$  and  $Y$  is:

$$\chi = 2^{-\theta}. \quad (4.19)$$

And the following special cases are hold:

1. If  $\theta \rightarrow \infty$ ,  $X$  and  $Y$  are asymptotically exact independent.
2. If  $\theta \rightarrow 0$ ,  $X$  and  $Y$  are asymptotically complete dependent.

Therefore, when studying the tail dependence structure of two variables, inference should focus on the estimation of  $\theta$ . As a last step before this inference, we present the joint distribution:

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \leq x, Y \leq y) \\ &= \int_0^\infty P(B \leq x^{\frac{1}{\psi_1}} a, C \leq y^{\frac{1}{\psi_2}} a) f_A(a) da \\ &= \frac{1}{\Gamma(\theta)} \int_0^\infty \left(1 - e^{-\alpha_1 a x^{\frac{1}{\psi_1}}}\right) \left(1 - e^{-\alpha_2 a y^{\frac{1}{\psi_2}}}\right) e^{-a} a^{\theta-1} da \\ &= 1 - \left(1 + \alpha_1 x^{\frac{1}{\psi_1}}\right)^{-\theta} - \left(1 + \alpha_2 y^{\frac{1}{\psi_2}}\right)^{-\theta} + \left(1 + \alpha_1 x^{\frac{1}{\psi_1}} + \alpha_2 y^{\frac{1}{\psi_2}}\right)^{-\theta}, \end{aligned} \quad (4.20)$$

and, consequently, the joint density is:

$$f_{X,Y}(x, y) = \theta(\theta + 1) \frac{\alpha_1 \alpha_2}{\psi_1 \psi_2} x^{\frac{1}{\psi_1}-1} y^{\frac{1}{\psi_2}-1} \left(1 + \alpha_1 x^{\frac{1}{\psi_1}} + \alpha_2 y^{\frac{1}{\psi_2}}\right)^{-\theta-2}. \quad (4.21)$$

It is worth noticing the difference between the joint distribution in (4.20) and the corresponding convolution model in (4.13), where we could only take either the tail of the marginal of  $X$  or that of  $Y$ . In the current model, we have the joint survival term given by:

$$P(X > x, Y > y) = \left(1 + \alpha_1 x^{\frac{1}{\psi_1}} + \alpha_2 y^{\frac{1}{\psi_2}}\right)^{-\theta},$$

which represents the dependence structure and makes possible as well the tail dependence, rather than only the exact and complete dependence. The next step is about inference for Model 1.

#### 4.2.2.3 Inference Exclusive to $\theta$ . Model 1

We use the MLE for Model 1, since this methodology works. However, in general, we will use the Bayesian framework.

Let  $\mathbf{D} = (\mathbf{X}_n, \mathbf{Y}_n)$  be a sample of  $n$  i. i. d. vectors, where  $\mathbf{X}_n = (x_1, \dots, x_n)^T$  and  $\mathbf{Y}_n = (y_1, \dots, y_n)^T$ . Define the parameter space as  $\Omega = \mathbb{R}_+^5$  and define  $\omega \in \Omega$  as:

$$\omega = \{\theta, \alpha_1, \alpha_2, \psi_1, \psi_2\}.$$

Then, from (4.21), the likelihood function of Model 1 is given by:

$$L(\omega | \mathbf{D}) = \left(\theta(\theta + 1) \frac{\alpha_1 \alpha_2}{\psi_1 \psi_2}\right)^n \prod_{i=1}^n x_i^{\frac{1}{\psi_1}-1} y_i^{\frac{1}{\psi_2}-1} \left(1 + \alpha_1 x_i^{\frac{1}{\psi_1}} + \alpha_2 y_i^{\frac{1}{\psi_2}}\right)^{-\theta-2},$$

and, consequently the log-likelihood function is:

$$\begin{aligned}
 l(\boldsymbol{\omega} | \mathbf{D}) &= \log L(\boldsymbol{\omega} | \mathbf{D}) \\
 &= n \log \theta + n \log(\theta + 1) + n \log \left( \frac{\alpha_1 \alpha_2}{\psi_1 \psi_2} \right) + \left( \frac{1}{\psi_1} - 1 \right) \sum_{i=1}^n \log x_i \\
 &\quad + \left( \frac{1}{\psi_2} - 1 \right) \sum_{i=1}^n \log y_i - (\theta + 2) \sum_{i=1}^n \log \left( 1 + \alpha_1 x_i^{\frac{1}{\psi_1}} + \alpha_2 y_i^{\frac{1}{\psi_2}} \right),
 \end{aligned} \tag{4.22}$$

As a first step, we fix the elements of  $\boldsymbol{\omega}$  as known, except for  $\theta$ . The aim of this section is to study the properties of the MLE of  $\theta$ . Therefore, we need to find this estimator as follows:

$$\frac{d}{d\theta} l(\theta | \mathbf{D}) = \frac{n}{\theta} + \frac{n}{\theta + 1} - \sum_{i=1}^n \log \left( 1 + \alpha_1 x_i^{\frac{1}{\psi_1}} + \alpha_2 y_i^{\frac{1}{\psi_2}} \right),$$

and denote:

$$M = \frac{1}{n} \sum_{i=1}^n \log \left( 1 + \alpha_1 x_i^{\frac{1}{\psi_1}} + \alpha_2 y_i^{\frac{1}{\psi_2}} \right),$$

then

$$\frac{d}{d\theta} l(\theta | \mathbf{D}) = 0 \iff \hat{\theta} = \frac{2 - M \pm (M^2 + 4)^{1/2}}{2M},$$

for which, the only positive solution for all the parameter space is:

$$\hat{\theta} = \frac{1}{M} + \frac{1}{2} \left( 1 + \left( \frac{2}{M} \right)^2 \right)^{1/2} - \frac{1}{2}, \tag{4.23}$$

and given that

$$\frac{d^2}{d\theta^2} l(\theta | \mathbf{D}) = -\frac{n}{\theta^2} - \frac{n}{(\theta + 1)^2} < 0, \tag{4.24}$$

we conclude  $\hat{\theta}$  provides a maximum.

Recall from (4.19) that the tail dependence is expressed as  $\chi = 2^{-\theta}$ , then the extreme tail dependence cases are given by:

- If  $M \rightarrow \infty$ , then  $\hat{\theta} \rightarrow 0$ , then  $\chi = 1$ . Therefore,  $X$  and  $Y$  are asymptotically complete dependent.
- If  $M \rightarrow 0$ , then  $\hat{\theta} \rightarrow \infty$ , then  $\chi = 0$ . Therefore,  $X$  and  $Y$  are asymptotically exact independent.

The first feature of  $\hat{\theta}$  is consistency, as the following theorem demonstrates.

**Theorem 4.2.2.** *Consistency of  $\hat{\theta}$*

*The MLE  $\hat{\theta}$  of  $\theta$  found in (4.23) is consistent, i.e.:*

$$\hat{\theta} \xrightarrow{P} \theta. \quad (4.25)$$

In Appendix B.4 the conditions for which  $\hat{\theta} \xrightarrow{P} \theta$  are established.

The second feature of  $\hat{\theta}$  is asymptotic normality. We establish this property in the following theorem (see Newey and McFadden (1994) or Schervish (1995)).

**Theorem 4.2.3.** *Asymptotic Normality of MLE*

*Let  $\hat{\varphi}$  be an MLE of a parameter  $\varphi$ . Then, under regularity conditions*

$$\sqrt{n}(\hat{\varphi} - \varphi) \xrightarrow{d} N(0, I^{-1}(\hat{\varphi})), \quad (4.26)$$

*where  $I(\cdot)$  is the Fisher information matrix.*

In Appendix B.5, we prove that the regularity conditions hold for (4.26).

For the construction of confidence intervals for  $\theta$ , we need to elaborate a bit further, since  $\theta > 0$ . Rearranging (4.26), we get that:

$$\frac{\hat{\varphi} - \varphi}{(n I(\hat{\varphi}))^{-1/2}} = \frac{\hat{\varphi} - \varphi}{SD(\hat{\varphi})} \xrightarrow{d} N(0, 1), \quad (4.27)$$

where  $SD$  stands for Standard Deviation.

Define  $g(\theta) = \log(\theta)$ . Then, by invariance of the MLE,  $g(\hat{\theta})$  is an MLE of  $g(\theta)$ . Therefore,

$$\frac{\log(\hat{\theta}) - \log(\theta)}{SD(\log(\hat{\theta}))} \xrightarrow{d} N(0, 1).$$

By the Taylor expansion of  $\log(\hat{\theta})$  around  $\theta$ , we get that:

$$\log(\hat{\theta}) \approx \log(\theta) + \frac{1}{\theta}(\theta - \hat{\theta}) - \frac{1}{2!\theta^2}(\theta - \hat{\theta})^2 + \dots$$

which leads to the following approximation:

$$\log(\hat{\theta}) \approx \log(\theta) + \frac{1}{\theta}(\theta - \hat{\theta}).$$

Therefore:

$$\text{Var}(\log(\hat{\theta})) \approx \frac{1}{\theta^2} \text{Var}(\hat{\theta}),$$

which means:

$$SD(\log(\hat{\theta})) \approx \frac{1}{\theta} SD(\hat{\theta}).$$

Hence,

$$\frac{\log(\hat{\theta}) - \log(\theta)}{\frac{SD(\log(\hat{\theta}))}{\theta}} \xrightarrow{d} N(0, 1).$$

This approximation leads to the following construction of the 95% confidence interval:

$$P\left(\hat{\theta} \exp\left\{-1.96 \frac{SD(\log(\hat{\theta}))}{\theta}\right\} < \theta < \hat{\theta} \exp\left\{1.96 \frac{SD(\log(\hat{\theta}))}{\theta}\right\}\right) \approx 0.95.$$

Then, for data  $\mathbf{D}$  and the log-likelihood defined as in (4.22), we make the following approximation:

$$\frac{SD(\log(\hat{\theta}))}{\theta} \approx \frac{(n \hat{I}(\hat{\theta}))^{-1/2}}{\hat{\theta}},$$

where

$$\hat{I}(\hat{\theta}) = -\frac{1}{n} \sum_{i=1}^n \frac{\delta^2}{\delta\theta^2} l(\theta | \mathbf{D}) \Big|_{\theta=\hat{\theta}}.$$

Therefore, the 95% confidence interval is given by:

$$\left[ \hat{\theta} \exp \left\{ -\frac{1.96}{(n\hat{I}(\hat{\theta}))^{1/2} \hat{\theta}} \right\}, \hat{\theta} \exp \left\{ \frac{1.96}{(n\hat{I}(\hat{\theta}))^{1/2} \hat{\theta}} \right\} \right].$$

We present an illustration for inference with this method. We simulated observations from model (4.16). We consider four levels of tail dependence:  $\theta \in \{5, 3, 1, 1/3, 0.2\}$ . In terms of heaviness of the marginal tails, we split data in heavy tails:  $\psi_1/\theta = 1$  and  $\psi_2/\theta = 1.5$ ; and light tails:  $\psi_1/\theta = 0.25$  and  $\psi_2/\theta = 0.2$ . We simulated 500 observations for each model, with 10 replications for each model. In Table C.1 we show the the estimations of  $\theta$ ,  $\chi$ , and 95% confidence intervals via the MLE method. We present, as well, the standard error of the replications. As it can be seen,  $\hat{\theta}$  is considerably close to all the true values of  $\theta$ , all the confidence intervals contain the true value  $\theta$ , and the standard errors are small.

#### 4.2.2.4 Summary of Model 1

Model 1 represents a framework that embraces a wider range of tail behavior than the previous models. Even more, consistency and asymptotic normality is achieved for the dependence parameter  $\theta$ . However, before going further on the inference, we need to address some drawbacks of the model.

If we look at the marginal tail in (4.17), we can approximate it by:

$$1 - P(X \leq x) \approx \alpha_1^{-\theta} x^{-\frac{\theta}{\psi_1}},$$

for large  $x$ . Then,  $\theta$  dictates both the marginal tail and the dependence structure. And, for large data, there exists an identification issue between  $\theta$  and  $\psi_1$  or  $\psi_2$ . Hence, the second property of (4.1) is violated in Model 1, since the marginal tails are linked to the dependence structure. Therefore, we need to isolate  $\theta$  only to model the joint behavior.

Another drawback of Model 1 is the fact that it does not factorize to the marginal effects when considering the tail independence case ( $\theta \rightarrow \infty$ ). In fact, if we proceed as in Johnson *et al.* (1994) (p.687):

$$\begin{aligned} P(X \leq \theta^{-\psi_1} x) &= 1 - \left(1 + \frac{x^{\frac{1}{\psi_1}}}{\theta}\right)^{-\theta} \\ &= 1 - \exp \left\{ -\theta \log \left(1 + \frac{x^{\frac{1}{\psi_1}}}{\theta}\right) \right\} \\ &= 1 - \exp \left\{ -\theta \left( \frac{x^{\frac{1}{\psi_1}}}{\theta} - \frac{1}{2} \left( \frac{x^{\frac{1}{\psi_1}}}{\theta} \right)^2 + \dots \right) \right\} \\ &= 1 - e^{-x^{\frac{1}{\psi_1}}}, \text{ as } \theta \rightarrow \infty, \end{aligned}$$

which means that the tail independent case implies directly that the marginal tails must be exclusively of the Gumbel type, which is not a heavy-tailed type.

In the next chapter, we introduce changes in Model 1 to deal with the drawbacks found here.

# Chapter 5

## The Bivariate Model

In this chapter we introduce the bivariate model, which will overcome the drawbacks of the models found in Chapter 4. This model will become the individual bivariate model of the mixture model introduced in the next chapter. In fact, it will determine the way to construct the parametric family of all the individual models in the mixture model.

In Section 5.1, we introduce the representation of the final model (Model 2). The marginal distribution derivation is found in Section 5.2. The joint distribution and the dependence structure are presented in Section 5.3.

In Section 5.5 we discuss what type of data Model 2 can be used for. We define the regions where the data is modelled. As well, we define two ways of dealing with data. However, we will only keep one of them for the rest of the thesis.

We present some similarities and comparisons of Model 2 with the literature

in Section 5.4.

Finally, in Section 5.6, we deal with inference for both the MLE and the Bayesian framework. This includes the introduction of the likelihood and posterior functions for the different ways of dealing with data.

## 5.1 Exponential-Gamma Model 2

As a modification of (4.16), we propose the following bivariate model:

$$\begin{aligned} X &= (1 + B/A)^{\theta\xi_1}; \\ Y &= (1 + C/A)^{\theta\xi_2}; \end{aligned} \tag{5.1}$$

with independent  $B \sim \text{Exp}(\alpha_1)$ ,  $C \sim \text{Exp}(\alpha_2)$ ,  $A \sim \text{Ga}(\theta, 1)$ ,  $\alpha_1, \alpha_2, \theta > 0$ , and  $\xi_1, \xi_2 > 0$ .

There are two fundamental changes to this model and they are related to eliminating the drawbacks of Model 1 in Section 4.1.1:

1. The appearance of  $\theta$  in the exponent. This will improve estimation performance, since it will separate the marginal from the joint estimation of the shape parameters.
2. The inclusion of the “1+” term. This will facilitate the model to factorize into the marginals in the tail independence case (i.e. as  $\theta \rightarrow \infty$ ).

The second change implies that we are only able to model data larger than

1. However, there is no loss of generality, since, as we will address later, Model

2 fits heavy and semi-heavy tailed distributions. Therefore, a change of scale on the observations does not affect the tail behavior. In a broad sense, instead of defining the origin in the support of the random vector  $(X, Y)$  in  $(0, 0)$ , we are defining it in  $(1, 1)$ .

## 5.2 Marginal Distributions: Model 2

In order to illustrate the effect of  $\theta$  in the exponent, and to sketch the changes we are making on the distributions, we compute the marginal distribution.

This is given by:

$$\begin{aligned}
 F_X(x) &= P(B < (x^{1/\theta\xi_1} - 1)A) \\
 &= \int_0^\infty P(B < (x^{1/\theta\xi_1} - 1)a \mid a) f_A(a) da \\
 &= 1 - \frac{1}{\Gamma(\theta)} \int_0^\infty \exp(-\alpha_1 a (x^{1/\theta\xi_1} - 1)) \exp(-a) a^{\theta-1} da \\
 &= 1 - \frac{1}{\Gamma(\theta)} \int_0^\infty \exp(-a(1 + \alpha_1(x^{1/\theta\xi_1} - 1))) a^{\theta-1} da \\
 &= 1 - \left(1 + \alpha_1(x^{\frac{1}{\theta\xi_1}} - 1)\right)^{-\theta},
 \end{aligned} \tag{5.2}$$

and so the marginal density function is given by:

$$f_X(x) = \left(\frac{\alpha_1}{\xi_1}\right) x^{\frac{1}{\theta\xi_1}-1} \left(1 + \alpha_1(x^{\frac{1}{\theta\xi_1}} - 1)\right)^{-\theta-1}.$$

With the introduction of  $\theta$  in the exponent, we eliminate its influence on the shape parameter of the marginal tail, since:

$$\begin{aligned}
 P(X > x) &= \left(1 + \alpha_1(x^{\frac{1}{\theta\xi_1}} - 1)\right)^{-\theta} \\
 &\approx \alpha_1^{-\theta} x^{-\frac{1}{\xi_1}}, \text{ as } x \rightarrow \infty.
 \end{aligned} \tag{5.3}$$

In Section 4.1.1, we saw that Model 1 has a Burr-XII marginal distribution, and recall that  $W \sim \text{Burr-XII}(c, k)$ , if

$$P(W \leq w) = 1 - (1 + w^c)^{-k}, \text{ for } w > 0,$$

or, including  $\delta$  and  $\gamma$ , location and scale parameters, respectively:

$$P(W \leq w) = 1 - \left(1 + \frac{(w - \delta)^c}{\gamma^c}\right)^{-k}, \text{ for } w > \delta.$$

Thus, the marginal model in (5.2) is not Burr-XII. However, because of the similarity, we will refer it as a quasi Burr-XII type.

It is possible to compute the  $r$ th-moment, similarly to Model 1, as follows:

For  $X$  defined as in (5.1), define  $Z = \alpha_1 \left(X^{\frac{1}{\theta\xi_1}} - 1\right)$ . Then

$$f_Z(z) = \theta(1 + z)^{-\theta-1}.$$

Therefore, the  $r$ th-moment of  $X$  is given by:

$$E(X^r) = E\left(\alpha_1^{-r\theta\xi_1}(1 + Z)^{r\theta\xi_1}\right) = \alpha_1^{-r\theta\xi_1}\theta \int_0^\infty (1 + z)^{r\theta\xi_1-\theta-1} dz,$$

which yields:

$$E(X^r) = \begin{cases} \infty & \text{if } r \geq \frac{1}{\xi_1} \\ \frac{\alpha_1^{r\theta\xi_1}}{r\xi_1} & \text{if } r < \frac{1}{\xi_1}. \end{cases}$$

Hence, the heaviness of the tail depends only on the shape parameter  $\xi_1$ .

Therefore, the marginals of the final model cover heavy tails.

The case when  $\theta \rightarrow \infty$  will play an important role in the coming sections.

Thus, we present how the marginal behaves as  $\theta \rightarrow \infty$ :

$$\begin{aligned}
 P(X > x) &= \left(1 + \alpha_1(x^{\frac{1}{\theta\xi_1}} - 1)\right)^{-\theta} \\
 &\approx \left(1 + \frac{\alpha_1 \log x^{\frac{1}{\xi_1}}}{\theta}\right)^{-\theta} \\
 &\approx \exp\left\{-\alpha_1 \log x^{\frac{1}{\xi_1}}\right\} \\
 &= x^{-\frac{\alpha_1}{\xi_1}}, \text{ as } \theta \rightarrow \infty.
 \end{aligned} \tag{5.4}$$

In the first approximation, we used the following:

$$\frac{z^\rho - 1}{\rho} \approx \log z, \text{ as } \rho \rightarrow 0. \tag{5.5}$$

The expression in (5.4) is an improvement from Model 1, where an asymptotic expression for  $P(X > x)$  when  $\theta \rightarrow \infty$  was not possible to achieve.

Similar to Burr-XII distributions, we can characterize the shape of the density via the mode. The derivative of the density in (5.3) is given by:

$$\begin{aligned}
 \frac{d}{dx} f_X(x) &= \frac{(1 - \theta\xi_1)\alpha_1}{\theta\xi_1^2} x^{\frac{1}{\theta\xi_1}-2} \left(1 + \alpha_1(x^{\frac{1}{\theta\xi_1}} - 1)\right)^{-\theta-1} \\
 &\quad - \frac{(\theta + 1)\alpha_1^2}{\theta\xi_1^2} x^{\frac{2}{\theta\xi_1}-2} \left(1 + \alpha_1(x^{\frac{1}{\theta\xi_1}} - 1)\right)^{-\theta-2},
 \end{aligned}$$

which is equal to zero if and only if

$$\frac{(1 - \theta\xi_1)}{\alpha_1(\theta + 1)} = \frac{x^{\frac{1}{\theta\xi_1}}}{1 - \alpha_1 + \alpha_1 x^{\frac{1}{\theta\xi_1}}}.$$

Then, the mode is given by

$$\tilde{x} = \left(\frac{(1 - \theta\xi_1)(1 - \alpha_1)}{\alpha_1\theta(1 + \xi_1)}\right)^{\theta\xi_1}.$$

Hence, the marginal density is:

- Unimodal, when  $\alpha_1(1 + \theta) + \theta\xi_1 < 1$ .

- L-shaped, when  $\alpha_1(1 + \theta) + \theta\xi_1 \geq 1$ .

We present two examples of each case in Figure D.4.

### 5.3 Joint Distribution and Dependence

Here, we introduce the joint survival and joint distributions for Model 2, as follows:

$$\begin{aligned}
 \bar{F}_{X,Y}(x, y) &= P(X > x, Y > y) \\
 &= P(B > (x^{1/\theta\xi_1} - 1)A, C > (y^{1/\theta\xi_2} - 1)A) \\
 &= \frac{1}{\Gamma(\theta)} \int_0^\infty \exp(-\alpha_1 a (x^{1/\theta\xi_1} - 1) - \alpha_2 a (y^{1/\theta\xi_2} - 1) - a) a^{\theta-1} da \\
 &= \frac{1}{\Gamma(\theta)} \int_0^\infty \exp(-a (1 + \alpha_1 x^{1/\theta\xi_1} - \alpha_1 + \alpha_2 y^{1/\theta\xi_2} - \alpha_2)) a^{\theta-1} da \\
 &= \left(1 + \alpha_1 (x^{\frac{1}{\theta\xi_1}} - 1) + \alpha_2 (y^{\frac{1}{\theta\xi_2}} - 1)\right)^{-\theta},
 \end{aligned} \tag{5.6}$$

and hence

$$f(x, y) = \left(\frac{(\theta + 1)\alpha_1\alpha_2}{\theta\xi_1\xi_2}\right) x^{\frac{1}{\theta\xi_1}-1} y^{\frac{1}{\theta\xi_2}-1} \left(1 + \alpha_1 (x^{\frac{1}{\theta\xi_1}} - 1) + \alpha_2 (y^{\frac{1}{\theta\xi_2}} - 1)\right)^{-\theta-2}. \tag{5.7}$$

From (5.6), we can see that the joint distribution is closed to marginalization, since:

$$\bar{F}_{X,Y}(x, 1) = \left(1 + \alpha_1 (x^{\frac{1}{\theta\xi_1}} - 1)\right)^{-\theta},$$

which is the marginal survival function of  $X$ .

The next issue to consider is the asymptotic properties of the model. Define, as in Theorem (2.2.3), and as in (2.30), the  $S$  and  $T$  transformations

as:

$$S = -\frac{1}{\log F_X(X)} \quad \text{and} \quad T = -\frac{1}{\log F_Y(Y)}. \quad (5.8)$$

Then, the joint survival distribution of  $S$  and  $T$  is given by:

$$\begin{aligned} \bar{F}_{S,T}(s, t) &= \left\{ (1 - e^{-1/s})^{-\frac{1}{\theta}} + (1 - e^{-1/t})^{-\frac{1}{\theta}} - 1 \right\}^{-\theta} \\ &\approx \left\{ s^{\frac{1}{\theta}} + t^{\frac{1}{\theta}} - 1 \right\}^{-\theta}, \end{aligned} \quad (5.9)$$

for large  $s$  and  $t$  (see Appendix B.6 for details).

The following theorem shows that the asymptotically complete dependence and the asymptotically exact independence rely only on  $\theta$ , as in Model 1.

**Theorem 5.3.1.** *Tail Dependence Characterization*

Define  $X$  and  $Y$  as in (5.1), then:

1. If  $\theta \rightarrow \infty$ , then  $X$  and  $Y$  are asymptotically exact independent;
2. If  $\theta \rightarrow 0$ , then  $X$  and  $Y$  are asymptotically complete dependent.

We present the proof in Appendix B.6. As can be seen from the proof, the distribution of the model splits into the product of both marginals when the independence case is considered. This is precisely the advantage of including the “1+” term in (5.1). In a similar result, we find that in the limit of  $\theta \rightarrow \infty$ , the joint survival function behaves as a complete dependence function.

As a final step, we need to introduce the dependence structure of Model 2 via the following theorem.

**Theorem 5.3.2.** *Tail Dependence of Model 2*

Define  $X$  and  $Y$  as in (5.1). Then, the tail dependence between  $X$  and  $Y$  is:

$$\chi = 2^{-\theta}. \quad (5.10)$$

Therefore, for the asymptotic dependence, inference should focus on the estimation of  $\theta$ . We present the proof of the theorem in Appendix B.7.

We note that the model, for finite  $\theta$ , is always modeling a tail dependence. However, for inference purposes, a large  $\theta$  would indicate a very small dependence, which could suggest exact independence.

In a Bayesian setting, where prediction takes on importance, forming a prediction with a posterior for  $\theta$  centered on a large value will effectively be equivalent to prediction assuming tail independence. Also, it avoids the decision as to whether to take a model as independent or not. We regard this as an advantage of the Bayesian Model.

## 5.4 Model 2 and the Literature

It is important to notice some similarities of Model 2 with models in the literature.

From Theorems 5.3.1 and 5.3.2, we established that:

- $\lim_{\theta \rightarrow 0} \chi = 1$ : complete dependence.
- $\lim_{\theta \rightarrow \infty} \chi = 0$ : exact independence and the joint distribution factorises into the marginal distributions.

These dependence limits are equivalent to those of Ledford and Tawn (1996) in (2.27). In contrast to Ledford and Tawn, we smooth between the extremal dependences via the posterior of  $\theta$ , since  $\chi$  only depends on this parameter. The problems in estimation and in the tests found in the classical approach can be avoided within the Bayesian framework, i.e. using the posterior distribution of  $\theta$ . This aspect of Bayesian inference will be presented in Section 5.6.

Regarding the dependence measures described previously, we have that for Model 2:

$$\bar{\chi} = \lim_{s \rightarrow \infty} \frac{2 \log P(S > s)}{\log P(S > s, T > s)} - 1 = \lim_{s \rightarrow \infty} \frac{2 \log s}{\log \theta + \log s} - 1 = 1, \quad (5.11)$$

for  $\theta < \infty$ . Therefore, this parameter suggests that there always exists a dependence between the variables for finite  $\theta$ . This result is in accordance with the following:

$$P(S > s, T > s) \sim 2^{-\theta} s^{-1},$$

as  $s \rightarrow \infty$ , which means that  $\eta = 1$ .

This agrees with our model, but it is not a disadvantage. Our claim is that the posterior distribution for  $\theta$  is all that is needed. Similar to a prior on  $\rho$ , a correlation parameter in a normal distribution, a prior on  $(-1, 1)$ , there is no point mass at 0. If independence holds, then a posterior accumulating at 0 will occur, but no point mass.

It is worth to stress that Model 2 assumes either tail dependence or exact independence of the variables. Therefore, no other type of independence is assumed.

Regarding the form of the joint distribution and its properties, the result in (5.10) might recall some of the following copulas (refer to Heffernan (2000) for the notation and the references): the BB1, BB4 and BB7 of Joe (1997) on both their upper and lower tail, the BB3 of Joe (1997) on its upper tail, and the lower tail of the Pareto copula. However, we will only focus on two similar models: the lower joint tail of the Clayton distribution found in Ledford and Tawn (1996) and the bivariate logistic extreme value distribution.

Clayton (1978) constructed a bivariate model to study disease incidence in parents and offsprings. For  $Z$  and  $W$ , random variables with joint distribution  $F$ , and joint density  $f$ , define the hazard functions as:

$$\lambda_Z(z; w) = \frac{d}{dz} \{-\log F(z, w)\} = \int_w^\infty f(z, v) dv / F(z, w),$$

and

$$\lambda_W(w; z) = \frac{d}{dw} \{-\log F(z, w)\} = \int_z^\infty f(u, w) du / F(z, w).$$

Define, as well, the bivariate failure rate as:

$$l(z, w) = \frac{f(z, w)}{F(z, w)}.$$

Clayton was looking for a family of distributions for which:

$$\frac{l(z, w)}{\lambda_Z(z; w)\lambda_W(w; z)} = \phi,$$

for a constant  $\phi > 0$ . In other terms, a joint distribution for which:

$$f(z, w) \int_z^\infty \int_w^\infty f(u, v) dv du = \phi \int_z^\infty f(u, w) du \int_w^\infty f(z, v) dv.$$

Clayton found that:

$$f(z, w) = \phi a'(z)b'(w) \left( 1 + (\phi - 1)[a(z) + b(w)] \right)^{-\frac{1}{\phi-1}-2}, \quad (5.12)$$

where  $a(\cdot)$  and  $b(\cdot)$  are non-decreasing functions with  $a(0) = b(0) = 0$ . Clayton suggests a parametric form for these functions depending on some regression coefficients. For the case of  $z$ , let the regression coefficient be  $\gamma$ , then:

$$a(z) \propto (z - k)^r \exp(\beta^T u),$$

with  $u$  a covariate of  $z$ , and similarly for  $w$ .

Model 2 is almost (5.12) when:

$$\phi = 1 + \frac{1}{\theta} \text{ or } \theta = \frac{1}{\phi - 1},$$

and thus, the joint densities are comparable. What differentiates Model 2 is the form of the functions  $a(\cdot)$  and  $b(\cdot)$ :

1. Their support is  $(1, \infty)$ .
2. Their parametric form depends on  $\theta$ , so that their scale parameters do not depend on  $\theta$ .

Consider the representation of the model when transforming the variables to Fréchet marginals given in (5.9), i.e., for large  $x$  and  $y$  or equivalently, for large  $s$  and  $t$ , we have:

$$P(X > x, Y > y) = P(S > s, T > t) \approx (s^{1/\theta} + t^{1/\theta} - 1)^{-\theta}. \quad (5.13)$$

This expression is equivalent to the Clayton distribution (when applied to copulas), which is classified as an asymptotic dependence case by Ledford and Tawn. The survival function is:

$$\bar{F}(s, t) = \left( \bar{G}^*(s)^{-1/\beta} + \bar{G}^*(t)^{-1/\beta} - 1 \right)^{-\beta}, \quad (5.14)$$

where  $\bar{G}^*$  is the unit Fréchet survivor function and  $\beta \geq 0$ . This representation is the same as that of (5.13).

On the other hand, the expression (5.13) is similar to the exponent measure  $V$ , see (2.17), of the bivariate logistic extreme value distribution, which is a distribution of componentwise maxima; and is expressed as:

$$V_\gamma(s, t) = \left( s^{-1/\gamma} + t^{-1/\gamma} \right)^\gamma,$$

where  $0 < \gamma < 1$ .

The same result is found in the Clayton distribution where  $\mathcal{L}(s) = 2^{-\theta}$  and in the bivariate logistic, where  $\mathcal{L}(s) = 2 - 2^{-\gamma}$ .

Regarding the equality of the models (5.13) and (5.14), we need to make clear a couple of points. First of all, the equality holds only for the survival function for large values. The second and most important point is that we have constructed a model that differs with copula literature. In Model 2, the dependence structure is dictated by the definition of the variables, so that it is linked to the marginals, whereas in the copula literature, the dependence and the marginal structure are split. Therefore, we are not explicitly working with copulas.

## 5.5 Data Characterization for Model 2

In this section we describe what type of data can be modelled by Model 2. As we mentioned in previous sections, we are modelling positive random variables, let them be  $X^*$  and  $Y^*$ . We describe two ways to proceed with extreme data.

Recall from Section 2.2.3.2, and as Figure D.1 shows, we have divided  $\mathbb{R}^2$  into four regions as follows:

$$\{R_{ij}^*; i = \mathbb{1}_{[X^* > u_{X^*}]}, j = \mathbb{1}_{[Y^* > u_{Y^*}]}\}, \quad (5.15)$$

for  $u_{X^*}$  and  $u_{Y^*}$ , positive high thresholds of  $X^*$  and  $Y^*$ , respectively.

As a first step, we transform the variables  $X^*$  and  $Y^*$  as follows:

$$X = X^* + 1, \text{ and } Y = Y^* + 1, \quad (5.16)$$

for  $(X^*, Y^*) \in \{(0, u_{X^*}), (0, u_{Y^*})\}^c$ .

The dependence structure of the original variables is conserved, since the location shift does not affect the value of  $\chi$ . The transformed variables have the following support:

$$(X, Y) \in \{(1, u_X) \times (1, u_Y)\}^c,$$

with  $u_X = u_{X^*} + 1$ , and  $u_Y = u_{Y^*} + 1$ . Then, with respect to the transformed variables,  $\mathbb{R}^2$  is divided into four regions as follows:

$$\{R_{ij}; i = \mathbb{1}_{[X > u_X]}, j = \mathbb{1}_{[Y > u_Y]}\}. \quad (5.17)$$

Then, these regions are represented in Figure D.1, with the origin in  $(1, 1)$ .

We consider that data in  $R_{00}^*$  (and as a consequence, transformed data in  $R_{00}$ ) do not contribute with any relevant information of the extremes. Hence, we focus in modelling data falling in regions  $R_{10}$ ,  $R_{01}$ , and  $R_{11}$ .

We assume  $X$  and  $Y$  have Model 2 as distribution. However, since we are ignoring the region  $(1, u_X) \times (1, u_Y)$ , we need to truncate the distribution by the following factor:

$$1 - F(u_X, u_Y) = \left(1 + \alpha_1 \left(u_X^{\frac{1}{\theta\xi_1}} - 1\right)\right)^{-\theta} + \left(1 + \alpha_2 \left(u_Y^{\frac{1}{\theta\xi_2}} - 1\right)\right)^{-\theta} - \left(1 + \alpha_1 \left(u_X^{\frac{1}{\theta\xi_1}} - 1\right) + \alpha_2 \left(u_Y^{\frac{1}{\theta\xi_2}} - 1\right)\right)^{-\theta}. \quad (5.18)$$

Consider a sample of size  $n$  of  $X$  and  $Y$ , say  $\mathbf{D}$ , is given. In this way,  $\mathbf{D} = \{(x_{i,i}), i = 1, \dots, n\}$ . Let  $\boldsymbol{\omega} = \{\theta, \alpha_1, \xi_1, \alpha_2, \xi_2\} \in \Omega = \mathbb{R}_+^5$ .

Firstly, we introduce the ‘‘censored data’’ approach. This approach is equivalent to the model expressed in (2.24) in Section 2.2.3.2. The difference is that we are not interested in observations falling on the region  $R_{00}$ , i.e. we discard the observations in  $R_{00}$ .

Then, the likelihood of the censored case, similarly to (2.24), is given by:

$$L(\boldsymbol{\omega} | \mathbf{D}) = \prod_{i=1}^n g(\boldsymbol{\omega} | (x_i, y_i)), \quad (5.19)$$

where

$$g(\boldsymbol{\omega} | (x_i, y_i)) = \begin{cases} (1 - F(u_X, u_Y))^{-1} \times f(x_i, y_i) & \text{if } (x_i, y_i) \in R_{11} \\ (1 - F(u_X, u_Y))^{-1} \times \frac{\delta F}{\delta X} \Big|_{(x_i, u_Y)} & \text{if } (x_i, y_i) \in R_{10} \\ (1 - F(u_X, u_Y))^{-1} \times \frac{\delta F}{\delta Y} \Big|_{(u_X, y_i)} & \text{if } (x_i, y_i) \in R_{01}, \end{cases}$$

where, from (5.6) and (5.7):

$$\begin{aligned}
 f(x_i, y_i) &= \left( \frac{(\theta+1)\alpha_1\alpha_2}{\theta\xi_1\xi_2} \right) x_i^{\frac{1}{\theta\xi_1}-1} y_i^{\frac{1}{\theta\xi_2}-1} \left( 1 + \alpha_1(x_i^{\frac{1}{\theta\xi_1}} - 1) + \alpha_2(y_i^{\frac{1}{\theta\xi_2}} - 1) \right)^{-\theta-2} \\
 \frac{\delta F}{\delta X} \Big|_{(x_i, u_Y)} &= \left( \frac{\alpha_1}{\xi_1} \right) x_i^{\frac{1}{\theta\xi_1}-1} \left( 1 + \alpha_1(x_i^{\frac{1}{\theta\xi_1}} - 1) + \alpha_2(u_Y^{\frac{1}{\theta\xi_2}} - 1) \right)^{-\theta-1} \\
 \frac{\delta F}{\delta Y} \Big|_{(u_X, y_i)} &= \left( \frac{\alpha_2}{\xi_2} \right) y_i^{\frac{1}{\theta\xi_2}-1} \left( 1 + \alpha_1(u_X^{\frac{1}{\theta\xi_1}} - 1) + \alpha_2(y_i^{\frac{1}{\theta\xi_2}} - 1) \right)^{-\theta-1}.
 \end{aligned}$$

On the other hand, we introduce the “non-censored data” approach. The transformed variables are the same as in (5.16), and the interest is as well in the regions  $R_{10}$ ,  $R_{01}$ , and  $R_{11}$ . For the same sample  $\mathbf{D}$  and same parameter vector  $\boldsymbol{\omega}$ , the likelihood function of the non-censored approach is given by:

$$\begin{aligned}
 L(\boldsymbol{\omega}|\mathbf{D}) &= \left( \frac{(\theta+1)\alpha_1\alpha_2}{\theta\xi_1\xi_2} \right)^n \times (1 - F(u_X, u_Y))^{-n} \\
 &\quad \prod_{i=1}^n x_i^{\frac{1}{\theta\xi_1}-1} y_i^{\frac{1}{\theta\xi_2}-1} \left( 1 + \alpha_1(x_i^{\frac{1}{\theta\xi_1}} - 1) + \alpha_2(y_i^{\frac{1}{\theta\xi_2}} - 1) \right)^{-\theta-2},
 \end{aligned} \tag{5.20}$$

for  $(x_i, y_i) \in \{R_{00}\}^c$ , for  $i = 1, \dots, n$ .

We divide extremes inference in two objectives: tail dependence determination and prediction. The first one consists only in the estimation of  $\theta$ , whereas the second one focuses on fitting Model 2 to predict values of  $X^*$  and  $Y^*$ .

Both the censored and the non-censored approaches can achieve both the tail dependence determination and the prediction. The only difference between both approaches is the form of the likelihood. Therefore, we opt to treat data with the non-censored approach for all the analysis in the thesis.

## 5.6 Inference for Model 2

In this section we explain the inference for Model 2, via both MLE and Bayesian methodologies. Although we have already stated we are using the Bayesian paradigm as the basis of inference of the thesis, we need to make a comparison of both methodologies. For this purpose, we compare the estimates of both the MLE and the Bayesian frameworks, in simulations of observations of Model 2. As we will show, the MLE has some problems with some parameter regions, whereas the Bayesian estimation overcomes these difficulties.

Finally, in Section 5.6.3, we present a general simulation case, where Model 2 is fitted to simulated data of different models.

### 5.6.1 MLE Approach

The only purpose of the simulation of Model 2 is the comparison of estimates of MLE and the Bayesian paradigm. Therefore, it is not necessary to focus on thresholding data. Therefore, we simulate observations of Model 2 in the region:  $(1, \infty)^2$ . In other words, we are setting  $u_X = u_Y = 1$  for the definitions in Section 5.5. As a consequence, the truncation factor in (5.18) is given by:  $1 - F(u_X, u_Y) = 1$ .

Recall the likelihood function in (5.20). The MLE has to be evaluated using numerical methods. The standard procedure is to use a Newton-Raphson algorithm to maximize the log-likelihood  $l(\omega|\mathbf{D})$ . However, the MLE has been

found to present some regularity problems in some regions for a wide variety of models; see, for example, Tiago de Oliveira (1980) and Tawn (1988). Tawn (1988) illustrated that for the independent case in the logistic model, the MLE had nonregular behavior. For a wider discussion on inference problems when considering the MLE, we refer the reader to the papers of Coles and Tawn (1991) and Tawn (1990).

We simulated observations from Model 2, using the representation in (5.1). The cases are similar to the illustration of Section 4.2.2.3.

We consider four levels of tail dependence:  $\theta \in \{5, 3, 1, 1/3, 0.2\}$ . In terms of heaviness of the marginal tails, we split data into light tails:  $\xi_1 = 0.1$  and  $\xi_2 = 0.15$ ; and heavy tails:  $\xi_1 = 1$  and  $\xi_2 = 1.5$ . Regarding the amount of data, we simulated large samples of 500 observations for each model and small samples, with 70 observations, with 10 replications for each model. We chose 70 as a low amount of data in accordance with the real data analysis made in Section 7.2. In Table C.2 we show the estimations of the parameters of Model 2, as well as  $\chi$ , via the MLE method. For the 95% confidence intervals of  $\theta$ , we assumed asymptotic normality, similarly to that described in Section 4.2.2.3. For the cases of low and semi low dependence ( $\theta = 5$  and  $\theta = 3$ , respectively), we presented the left 95% confidence intervals for  $\chi$ . For the rest of the cases, centered 95% confidence intervals of  $\chi$  were given. We present, as well, the standard error of the replications.

As it can be seen, in the low dependence case, the point estimation is poor

for all the parameters. The specific case of the estimation of  $\theta$  could be fixed with a reparametrization. However, the left 95% confidence intervals of  $\chi$  are too wide in the small sampling, for example (0, 0.54) in the light tails case (compared to the Bayesian credibility intervals of the following section), which should not be affected by a reparametrization. For the large sampling and low dependence cases, although the point estimations are not accurate enough, the confidence intervals show a good precision. This fact, as well, might suggest a reparametrization in order to improve the point estimations.

For the rest of the cases, the point estimations and the centered confidence intervals seem accurate for any sampling size.

In order to test the point estimation of the MLE of this parametrization, we increased the sampling size to 1000 observations, only for the low dependence cases. We present the estimation in Figure C.3. The increase of the amount of data makes the MLE method produce more accurate estimation. This fact will be important for the Bayesian estimation as well, since, for large sample sizes, both methods are effectively equivalent.

We did not consider any reparametrization, since we found the Bayesian estimates worked well for all the cases, as we show in the following section.

### 5.6.2 Bayesian Estimation

Given the likelihood function in (5.20), we can express the posterior model as follows:

$$p_{\Omega}(\boldsymbol{\omega} | \mathbf{D}) \propto \pi(\boldsymbol{\omega}) \times \left( \frac{(\theta + 1)\alpha_1\alpha_2}{\theta\xi_1\xi_2} \right)^n \times (1 - F(u_X, u_Y))^{-n} \\ \prod_{i=1}^n x_i^{\frac{1}{\theta\xi_1}-1} y_i^{\frac{1}{\theta\xi_2}-1} \left( 1 + \alpha_1(x_i^{\frac{1}{\theta\xi_1}} - 1) + \alpha_2(y_i^{\frac{1}{\theta\xi_2}} - 1) \right)^{-\theta-2}; \quad (5.21)$$

where  $\pi(\boldsymbol{\omega})$  is the prior density distribution of the parameters, and where  $1 - F(u_X, u_Y)$  is given by (5.18).

The parameter space is  $\Omega = \mathbb{R}_+^5$  and all the parameters are assumed to be independent. We considered independent exponential distributions as priors.

We define the log-posterior as  $lp(\boldsymbol{\omega} | \mathbf{D}) = \log(p_{\Omega}(\boldsymbol{\omega} | \mathbf{D}))$ . We assume independent prior distributions and express them as  $\pi(\omega_j) = \delta_{\omega_j} e^{-\omega_j \delta_{\omega_j}}$ , where  $\omega_j$  is an element in  $\boldsymbol{\omega}$ . For example,  $\pi(\omega_{\theta})$  is the prior of the parameter  $\theta$ , with hyperparameter  $\delta_{\omega_{\theta}}$ .

Define the subset  $\overline{\omega_j}$  as the elements in  $\boldsymbol{\omega}$ , excluding  $\omega_j$ . Therefore, the conditional log-posterior distributions can be written as follows:

$$lp(\theta | \mathbf{D}, \overline{\omega_{\theta}}) = n \log \left( \frac{\theta + 1}{\theta} \right) + \frac{1}{\theta\xi_1} \sum \log(x_i) - \delta_{\theta}\theta \\ + \frac{1}{\theta\xi_2} \sum \log(y_i) + K_{\theta} - n \log(1 - F(u_X, u_Y)) \\ - (\theta + 2) \sum \log \left( 1 + \alpha_1(x_i^{\frac{1}{\theta\xi_1}} - 1) + \alpha_2(y_i^{\frac{1}{\theta\xi_2}} - 1) \right),$$

and

$$\begin{aligned}
 lp(\alpha_1 | \mathbf{D}, \overline{\omega_{\alpha_1}}) = & n \log(\alpha_1) - \delta_{\alpha_1} \alpha_1 - n \log(1 - F(u_X, u_Y)) + K_{\alpha_1} \\
 & - (\theta + 2) \sum \log \left( 1 + \alpha_1 (x_i^{\frac{1}{\theta \xi_1}} - 1) + \alpha_2 (y_i^{\frac{1}{\theta \xi_2}} - 1) \right),
 \end{aligned} \tag{5.22}$$

and

$$\begin{aligned}
 lp(\xi_1 | \mathbf{D}, \overline{\omega_{\xi_1}}) = & -n \log(\xi_1) + \frac{1}{\theta \xi_1} \sum \log(x_i) - \delta_{\xi_1} \xi_1 - n \log(1 - F(u_X, u_Y)) \\
 & - (\theta + 2) \sum \log \left( 1 + \alpha_1 (x_i^{\frac{1}{\theta \xi_1}} - 1) + \alpha_2 (y_i^{\frac{1}{\theta \xi_2}} - 1) \right) + K_{\xi_1},
 \end{aligned}$$

and

$$\begin{aligned}
 lp(\alpha_2 | \mathbf{D}, \overline{\omega_{\alpha_2}}) = & n \log(\alpha_2) - \delta_{\alpha_2} \alpha_2 - n \log(1 - F(u_X, u_Y)) + K_{\alpha_2} \\
 & - (\theta + 2) \sum \log \left( 1 + \alpha_1 (x_i^{\frac{1}{\theta \xi_1}} - 1) + \alpha_2 (y_i^{\frac{1}{\theta \xi_2}} - 1) \right),
 \end{aligned}$$

and

$$\begin{aligned}
 lp(\xi_2 | \mathbf{D}, \overline{\omega_{\xi_2}}) = & -n \log(\xi_2) + \frac{1}{\theta \xi_2} \sum \log(y_i) - \delta_{\xi_2} \xi_2 - n \log(1 - F(u_X, u_Y)) \\
 & - (\theta + 2) \sum \log \left( 1 + \alpha_1 (x_i^{\frac{1}{\theta \xi_1}} - 1) + \alpha_2 (y_i^{\frac{1}{\theta \xi_2}} - 1) \right) + K_{\xi_2},
 \end{aligned}$$

where the  $K_{\omega_j}$  term represents the logarithm of the normalizing factor of the conditional log-posterior of  $\omega_j$ .

We chose the Metropolis-Hastings within Gibbs sampler (see, for example, Tierney (1994)) method to draw samples. The proposal distribution we pick in each case was a log-normal distribution. The standard deviation was tuned such that the acceptance rate was in the interval  $[0.35, 0.40]$ , see for example Gelman *et al.* (2003).

The simulation data consisted of the same simulations as in the MLE case. Therefore, the thresholds are set as  $u_X = u_Y = 1$ .

We drew samples from the conditional log-posteriors defined in (5.22). For every sample of  $\omega$ , we sampled a prediction of  $X$  and  $Y$  with the representation in (5.1), and we denote the predictive sample as  $\mathbf{X}^S$ . The run sizes were 20,000.

In Figure D.5, we plotted the empirical means for different sizes of an MCMC sample of  $\theta$ . The first two plots show that, after 1000 samples, the means have stabilized. For the third and the fourth examples, the stabilization levels were around 3000 and 2000, respectively. Therefore, we chose these as burn-in levels. In general, the burn-in levels were not larger than 20% of the run size.

Let  $\omega^S$  be the MCMC sampling of  $\omega$ , given the data. We provide the parameter estimate with both the mean and the median of  $\omega^S$ . The results were very similar. However, we found that the median approach was fitting better to the simulation case when considering that the variables were generated by our model. Thus, we select onwards a median criteria for Bayes inference.

The general form of the table describing the Bayesian inference includes the estimation for all the parameter space and additionally, we give the central 95% credibility intervals for the parameter  $\theta$ , and for the tail dependence,  $\chi$ , except for the low dependence case, where the left 95% credibility interval is presented.

We chose unit exponential distributions for all the priors in an attempt to be non informative but not excessively, except for  $\theta$ . Since the main purpose of our model is the tail dependence structure, we will only discuss the selection

of the value of the hyperparameter  $\delta_\theta$ . Any prior belief on a bivariate model can be seen as a belief in the strength of the tail dependence. Therefore, we must evaluate how sensitive  $\chi$  is to the hyperparameter  $\delta_\theta$ , i.e. we need to compute the prior expected value of  $\chi$  and see how related they are:

$$\begin{aligned} E_{\pi_{\delta_\theta}}[\chi] &= \int 2^{-\theta} \pi_{\delta_\theta}(\delta_\theta) d\theta \\ &= \int_0^\infty 2^{-\theta} \delta_\theta e^{-\delta_\theta \theta} d\theta \\ &= \frac{\delta_\theta}{\delta_\theta + \log(2)}. \end{aligned} \tag{5.23}$$

This expression makes very simple the elicitation on the tail dependence structure. We set  $\delta_\theta = \log(2)$ , since this means  $E[\chi] = 0.5$ , i.e., our initial belief is that there exists a medium tail dependence structure. For the rest of the parameters, we chose unitary exponential prior distributions.

If we would like to consider a non-informative prior, we would need to keep the exponential as flat as possible. Then, we would need to consider the limit case in (5.23) when  $\delta_\theta \rightarrow \infty$ . When a flat prior is considered, the prior tail dependence structure can be expressed as:

$$\lim_{\delta_\theta \rightarrow \infty} E[2^{-\theta}] = 1.$$

Therefore, fixing a flat prior implies a belief of a complete tail dependence. This result can be related with the failure cases of the MLE: if a flat prior is chosen, the prior belief is a complete tail dependence. Therefore, it would need a great amount of data to get realistic estimation in the case of actual tail independence.

We show in Table C.5 the estimates of the simulation data. The improvement in estimation is evident for those cases where the MLE did not show great performance (i.e., the low dependence structure). The point estimations have an acceptable accuracy for all the dependence cases, and the credibility intervals for the low and semi low dependence cases for the small sample size are smaller than the confidence intervals of the MLE. The only issue to address is that in the semi low dependence cases, the estimates of  $\theta$  for the small sample is more accurate than the large sample. The reason for this is that the size of the replication is low. Besides, the credibility interval for  $\chi$  is more narrow in both cases, which reveals the improvement in the estimation when increasing the sample size.

Therefore, we can proceed to apply the Bayesian paradigm with more confidence for any size of data and any type of asymptotic dependence.

### 5.6.3 General Simulation Case

In this section, our interest relies on fitting our model with any kind of tails. Therefore, we simulated data coming from different types of GPD (recall that the GPD is the limit distribution of the excesses over a high threshold). The purpose of this case is to look at the responses that our model has to different types of tail dependence structures and different types of tail heaviness.

We drew 1000 bivariate samples, with 10 replications for each model, from the following dependence structures:

- Independently.
- Tail dependence: Clayton copula (recall 5.12) considering four levels of tail dependence:  $\phi \in \{1.2, 4/3, 2, 4, 6\}$  (which correspond to  $\theta \in \{5, 3, 1, 1/3, 0.2\}$ ).
- Tail independence: Gaussian copula with  $\rho = 0.5$ .

In terms of heaviness of the marginal GPD tails, we split data in light tails:  $\xi_1 = 0.1$  and  $\xi_2 = 0.15$ ; and heavy tails:  $\xi_1 = 1$  and  $\xi_2 = 1.5$ .

The Bayesian inference is carried out as described in Section 5.6.2.

Define the simulated variables as  $X^*$  and  $Y^*$ . Then, we only model the observations in the region:

$$\{(0, u_{X^*}), (0, u_{Y^*})\}^c,$$

where  $u_{X^*}$  and  $u_{Y^*}$  were chosen as the 95% quantile of  $X^*$  and  $Y^*$ , respectively.

We transformed the data as in (5.16):

$$X = X^* + 1, \text{ and } Y = Y^* + 1, \quad (5.24)$$

for  $(X^*, Y^*) \in \{(0, u_{X^*}), (0, u_{Y^*})\}^c$ .

Then, we drew MCMC simulations via the method explained in Section 5.6.2. We took a sample of 10,000 simulations with a burn-in of 500. The log-normal distribution was chosen as the proposal distribution.

The summary of the Bayesian inference is presented in Table C.6, which shows the following features:

- The exact independence case is estimated with a large  $\theta$ .
- Slight over estimation of  $\chi$  for low tail dependence cases. However,  $\chi$  is contained in the credibility intervals.
- Accurate estimation of  $\chi$  for medium and high tail dependence cases.
- Over estimation of  $\chi$  for tail independence case.

The accuracy of the estimations is not a big issue, since the amount of data in  $R_{00}^c$  was between 57 and 97 observations. The only issue that seems to be important is the over estimation of  $\chi$  in the tail independence case. However, as we stated before, we are not modelling other type of tail independence, rather than the exact independence.

However, as the main purpose of the model is the prediction, we study in Figures D.6 and D.7 how the model predicts in some scenarios.

The first scenario is the exact independence (heavy tailed). On the left side of Figure D.6 we plot the transformed data (recall (5.24)), whereas in the right side, the predictive sample, both in log-log scale. It can be seen the behavior of a lack of association in the tails, which is confirmed in the plots of both the empirical and predictive  $\chi(u)$ , in the first row of Figure D.7.

The following scenario is the tail independence case, with heavy tail. The figures show that the model tends to overestimate the dependence. The behavior of the sampling has almost as much association as the next scenario: tail dependence when  $\chi = 0.125$ . As we noted in this chapter, Model 2 does

not cover different types of tail independence. Therefore, we have to keep in mind that, for those cases, the prediction will tend to show a low dependence behavior.

Finally, the last two scenarios denote low and high dependence. In both of them, both the scatter plots and the plots of  $\chi(u)$  denote a low and a strong dependence among the samplings, respectively.

It is worth to notice that the parameters and predictive samplings are run at the same time. Therefore, avoiding a two-stage inference. This feature will be kept in the inference of the mixture model.

In summary, we presented in this section the form of a bivariate model, which is the base of the multivariate mixture model that we are presenting in the next section. The model has a simple dependence structure and is closed to marginalization. For the construction of the multivariate model, we will try to keep this features in higher dimensions. As well, the Bayesian paradigm turned out to be the better approach for inference, which is a desirable feature for the mixture models.

# Chapter 6

## The Multivariate Model

We present in this chapter a generalization of Model 2 to the multivariate setting. The way to generalize is not trivially achieved from (5.1). We anticipated that the proposed model is a mixture of models of different tail dependence structure. Inference with a mixture model can be seen either as a model selection or as a model averaging problem. This fact will play an important role in this chapter, since the objective differs substantially from one option to the other.

Section 6.1 presents some dependence definitions for the multivariate case. Section 6.2 deals with the trivariate model, its properties and inference. Similarly to Model 2, we establish a preference over one method of dealing with data. In Section 6.3 we introduce the general multivariate model. Section 6.4 presents the similarities of the multivariate model with some relevant models in the literature. Finally, Section 6.5 contains a discussion of the multivariate

models.

## 6.1 Multivariate Dependence

We need to define different types of dependence within a multivariate model. We treated extensively the parameter  $\chi$ , which measures the tail dependence between two variables in Chapters 2, 4, and 5. A multivariate expression of the extreme cases of  $\chi$  (i.e. considering the distribution of the tails, rather than the distribution of the maximum) was introduced by de Haan and Resnick (1977).

**Definition 6.1.1.** *Multivariate complete tail dependence and exact independence.*

Let  $\mathbf{S} = (S_1, \dots, S_d)^1$  be a vector of random variables with the same univariate marginal distribution.

Then, there is  $d$ -dimensional exact tail independence if for each subset  $\{j_1, \dots, j_r\}$  of  $\{1, \dots, d-1\}$ :

$$\lim_{s \rightarrow \infty} P(S_{j_1} > s, \dots, S_{j_r} > s | S_k > s) = 0. \quad (6.1)$$

for  $k \notin \{j_1, \dots, j_r\}$ .

---

<sup>1</sup>Note that the vector notation has changed from Section 2.2.1. The reason for this is that in the current chapter, there is no risk of confusion with other representations (as was the case for order statistics).

Similarly, there is  $d$ -dimensional complete tail dependence if:

$$\lim_{s \rightarrow \infty} P(S_2 > s, \dots, S_d > s | S_1 > s) = 1.$$

Note that the conditional could be on any variable, since they have the same univariate marginal distribution.

In the exact tail independence case, this means that any subset of  $\mathbf{S}$  is made up of exact tail independent random variables. On the other hand, in the complete tail dependence case, the  $d$  components must be complete tail dependent.

These definitions denote the complete tail dependence and exact independence cases. However, we can extend the complete tail dependence concept to define the tail dependence case.

**Definition 6.1.2.** *Multivariate tail dependence*

Let  $\mathbf{S} = (S_1, \dots, S_d)$  be a vector of random variables with the same univariate marginal distribution. Then, there is  $d$ -dimensional tail dependence, if for each subset  $\{j_1, \dots, j_r\}$  of  $\{2, \dots, d\}$ :

$$0 < \lim_{s \rightarrow \infty} P(S_{j_2} > s, \dots, S_{j_r} > s | S_{j_1} > s) \leq 1.$$

Hence, if there is  $d$ -dimensional tail dependence in  $\mathbf{S}$ , a vector of  $d$  variables, then there is  $(d - j)$ -dimensional tail dependence in any subset of  $\mathbf{S}$  of size  $(d - j)$ , for  $j = 1, \dots, (d - 1)$ .

However, we need to define the tail dependence in a different way. Assume we have  $\mathbf{S} = (S_1, \dots, S_d)$ , a vector of random variables. We are aiming to

model these variables with a joint distribution  $F$ . Then,  $F$  can fix any type of tail dependence on different disjoint subsets of  $\mathbf{S}$ . For example, half of them can be exact tail independent and the other half, tail dependent. Thus, we find it more convenient to define tail dependence in terms of the joint distribution  $F$ . We define this dependence as follows:

**Definition 6.1.3.** *M-tuple Tail Dependence*

Let  $\mathbf{S} = (S_1, \dots, S_d)$ , for  $d \geq 2$ , be a vector of random variables with the same marginal distribution and a joint distribution  $F$ . Then, the distribution  $F$  is said to have  $m$ -tuple tail dependence, if there exists an  $m$  ( $1 \leq m \leq d$ ), for which the following holds:

$$\sup_{k \leq d} \left\{ k : \forall \{j_1, \dots, j_k\} \in \{1, \dots, d\}, \right. \quad (6.2)$$

$$\left. 0 < \lim_{s \rightarrow \infty} P(S_{j_2} > s, \dots, S_{j_k} > s \mid S_{j_1} > s) \leq 1 \right\} = m.$$

Therefore, an  $m$ -tuple dependence of a model of  $d$  variables is the maximum number of tail dependent (including complete dependent) variables. In connection with the definition of de Haan and Resnick (1977),  $F_{\mathbf{S}}$  has  $m$ -tuple dependence if the maximum dimension of tail dependence of all the subsets of  $\mathbf{S}$  is  $m$ , for  $0 < m \leq d$ .

Recall that for any r.v.  $(S_1, \dots, S_r)$ :

$$P(S_1 > s, \dots, S_r > s) \leq P(S_1 > s, \dots, S_{k-1} > s, S_{k+1} > s, \dots, S_r > s),$$

for some  $k \in \{1, \dots, r\}$ .

Assume  $F$  has  $m$ -tuple tail dependence and that  $\{j_1, \dots, j_m\} \in \{1, \dots, d\}$

are tail dependent elements of  $\mathbf{S}$ . Then it is clear that  $F$  has a  $(m - k)$ -dimensional tail dependence, for  $k = 2, \dots, (m - 1)$ , i.e.

$$0 < \lim_{s \rightarrow \infty} P(S_{j^*_2} > s, \dots, S_{j^*_{m-k}} > s | S_{j^*_1} > s) \leq 1,$$

for  $\{j^*_1, \dots, j^*_{m-k}\}$ , any subset of size  $(m - k)$  of  $\{j_1, \dots, j_m\}$ .

The task of a multivariate model is to determine a joint model  $F$  which works for any combination of  $m$  tail dependent variables. Therefore, high flexibility for the model is required. As we will expose later, we propose a mixture model to achieve this flexibility.

Finally, we introduce some notation for multivariate tail dependence.

**Definition 6.1.4.** *Notation of Tail Dependence.*

Let  $\mathbf{S} = (S_1, \dots, S_d)$  be a vector of random variables with the same univariate marginal distribution. Then, for  $\{j_1, \dots, j_m\}$  a subset of  $\{1, \dots, d\}$ , define:

$$\chi_{\{j_1, \dots, j_m\}} = \lim_{s \rightarrow \infty} P(S_{j_2} > s, \dots, S_{j_m} > s | S_{j_1} > s).$$

And in the case when  $m = d$ , we can simplify the notation as:

$$\chi_{\{j_1, \dots, j_d\}} = \chi_d.$$

As well, for  $\{j^*_1, \dots, j^*_k\}$ , a subset of  $\{1, \dots, d\}$  and a disjoint set of  $\{j_1, \dots, j_m\}$ , define:

$$\chi_{\{[j_1, \dots, j_m], [j^*_1, \dots, j^*_k]\}} = \lim_{s \rightarrow \infty} P(S_{j_1} > s, \dots, S_{j_m} > s | S_{j^*_1} > s, \dots, S_{j^*_k} > s).$$

Note that  $\chi_2 = \chi$  for the bivariate case.

The tail dependence definitions that we introduced in this section, will help to construct our generalization of Model 2. In order to ease the explanation of the multivariate framework, we firstly present the trivariate model.

## 6.2 Trivariate Model

Here, we present a trivariate model, which helps on settling the way to generalize Model 2 to the multivariate setting.

### 6.2.1 Trivariate Model Motivation

Let  $\mathbf{X} = (X_1, X_2, X_3)$  be a vector of random variables. Similarly to Model 2, we are aiming to find a model for regions of the form:

$$R = \{(1, u_{X_1}) \times (1, u_{X_2}) \times (1, u_{X_3})\}^c,$$

for  $u_{X_j}$ , a high threshold of  $X_j$ , for  $j = 1, 2, 3$ .

A desirable feature of a multivariate model is the recovery of the bivariate structure of Model 2, when marginalizing all the variables, save two. Therefore, as a first attempt of representation, and prompted by (5.1), we consider the representation of these variables as follows:

$$\begin{aligned} X_1 &= (1 + B_1/A_1)^{\theta_1 \xi_1}; \\ X_2 &= (1 + B_2/A_2)^{\theta_2 \xi_2}; \\ X_3 &= (1 + B_3/A_3)^{\theta_3 \xi_3}; \end{aligned} \tag{6.3}$$

In Chapter 5, we saw that the subvariable  $B$  represented the scale of each variable, whereas the parameter  $\xi$  represented the marginal tail. Therefore, it is reasonable to select them as: independent  $B_i \sim \text{Exp}(\alpha_i)$ ,  $\theta, \alpha_1, \xi_i > 0$ , i.e. they belong exclusively to  $X_i$ , for  $i = 1, 2, 3$ .

The subvariable  $A$  though, must be modified, since it represents the dependence among the variables. One requirement is that  $A_k = A_j$ , when  $X_k$  and  $X_j$  are tail dependent. Let  $A_j \sim \text{Ga}(\theta_j, 1)$  with joint density  $f_{\mathbf{A}}$ , for  $j = 1, 2, 3$  and where  $\mathbf{A} = (A_1, A_2, A_3)$ .

In representation (6.3), the marginal distributions are equal to those of Model 2, i.e. they are quasi Burr-XII type, and are as follows:

$$F_{X_j}(x) = 1 - \left(1 + \alpha_j \left(x^{\frac{1}{\theta_j \xi_j}} - 1\right)\right)^{-\theta_j}, \quad (6.4)$$

with  $x > 1$ , and for  $j = 1, 2, 3$ .

The first issue to study is the bivariate structure. Recall that in Model 2, we assumed that theoretically, there exists tail dependence between two variables, and that this dependence was measured by a single parameter  $\theta$ . In order to get that representation, we needed to set a common subvariable  $A$  for both variables. This feature will help us to develop the form of the joint distribution. However, we firstly need to clarify a tail dependence issue in the generalization.

When working with only two variables, and when the main purpose of the study is the structure of the tails, it is convenient for us to model the variables as tail dependent when  $\theta$  is finite (see Section 5.3). However, in

multivariate analysis with dimension larger than 2, it is helpful to differentiate the tail dependent and the exact tail independent variables, and to represent the exact tail independent variables in the joint model directly as the product of their marginal distributions.

In order to study the bivariate structure, let  $X_1$  and  $X_2$  be the variables of interest. We can define the joint model of  $A_1$  and  $A_2$  if we marginalize with respect to  $A_3$ :

$$f_{A_1, A_2}(a_1, a_2) = \int_0^\infty f_{\mathbf{A}}(a_1, a_2, a_3) da_3.$$

We propose to model it as:

$$f_{A_1, A_2}(a_1, a_2) = w_1 \frac{e^{-a}}{\Gamma(\theta)} a^{\theta-1} \mathbf{1}_{[a_1=a_2=a]} + w_2 \frac{e^{-(a_1+a_2)}}{\Gamma(\theta_1)\Gamma(\theta_2)} a_1^{\theta_1-1} a_2^{\theta_2-1} \mathbf{1}_{[a_1 \neq a_2]}, \quad (6.5)$$

where  $0 \leq w_1 \leq 1$ ,  $0 \leq w_2 \leq 1$ , and  $w_1 + w_2 = 1$ . Therefore, the joint survival

distribution,  $\bar{F}^*$  of the variables is:

$$\begin{aligned}
\bar{F}_{X_1, X_2}^*(x_1, x_2) &= P(X_1 > x_1, X_2 > x_2) \\
&= P\left(B_1 > \left(x_1^{\frac{1}{\theta_1 \xi_1}} - 1\right)A_1, B_2 > \left(x_2^{\frac{1}{\theta_2 \xi_2}} - 1\right)A_2\right) \\
&= \iint_0^\infty P\left(B_1 > \left(x_1^{\frac{1}{\theta_1 \xi_1}} - 1\right)a_1, B_2 > \left(x_2^{\frac{1}{\theta_2 \xi_2}} - 1\right)a_2 \mid a_1, a_2\right) \\
&\quad \times f_{A_1, A_2}(a_1, a_2) da_1 da_2 \\
&= \frac{w_1}{\Gamma(\theta)} \int_0^\infty e^{-a \left[1 + \alpha_1 \left(x_1^{\frac{1}{\theta_1 \xi_1}} - 1\right) + \alpha_2 \left(x_2^{\frac{1}{\theta_2 \xi_2}} - 1\right)\right]} a^{\theta-1} da \\
&\quad + \frac{w_2}{\Gamma(\theta_1)\Gamma(\theta_2)} \int_0^\infty \int_0^\infty e^{-a_1 \left[1 + \alpha_1 \left(x_1^{\frac{1}{\theta_1 \xi_1}} - 1\right)\right]} a_1^{\theta_1-1} \\
&\quad \times e^{-a_2 \left[1 + \alpha_2 \left(x_2^{\frac{1}{\theta_2 \xi_2}} - 1\right)\right]} a_2^{\theta_2-1} da_1 da_2 \\
&= w_1 \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta_1 \xi_1}} - 1\right) + \alpha_2 \left(x_2^{\frac{1}{\theta_2 \xi_2}} - 1\right)\right)^{-\theta} \\
&\quad + w_2 \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta_1 \xi_1}} - 1\right)\right)^{-\theta_1} \left(1 + \alpha_2 \left(x_2^{\frac{1}{\theta_2 \xi_2}} - 1\right)\right)^{-\theta_2}, \tag{6.6}
\end{aligned}$$

for  $x_1, x_2 > 1$ .

Thus, the joint survival function is a mixture of Model 2 and the product of the survival marginals as in (6.4). However, the model in (6.6) produces the following univariate survival marginal:

$$\begin{aligned}
\bar{F}_{X_1}^*(x_1) &= \bar{F}_{X_1, X_2}^*(x_1, 1) \\
&= w_1 \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta_1 \xi_1}} - 1\right)\right)^{-\theta} + w_2 \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta_1 \xi_1}} - 1\right)\right)^{-\theta_1}, \tag{6.7}
\end{aligned}$$

and similarly for  $\bar{F}_{X_2}^*$ . This expression differs to that one in (6.4), unless  $\theta = \theta_1$ . Hence, in order to maintain dimensional coherence, we propose the

joint density of  $\mathbf{A}$  as:

$$f_{A_1, A_2}(a_1, a_2) = w_1 \frac{e^{-a}}{\Gamma(\theta)} a^{\theta-1} \mathbb{1}_{[a_1=a_2=a]} + w_2 \frac{e^{-(a_1+a_2)}}{\Gamma(\theta)^2} a_1^{\theta-1} a_2^{\theta-1} \mathbb{1}_{[a_1 \neq a_2]}, \quad (6.8)$$

where  $0 \leq w_1 \leq 1$ ,  $0 \leq w_2 \leq 1$ , and  $w_1 + w_2 = 1$ . In this way, the joint survival distribution,  $\bar{F}^*$  is expressed as:

$$\begin{aligned} \bar{F}_{X_1, X_2}^*(x_1, x_2) &= w_1 \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta \xi_1}} - 1 \right) + \alpha_2 \left( x_2^{\frac{1}{\theta \xi_2}} - 1 \right) \right)^{-\theta} \\ &\quad + w_2 \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta \xi_1}} - 1 \right) \right)^{-\theta} \left( 1 + \alpha_2 \left( x_2^{\frac{1}{\theta \xi_2}} - 1 \right) \right)^{-\theta}. \end{aligned} \quad (6.9)$$

In this way, the survival marginals are as in (6.4), since the marginals of (6.8) are gamma densities.

Although we have set a unique  $\theta$  for the marginals and for the dependence structure in model (6.9), the nature of the mixture model has not changed: it is a mixture of a tail dependence and an exact tail independence. Even more, as we saw in expression (5.3),  $\theta$  does not affect the marginal tail.

From now on, we will interpret the expression (6.9), as an average of models. This is a standard topic in a Bayesian framework, see for example Leamer (1978) and Raftery *et al.* (1995). In this type of model, there are two main options: treat it as a model selection problem or as a model averaging problem. For both methodologies, it is necessary to estimate the weights  $w_j$ . However, in model selection, only the model with the largest value of weight  $w_j$  is selected. In contrast, the model averaging takes into account the uncertainty in all the models, so that no model is discarded.

The methodology that we develop in Section 6.2.3 will allow us to select either case. However, we opt to follow the criteria:

- Treat the problem with model selection framework if the aim is to select a type of tail dependence.
- Treat the problem with model averaging framework in any other case, for example, if the aim is prediction.

It is clear that, if we opt for the model selection framework, the tail dependence structure is either exact tail independence or tail dependence, according to the selected model.

For the case of model averaging, we need to extend on the computation of the tail dependence structure, which for model (6.9) is shown in the following theorem.

**Theorem 6.2.1.** *Tail dependence structure of  $\bar{F}^*$*

*Let  $X_1$  and  $X_2$  be random variables with representation as in (6.3), and for which the joint density of the subvariables  $A_1$  and  $A_2$  is given by (6.8), i.e.  $X_1$  and  $X_2$  have (6.9) as joint survival function. Then, the tail dependence parameter  $\chi$ , denote for  $\bar{F}^*$  as  $\chi^*$ , is given by:*

$$\chi^* = w_1 2^{-\theta}. \quad (6.10)$$

We present the proof in Appendix B.8.

This result implies the following tail dependence structure for  $X_1$  and  $X_2$ :

- Exact tail independence if  $\theta \rightarrow \infty$ .
- Exact tail independence if  $w_1 = 0$ .
- Tail dependence in any other case.

The mixture model, seen as model selection, implies a model identifiability issue, since:

$$\lim_{\theta \rightarrow \infty} \bar{F}_{X_1, X_2}^*(x_1, x_2) = w_1 x_1^{-\frac{\alpha_1}{\xi_1}} x_2^{-\frac{\alpha_2}{\xi_2}} + w_2 x_1^{-\frac{\alpha_1}{\xi_1}} x_2^{-\frac{\alpha_2}{\xi_2}}.$$

Therefore, for large  $\theta$ , both models are equivalent. This is why tail independence is achieved when  $\theta \rightarrow \infty$  or when  $w_1 = 0$ . A practical way to overcome this issue is to truncate  $\theta$  to a large value, for example  $\theta \in (0, 5)$ . In this way,  $w_1$  should be expected to be close to 0 for the independence case. However, we choose to keep  $\theta > 0$  to study the theoretical properties of the representation.

For inference purposes, we find it convenient to restructure the marginal bivariate survival function as follows:

$$\begin{aligned} \bar{F}_{X_1, X_2}(x_1, x_2) &= w_1 \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta \xi_1}} - 1 \right) + \alpha_2 \left( x_2^{\frac{1}{\theta \xi_2}} - 1 \right) \right)^{-\theta} \\ &\quad + w_2 x_1^{-\frac{\alpha_1}{\xi_1}} x_2^{-\frac{\alpha_2}{\xi_2}}. \end{aligned} \tag{6.11}$$

for  $x_1, x_2 > 1$ .

Hence, we are fixing the exact tail independent model as the product of exact independent marginals when  $\theta \rightarrow \infty$ .

This representation drops the dimensional coherence. However, we do not find this feature as critical for inference. The purpose of the introduction of this

representation is the ease of inference. Therefore, expressing the independent component without the presence of  $\theta$  will facilitate the inference algorithm. As well, the exclusion of  $\theta$  in the independent component does not result in a severe inconvenience, since  $\theta$  is only modelling the dependence. Hence, it is reasonable to leave only the product of marginal tails in this component.

The correspondent univariate marginal survival function of  $X_1$  is given by:

$$\bar{F}_{X_1}(x) = \bar{F}_{X_1, X_2}(x_1, 1) = w_1 \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta \xi_1}} - 1 \right) \right)^{-\theta} + w_2 x_1^{-\frac{\alpha_1}{\xi_1}},$$

and similarly for  $\bar{F}_{X_2}$ .

Although we are dropping the dimensional coherence of model (6.9), the choice of model (6.11) will facilitate the inference.

For the model average choice, we have that the tail dependence is the same as in Theorem 6.2.1. This means, if  $\chi$  is the tail dependence parameter associated to the model in (6.11), then:

$$\chi = 2^{-\theta}. \tag{6.12}$$

The bivariate marginal model implied by (6.11) will give the motivation for the trivariate model: a mixture model of different tail dependence components. Thus, we need to define the different types of tail dependence components involved in a trivariate analysis.

## 6.2.2 Trivariate structure

Motivated by the bivariate mixture model in (6.11), we aim to construct the trivariate generalization via a mixture model. First of all, we need to introduce some notation.

Define the following latent variables, regarding tail dependence:

- $TI$ , if  $F_{\mathbf{X}}$  has triple exact tail independence.
- $DD_{j,k}(\theta)$ , if  $F_{\mathbf{X}}$  has double tail dependence, and  $X_j$  and  $X_k$  are tail dependent, for  $j \neq k$  and  $j, k \in \{1, 2, 3\}$ .
- $TD(\theta)$ , if  $F_{\mathbf{X}}$  has triple tail dependence.

Therefore, the general trivariate model we propose has the following form:

$$\bar{F}_{\mathbf{X}} = w_1 \bar{F}_{\mathbf{X}|TI} + w_2 \bar{F}_{\mathbf{X}|DD_{1,2}} + w_3 \bar{F}_{\mathbf{X}|DD_{1,3}} + w_4 \bar{F}_{\mathbf{X}|DD_{2,3}} + w_5 \bar{F}_{\mathbf{X}|TD}, \quad (6.13)$$

with  $0 \leq w_j \leq 1$ , for  $j = 1, \dots, 5$  and  $\sum_{k=1}^5 w_k = 1$ .

As in the previous section, we construct a model with dimensional coherence.

The next step is to characterize each of the mixture components in (6.13). The first one is the simplest: the model with triple exact independence. The joint distribution is the product of the marginals as in the second factor in

(6.9), as we now write down:

$$\begin{aligned}\bar{F}_{\mathbf{X}|TI}(x_1, x_2, x_3) &= P(X_1 > x_1, X_2 > x_2, X_3 > x_3 | TI) \\ &= \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta \xi_1}} - 1\right)\right)^{-\theta} \left(1 + \alpha_2 \left(x_2^{\frac{1}{\theta \xi_2}} - 1\right)\right)^{-\theta} \\ &\quad \times \left(1 + \alpha_3 \left(x_3^{\frac{1}{\theta \xi_3}} - 1\right)\right)^{-\theta}.\end{aligned}\quad (6.14)$$

The second component to develop is the generic model with double tail dependence:  $DD_{j,k}(\theta)$ , for  $j \neq k$ , and  $j, k \in \{1, 2, 3\}$ . This means  $X_j$  and  $X_k$  are tail dependent, and  $X_r$  is exact tail independent. For example, let  $DD_{1,2}(\theta)$  be the model to illustrate. Then, motivated by (6.9), the form of this component is as follows:

$$\begin{aligned}\bar{F}_{\mathbf{X}|DD_{1,2}}(x_1, x_2, x_3) &= \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta \xi_1}} - 1\right) + \alpha_2 \left(x_2^{\frac{1}{\theta \xi_2}} - 1\right)\right)^{-\theta} \\ &\quad \times \left(1 + \alpha_3 \left(x_3^{\frac{1}{\theta \xi_3}} - 1\right)\right)^{-\theta}.\end{aligned}\quad (6.15)$$

Finally, for the last component in the mixture model, the model with triple tail dependence, it is clear that we need to set a common  $\theta$  for  $X_1$ ,  $X_2$ , and  $X_3$ . This means:

$$f_{\mathbf{A}|TD}(a_1, a_2, a_3) = \frac{1}{\Gamma(\theta)} e^{-a} a^{\theta-1} \mathbb{1}_{[a_1=a_2=a_3=a]}.$$

Thus, the joint survival of the model with triple dependence is expressed as:

$$\begin{aligned}
\bar{F}_{\mathbf{X}|TD}(x_1, x_2, x_3) &= P(X_1 > x_1, X_2 > x_2, X_3 > x_3 | TD) \\
&= P\left(B_1 > \left(x_1^{\frac{1}{\theta\xi_1}} - 1\right)A_1, B_2 > \left(x_2^{\frac{1}{\theta\xi_2}} - 1\right)A_2, \right. \\
&\quad \left. B_3 > \left(x_3^{\frac{1}{\theta\xi_3}} - 1\right)A_3\right) \\
&= \int \int \int_0^\infty P\left(B_1 > \left(x_1^{\frac{1}{\theta\xi_1}} - 1\right)a_1, B_2 > \left(x_2^{\frac{1}{\theta\xi_2}} - 1\right)a_2, \right. \\
&\quad \left. B_3 > \left(x_3^{\frac{1}{\theta\xi_3}} - 1\right)a_3 \mid a_1\right) f_{\mathbf{A}|TD}(a_1, a_2, a_3) da_1 da_2 da_3 \\
&= \frac{1}{\Gamma(\theta)} \int_0^\infty e^{-a \left[1 + \alpha_1 \left(x_1^{\frac{1}{\theta\xi_1}} - 1\right) + \alpha_2 \left(x_2^{\frac{1}{\theta\xi_2}} - 1\right) + \alpha_3 \left(x_3^{\frac{1}{\theta\xi_3}} - 1\right)\right]} a^{\theta-1} da \\
&= \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta\xi_1}} - 1\right) + \alpha_2 \left(x_2^{\frac{1}{\theta\xi_2}} - 1\right) + \alpha_3 \left(x_3^{\frac{1}{\theta\xi_3}} - 1\right)\right)^{-\theta}.
\end{aligned} \tag{6.16}$$

Suppose the model selection approach is chosen. Then, the asymptotics of the triple dependence model are given by:

$$P(S_1 > s, S_2 > s, S_3 > s | TD) \approx 3^{-\theta} s^{-1}, \text{ as } s \rightarrow \infty,$$

where  $S_j$  is defined as

$$S_j = -\frac{1}{\log F_{X_j|TD}(X_j)}, \tag{6.17}$$

for  $j = 1, 2, 3$ . Therefore, from Definition 6.1.4, the tail dependence parameter  $\chi$  for the triple dependence model is computed as:

$$\chi_{3|TD} = 3^{-\theta}, \tag{6.18}$$

and

$$\chi_{\{1,2\}|TD} = \chi_{\{1,3\}|TD} = \chi_{\{2,3\}|TD} = 2^{-\theta},$$

and

$$\chi_{\{[3],[1,2]\} | TD} = \lim_{s \rightarrow \infty} P(S_3 > s | S_1 > s, S_2 > s, TD) = \left(\frac{2}{3}\right)^\theta,$$

which is the same for  $\chi_{\{[2],[1,3]\} | TD}$  and  $\chi_{\{[1],[2,3]\} | TD}$ .

If we combine (6.14), (6.15), and (6.16), we can introduce the trivariate model with dimensional coherence, setting  $\mathbf{x} = (x_1, x_2, x_3)$ , as:

$$\begin{aligned} \bar{F}_{\mathbf{x}}^*(\mathbf{x}) = & w_1 \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta \xi_1}} - 1\right)\right)^{-\theta} \left(1 + \alpha_2 \left(x_2^{\frac{1}{\theta \xi_2}} - 1\right)\right)^{-\theta} \left(1 + \alpha_3 \left(x_3^{\frac{1}{\theta \xi_3}} - 1\right)\right)^{-\theta} \\ & + w_2 \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta \xi_1}} - 1\right) + \alpha_2 \left(x_2^{\frac{1}{\theta \xi_2}} - 1\right)\right)^{-\theta} \left(1 + \alpha_3 \left(x_3^{\frac{1}{\theta \xi_3}} - 1\right)\right)^{-\theta} \\ & + w_3 \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta \xi_1}} - 1\right) + \alpha_3 \left(x_3^{\frac{1}{\theta \xi_3}} - 1\right)\right)^{-\theta} \left(1 + \alpha_2 \left(x_2^{\frac{1}{\theta \xi_2}} - 1\right)\right)^{-\theta} \\ & + w_4 \left(1 + \alpha_2 \left(x_2^{\frac{1}{\theta \xi_2}} - 1\right) + \alpha_3 \left(x_3^{\frac{1}{\theta \xi_3}} - 1\right)\right)^{-\theta} \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta \xi_1}} - 1\right)\right)^{-\theta} \\ & + w_5 \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta \xi_1}} - 1\right) + \alpha_2 \left(x_2^{\frac{1}{\theta \xi_2}} - 1\right) + \alpha_3 \left(x_3^{\frac{1}{\theta \xi_3}} - 1\right)\right)^{-\theta}, \end{aligned} \quad (6.19)$$

with  $0 \leq w_k \leq 1$ , for  $k = 1, \dots, 5$  and  $\sum_{k=1}^5 w_k = 1$ .

The model in (6.19) is achieved by the representation in (6.3), with:

$$\begin{aligned} f_{\mathbf{A}}(a_1, a_2, a_3) = & w_1 \left(\frac{1}{\Gamma(\theta)^3}\right) e^{-(a_1+a_2+a_3)} (a_1 a_2 a_3)^{\theta-1} \\ & + w_2 \left(\frac{1}{\Gamma(\theta)^2}\right) e^{-a} a^{\theta-1} \mathbf{1}_{[a_1=a_2=a]} (e^{-a_3} a_3^{\theta-1}) \\ & + w_3 \left(\frac{1}{\Gamma(\theta)^2}\right) e^{-a} a^{\theta-1} \mathbf{1}_{[a_1=a_3=a]} (e^{-a_2} a_2^{\theta-1}) \\ & + w_4 \left(\frac{1}{\Gamma(\theta)^2}\right) e^{-a} a^{\theta-1} \mathbf{1}_{[a_2=a_3=a]} (e^{-a_1} a_1^{\theta-1}) \\ & + w_5 \left(\frac{1}{\Gamma(\theta)}\right) e^{-a} a^{\theta-1} \mathbf{1}_{[a_1=a_2=a_3=a]}. \end{aligned}$$

In order to show the dimensional coherence of model (6.19), we now show the

form of a bivariate marginal function:

$$\begin{aligned}
 \overline{F}_{X_1, X_2}^*(x_1, x_2) &= \overline{F}_{\mathbf{X}}^*(x_1, x_2, 1) \\
 &= w_* \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta \xi_1}} - 1\right)\right)^{-\theta} \left(1 + \alpha_2 \left(x_2^{\frac{1}{\theta \xi_2}} - 1\right)\right)^{-\theta} \\
 &\quad + (1 - w_*) \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta \xi_1}} - 1\right) + \alpha_2 \left(x_2^{\frac{1}{\theta \xi_2}} - 1\right)\right)^{-\theta},
 \end{aligned} \tag{6.20}$$

where  $w_* = w_1 + w_3 + w_4$ . This representation is equivalent to that in (6.9), for which we already know it has dimensional coherence.

For the Model Selection approach, the tail dependence structure depends on the selection of the model. However, for the Model Averaging approach, we present the dependence structure in the following theorem.

**Theorem 6.2.2.** *Trivariate tail dependence structure*

Let  $\mathbf{X} = (X_1, X_2, X_3)$  be a random vector with joint survival function as in (6.19). Then, from Definition 6.1.4 the following holds:

- The tail dependence parameter  $\chi_3$  of the model is given by:

$$\chi_3^* = w_5 3^{-\theta}. \tag{6.21}$$

- The pairwise tail dependencies are given by:

$$\begin{aligned}
 \chi_{\{1,2\}}^* &= (w_2 + w_5) 2^{-\theta}, \\
 \chi_{\{1,3\}}^* &= (w_3 + w_5) 2^{-\theta}, \\
 \chi_{\{2,3\}}^* &= (w_4 + w_5) 2^{-\theta}.
 \end{aligned} \tag{6.22}$$

- *And, finally:*

$$\begin{aligned}
 \chi_{\{[1],[2,3]\}}^* &= \left( \frac{w_5}{w_4 + w_5} \right) \left( \frac{2}{3} \right)^\theta, \\
 \chi_{\{[2],[1,3]\}}^* &= \left( \frac{w_5}{w_3 + w_5} \right) \left( \frac{2}{3} \right)^\theta, \\
 \chi_{\{[3],[1,2]\}}^* &= \left( \frac{w_5}{w_2 + w_5} \right) \left( \frac{2}{3} \right)^\theta.
 \end{aligned} \tag{6.23}$$

The proof can be found in Appendix B.9. This theorem gives the following tail dependence structure for  $\bar{F}_{\mathbf{X}}^*$ :

- $\bar{F}_{\mathbf{X}}^*$  has triple tail dependence if  $\theta < \infty$  and  $w_5 \neq 0$ .
- $\bar{F}_{\mathbf{X}}^*$  has, at most, double tail dependence in any other case.

The trivariate model introduced in (6.19) has advantageous properties. From (6.22) we can notice that, although  $\theta$  is common to all the individual models with tail dependence, the strength of tail dependence varies from pair to pair. Other advantages are dimensional coherence and easy of computation of the tail dependence parameters for the Model Averaging approach. However, the inference methodology of Section 6.2.3 is more convenient to be applied

with a modified version of (6.19), motivated by (6.11), which is expressed as:

$$\begin{aligned}
 \bar{F}_{\mathbf{X}}(\mathbf{x}) = & w_1 \left( x_1^{-\frac{\alpha_1}{\xi_1}} x_2^{-\frac{\alpha_2}{\xi_2}} x_3^{-\frac{\alpha_3}{\xi_3}} \right) \\
 & + w_2 \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta\xi_1}} - 1 \right) + \alpha_2 \left( x_2^{\frac{1}{\theta\xi_2}} - 1 \right) \right)^{-\theta} x_3^{-\frac{\alpha_3}{\xi_3}} \\
 & + w_3 \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta\xi_1}} - 1 \right) + \alpha_3 \left( x_3^{\frac{1}{\theta\xi_3}} - 1 \right) \right)^{-\theta} x_2^{-\frac{\alpha_2}{\xi_2}} \\
 & + w_4 \left( 1 + \alpha_2 \left( x_2^{\frac{1}{\theta\xi_2}} - 1 \right) + \alpha_3 \left( x_3^{\frac{1}{\theta\xi_3}} - 1 \right) \right)^{-\theta} x_1^{-\frac{\alpha_1}{\xi_1}} \\
 & + w_5 \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta\xi_1}} - 1 \right) + \alpha_2 \left( x_2^{\frac{1}{\theta\xi_2}} - 1 \right) + \alpha_3 \left( x_3^{\frac{1}{\theta\xi_3}} - 1 \right) \right)^{-\theta},
 \end{aligned} \tag{6.24}$$

with  $0 \leq w_k \leq 1$ , for  $k = 1, \dots, 5$  and  $\sum_{k=1}^5 w_k = 1$ .

Similarly to what happened to the model (6.11), we drop dimensional coherence, and the tail dependence asymptotics are just approximations of those of model (6.19). However, we find model (6.19) more convenient for inference purposes, although we pay the price of dropping the features of dimensional coherence and the ease of tail dependence structure computation for the Model Averaging approach.

Therefore, for  $F_{\mathbf{X}}$  defined as in (6.24), we approximate the tail dependence parameters with the results in Theorem 6.2.2.

As it occurred with model (6.19), the approximation is motivated by the fact that representing any exact independent factor as  $\left( 1 + \alpha_j \left( x_j^{\frac{1}{\theta\xi_j}} - 1 \right) \right)^{-\theta}$ , or as  $\left( x_j^{-\frac{\alpha_j}{\xi_j}} \right)$ , should not affect the dependence structure among the variables, for  $j \in \{1, 2, 3\}$ .

In the next section, we deal with the inference of the trivariate model,

which will use model (6.24) as basis.

### 6.2.3 Trivariate Inference

Recall the model in (6.24) is defined by five components or models. We denote the models, recalling the latent variables in (6.13), as follows:

- $M_1 = F_{\mathbf{X}|TI}$ .
- $M_2 = F_{\mathbf{X}|DD_{1,2}}$ .
- $M_3 = F_{\mathbf{X}|DD_{1,3}}$ .
- $M_4 = F_{\mathbf{X}|DD_{2,3}}$ .
- $M_5 = F_{\mathbf{X}|TD}$ .

Define the parameter space of model  $M_k$  as  $\Omega_k$ , for  $k = 1, \dots, 5$ , then define the following parameters:

- $\omega'_1 = \{\alpha_1, \alpha_2, \alpha_3, \xi_1, \xi_2, \xi_3\} \in \Omega'_1 = \mathbb{R}_+^6$ .
- $\omega_2 = \omega_3 = \omega_4 = \omega_5 = \{\theta, \alpha_1, \alpha_2, \alpha_3, \xi_1, \xi_2, \xi_3\} \in \Omega_2 = \mathbb{R}_+^7$ .

Therefore,  $\Omega_2 = \Omega_3 = \Omega_4 = \Omega_5$ .

#### 6.2.3.1 General Details of Reversible Jump

The way we are making inference is via the Reversible Jump MCMC method, which was described in Section 3.3. A jump from model  $M_k$  to model  $M_{k^*}$

depends on the posterior distributions for each model, and the distribution of an auxiliary r.v.  $u$ , which deals with the change in dimensions from  $\Omega_k$  to  $\Omega_{k^*}$  (see expression (3.7)).

For jumps involving only  $k = 2, \dots, 5$ , there is no change of dimension. Therefore, if a jump from model  $k$  to model  $k^*$  is proposed, then:

$$\omega_{k^*} = g_{k,k^*}(\omega_k) = \omega_k,$$

for  $k$  and  $k^*$  in  $\{2, \dots, 5\}$ .

Although  $\Omega'_1$  has a different dimension, we can add  $\theta$  to modify the parameter space, but not the posterior of  $M_1$ , as follows:

$$\omega_1 = (\omega'_1, \theta),$$

and then, we can define the change of dimension similarly to the previous cases, i.e.:

$$\omega_1 = g_{k,1}(\omega_k) = \omega_k, \text{ and } \omega_k = g_{1,k}(\omega_1) = \omega_1,$$

for  $k = 2, \dots, 5$ . Thus, the modified parameter space of  $M_1$  would be  $\Omega_1 = \mathbb{R}^7$ .

This means:

$$\Omega = \Omega_1 = \dots = \Omega_5 = \mathbb{R}_+^7,$$

and then, we define the unique parameter vector as:

$$\omega = \{\theta, \alpha_1, \alpha_2, \alpha_3, \xi_1, \xi_2, \xi_3\} \in \Omega.$$

Hence, with this modification, the deterministic functions  $g_{k,k^*}$  are equal to the identity, so there is no need of the auxiliary r.v. to cope with the change

of parameter space, and:

$$\left| \frac{\nabla g_{k,k^*}(\omega_k)}{\nabla \omega_k} \right| = 1.$$

If we propose model  $M_{k^*}$  with probability  $1/4$  for  $k^* = 1, \dots, 5$ , then

$$J_{k^*,k} = J_{k,k^*}.$$

Therefore, the jump ratio defined in (3.7) only depends on the ratio of the posterior (or log-posterior) distributions.

Recall from Section (3.3) that the Reversible Jump MCMC consists of two steps: a jump and a single sample from the model in turn. We use the Metropolis-Hastings within Gibbs sampler method to draw samples, as in Model 2 (recall Section 3.2.4). Therefore, we need two pieces of information for each model: the log-posterior distributions for the jump part and the conditional log-posterior distributions for the sampling.

Prior to describing the distributions of each model, we need to establish the form of the prior distributions. Similarly to Model 2, in Section 5.6.2, we define the priors as independent exponential distributions.

For the  $M_5$  case, similarly to the bivariate model (see (5.23)), we have that:

$$\begin{aligned} E_{\pi_{\delta_\theta}}[\chi_3] &= \int 3^{-\theta} \pi_{\delta_\theta | M_5}(\delta_\theta) d\theta \\ &= \int_0^\infty 3^{-\theta} \delta_\theta e^{-\delta_\theta \theta} d\theta \\ &= \frac{\delta_\theta}{\delta_\theta + \log(3)}. \end{aligned}$$

Then, if we set our initial belief as  $E[\chi_3] = 0.5$ , then  $\delta_\theta = \log(3)$ .

For the double tail dependent models, we set  $\delta_\theta = \log(2)$ , and for the rest of the parameters, we set the priors as unit exponential distributions.

Therefore, the priors are expressed as:

$$\pi(\theta) = \log(3) \exp \{-\log(3) \theta\}, \text{ for } M_5,$$

$$\pi(\theta) = \log(2) \exp \{-\log(2) \theta\}, \text{ for } M_2, M_3, M_4,$$

$$\pi(\theta) = 1, \text{ for } M_1,$$

and for any model:

$$\pi(\alpha_r) = \exp \{-\alpha_r\},$$

$$\pi(\xi_r) = \exp \{-\xi_r\},$$

for  $r = 1, 2, 3$ .

Before moving to the characterization of each model, we need to discuss an important issue of model (6.24), which we addressed in the previous section.

We know from Model 2 that:

$$\lim_{\theta \rightarrow \infty} M_k = M_1,$$

for  $k = 2, \dots, 4$ . As well, it can be easily verified from Part 1 of the proof in Appendix B.6, that

$$\begin{aligned} \lim_{\theta \rightarrow \infty} \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta \xi_1}} - 1 \right) + \alpha_2 \left( x_2^{\frac{1}{\theta \xi_2}} - 1 \right) + \alpha_3 \left( x_3^{\frac{1}{\theta \xi_3}} - 1 \right) \right)^{-\theta} \\ = \left( x_1^{-\frac{\alpha_1}{\xi_1}} x_2^{-\frac{\alpha_2}{\xi_2}} x_3^{-\frac{\alpha_3}{\xi_3}} \right). \end{aligned}$$

Therefore:

$$\lim_{\theta \rightarrow \infty} M_5 = M_1.$$

Hence, model (6.24) has an identification issue when the variables are exact independent. This fact is not a problem when the inference approach is model averaging (on which we will extend later). However, when the model selection

approach is selected, there is a strong equivalence between a tail dependence model with a large  $\theta$  and model  $M_1$ . We will deal with this issue further to make it clear, after we have exposed the inference.

### 6.2.3.2 Data Characterization for Trivariate Model

Recall from Section 5.5 that we are interested in positive variables, let them be  $X_1^*$ ,  $X_2^*$ , and  $X_3^*$ , and that there are two different objectives in inference: tail dependence determination and prediction.

Similarly to Section 5.5, we divide  $\mathbb{R}^3$  in the following regions:

$$\left\{ R_{ijk}^*; i = \mathbb{1}_{[X_1^* > u_{X_1^*}]}, j = \mathbb{1}_{[X_2^* > u_{X_2^*}]}, k = \mathbb{1}_{[X_3^* > u_{X_3^*}]} \right\}, \quad (6.25)$$

for  $u_{X_1^*}$ ,  $u_{X_2^*}$  and  $u_{X_3^*}$ , positive high thresholds of  $X_1^*$ ,  $X_2^*$ , and  $X_3^*$ , respectively. Then we only consider data in the region  $\{R_{000}\}^c$  and consider, as well, the following transformations:

$$X_1 = X_1^* + 1, \quad X_2 = X_2^* + 1, \quad \text{and} \quad X_3 = X_3^* + 1, \quad (6.26)$$

for  $(X_1^*, X_2^*, X_3^*) \in \{(0, u_{X_1^*}), (0, u_{X_2^*}), (0, u_{X_3^*})\}^c$ .

Therefore, the division of  $\mathbb{R}^3$  for the transformed variables is:

$$\left\{ R_{ijk}; i = \mathbb{1}_{[X_1 > u_{X_1}]}, j = \mathbb{1}_{[X_2 > u_{X_2}]}, k = \mathbb{1}_{[X_3 > u_{X_3}]} \right\}, \quad (6.27)$$

for  $u_{X_1} = u_{X_1^*} + 1$ ,  $u_{X_2} = u_{X_2^*} + 1$  and  $u_{X_3} = u_{X_3^*} + 1$ .

We model the vector  $(X_1, X_2, X_3)$  with the mixture model in (6.24). As we have already stated, there are two options to deal with a mixture model:

assume only one model is correct and treat the problem as model selection or recognise an uncertainty spread over different models, which is the target of the model averaging approach.

Consider the sample  $\mathbf{D}$  of size  $n$  of model (6.24), such that:

$$\mathbf{D} = \{(x_{1,i}, x_{2,i}, x_{3,i}), i = 1, \dots, n\},$$

and let  $\bar{\omega}_j$  be the subset:  $\bar{\omega}_j = \omega \setminus \{\omega_j\}$ , for  $\omega_j \in \omega$ .

Similarly to Section 5.5, we need to truncate every model, since we are ignoring the region  $R_{000}$ . Define  $\mathbf{u} = (u_{X_1}, u_{X_2}, u_{X_3})$ , then the density of the non-censored trivariate Model  $j$  is as follows:

$$f_{M_j}(\mathbf{x}|\omega) = \frac{f_{M_j}^*(\mathbf{x}) \mathbb{1}_{[x_1 > u_{X_1}, x_2 > u_{X_2}, x_3 > u_{X_3}]}}{1 - F_{M_j}^*(\mathbf{u})}, \quad (6.28)$$

for  $f_{M_j}^*$ , the density of trivariate model  $j$  of the mixture model (6.24) and for  $F_{M_j}^*$ , its trivariate joint distribution, for  $j = 1, \dots, 5$ .

In order to compute the denominator in (6.28), we use the following relation for any trivariate joint function:

$$1 - F_{\mathbf{X}} = \bar{F}_{X_1} + \bar{F}_{X_2} + \bar{F}_{X_3} - \bar{F}_{X_1, X_2} - \bar{F}_{X_1, X_3} - \bar{F}_{X_2, X_3} + \bar{F}_{\mathbf{X}}, \quad (6.29)$$

where  $\bar{F}_{X_j}$  is the univariate marginal survival function of  $X_j$ , and  $\bar{F}_{X_j, X_k}$  is the marginal bivariate survival function of  $X_j$  and  $X_k$ , for  $j$  and  $k$  in  $\{1, 2, 3\}$ .

### Trivariate *M1. TI*.

For the triple independent model, we know that the trivariate joint distri-

bution satisfies the following:

$$1 - F_{M_1}(\mathbf{u}) = u_{X_1}^{-\frac{\alpha_1}{\xi_1}} + u_{X_2}^{-\frac{\alpha_2}{\xi_2}} + u_{X_3}^{-\frac{\alpha_3}{\xi_3}} - u_{X_1}^{-\frac{\alpha_1}{\xi_1}} u_{X_2}^{-\frac{\alpha_2}{\xi_2}} - u_{X_1}^{-\frac{\alpha_1}{\xi_1}} u_{X_3}^{-\frac{\alpha_3}{\xi_3}} - u_{X_2}^{-\frac{\alpha_2}{\xi_2}} u_{X_3}^{-\frac{\alpha_3}{\xi_3}} + u_{X_1}^{-\frac{\alpha_1}{\xi_1}} u_{X_2}^{-\frac{\alpha_2}{\xi_2}} u_{X_3}^{-\frac{\alpha_3}{\xi_3}}.$$

Therefore, the joint density function for the triple independent model is:

$$f_{M_1}(\mathbf{x}|\boldsymbol{\omega}) = \left( \frac{\alpha_1 \alpha_2 \alpha_3}{\xi_1 \xi_2 \xi_3} \right) \frac{x_1^{-\left(\frac{\alpha_1}{\xi_1}+1\right)} x_2^{-\left(\frac{\alpha_2}{\xi_2}+1\right)} x_3^{-\left(\frac{\alpha_3}{\xi_3}+1\right)}}{1 - F_{M_1}(\mathbf{u})}.$$

Therefore, the log-posterior is the following:

$$lp(\boldsymbol{\omega}|\mathbf{D}, M_1) = n \log \left( \frac{\alpha_1 \alpha_2 \alpha_3}{\xi_1 \xi_2 \xi_3} \right) - \frac{\alpha_1}{\xi_1} \sum_{i=1}^n \log x_{1,i} - \frac{\alpha_2}{\xi_2} \sum_{i=1}^n \log x_{2,i} + K_{M_1} - \frac{\alpha_3}{\xi_3} \sum_{i=1}^n \log x_{3,i} - n \log (1 - F_{M_1}(\mathbf{u})) + \log \pi_{M_1}(\boldsymbol{\omega}),$$

where  $K_{M_1}$  is the logarithm of the normalizing constant of the posterior distribution.

Hence, the conditional log-posterior distributions are as follows:

$$lp(\alpha_1|\mathbf{D}, M_1, \bar{\alpha}_1) = n \log \alpha_1 - \left( \frac{\alpha_1}{\xi_1} \right) \sum_{i=1}^n \log x_{1,i} - n \log (1 - F_{M_1}(\mathbf{u})) - \alpha_1 + K_{\alpha_1},$$

$$lp(\alpha_2|\mathbf{D}, M_1, \bar{\alpha}_2) = n \log \alpha_2 - \left( \frac{\alpha_2}{\xi_2} \right) \sum_{i=1}^n \log x_{2,i} - n \log (1 - F_{M_1}(\mathbf{u})) - \alpha_2 + K_{\alpha_2},$$

$$lp(\alpha_3|\mathbf{D}, M_1, \bar{\alpha}_3) = n \log \alpha_3 - \left( \frac{\alpha_3}{\xi_3} \right) \sum_{i=1}^n \log x_{3,i} - n \log (1 - F_{M_1}(\mathbf{u})) - \alpha_3 + K_{\alpha_3},$$

$$lp(\xi_1|\mathbf{D}, M_1, \bar{\xi}_1) = -n \log \xi_1 - \left( \frac{\alpha_1}{\xi_1} \right) \sum_{i=1}^n \log x_{1,i} - n \log (1 - F_{M_1}(\mathbf{u})) - \xi_1 + K_{\xi_1},$$

$$lp(\xi_2|\mathbf{D}, M_1, \bar{\xi}_2) = -n \log \xi_2 - \left(\frac{\alpha_2}{\xi_2}\right) \sum_{i=1}^n \log x_{2,i} - n \log(1 - F_{M_1}(\mathbf{u})) \\ - \xi_2 + K_{\xi_2},$$

$$lp(\xi_3|\mathbf{D}, M_1, \bar{\xi}_3) = -n \log \xi_3 - \left(\frac{\alpha_3}{\xi_3}\right) \sum_{i=1}^n \log x_{3,i} - n \log(1 - F_{M_1}(\mathbf{u})) \\ - \xi_3 + K_{\xi_3},$$

where the  $K_{\omega_j}$  term represents the logarithm of the normalizing factor of the conditional posterior of  $\omega_j$ .

### Trivariate $M2$ . $DD_{1,2}$ .

For the double dependent model  $M2$ , the joint distribution holds the following:

$$1 - F_{M_2}(\mathbf{u}) = \left(1 + \alpha_1 \left(u_{X_1}^{\frac{1}{\theta\xi_1}} - 1\right)\right)^{-\theta} + \left(1 + \alpha_2 \left(u_{X_2}^{\frac{1}{\theta\xi_2}} - 1\right)\right)^{-\theta} + u_{X_3}^{-\frac{\alpha_3}{\xi_3}} \\ - \left(1 + \alpha_1 \left(u_{X_1}^{\frac{1}{\theta\xi_1}} - 1\right) + \alpha_2 \left(u_{X_2}^{\frac{1}{\theta\xi_2}} - 1\right)\right)^{-\theta} \\ - \left(1 + \alpha_1 \left(u_{X_1}^{\frac{1}{\theta\xi_1}} - 1\right)\right)^{-\theta} u_{X_3}^{-\frac{\alpha_3}{\xi_3}} - \left(1 + \alpha_2 \left(u_{X_2}^{\frac{1}{\theta\xi_2}} - 1\right)\right)^{-\theta} u_{X_3}^{-\frac{\alpha_3}{\xi_3}} \\ + \left(1 + \alpha_1 \left(u_{X_1}^{\frac{1}{\theta\xi_1}} - 1\right) + \alpha_2 \left(u_{X_2}^{\frac{1}{\theta\xi_2}} - 1\right)\right)^{-\theta} u_{X_3}^{-\frac{\alpha_3}{\xi_3}}.$$

Hence, the joint density function in this case is as follows:

$$f_{M_2}(\mathbf{x}|\boldsymbol{\omega}) = \left(\frac{(\theta+1)\alpha_1\alpha_2\alpha_3}{\theta\xi_1\xi_2\xi_3}\right) \frac{x_3^{-\frac{\alpha_3}{\xi_3}-1} x_1^{\frac{1}{\theta\xi_1}-1} x_2^{\frac{1}{\theta\xi_2}-1}}{1 - F_{M_2}(\mathbf{u})} \\ \times \left(1 + \alpha_1 \left(x_1^{\frac{1}{\theta\xi_1}} - 1\right) + \alpha_2 \left(x_2^{\frac{1}{\theta\xi_2}} - 1\right)\right)^{-\theta-2};$$

Then, the log-posterior is expressed as:

$$lp(\boldsymbol{\omega}|\mathbf{D}, M_2) = n \log \left(\frac{\alpha_1\alpha_2\alpha_3}{\xi_1\xi_2\xi_3}\right) + n \log \left(\frac{\theta+1}{\theta}\right) - \frac{\alpha_3}{\xi_3} \sum_{i=1}^n \log x_{3,i} \\ + \frac{1}{\theta\xi_1} \sum_{i=1}^n \log x_{1,i} + \frac{1}{\theta\xi_2} \sum_{i=1}^n \log x_{2,i} \\ - (\theta+2) \sum_{i=1}^n \log \left(1 + \alpha_1 \left(x_{1,i}^{\frac{1}{\theta\xi_1}} - 1\right) + \alpha_2 \left(x_{2,i}^{\frac{1}{\theta\xi_2}} - 1\right)\right) \\ - n \log(1 - F_{M_2}(\mathbf{u})) + \log \pi_{M_2}(\boldsymbol{\omega}) + K_{M_2},$$

where  $K_{M_2}$  is the logarithm of the normalizing constant of the posterior distribution.

If we define:

$$S_2(\boldsymbol{\omega}) = (\theta + 2) \sum_{i=1}^n \log \left( 1 + \alpha_1 \left( x_{1,i}^{\frac{1}{\theta \xi_1}} - 1 \right) + \alpha_2 \left( x_{2,i}^{\frac{1}{\theta \xi_2}} - 1 \right) \right).$$

Then, the conditional log-posterior distributions are as follows:

$$\begin{aligned} lp(\theta | \mathbf{D}, M_2, \bar{\theta}) &= n \log \left( \frac{\theta+1}{\theta} \right) + \frac{1}{\theta \xi_1} \sum_{i=1}^n \log x_{1,i} + \frac{1}{\theta \xi_2} \sum_{i=1}^n \log x_{2,i} \\ &\quad - S_2(\boldsymbol{\omega}) - n \log(1 - F_{M_2}(\mathbf{u})) - \log(2) \theta + K_{\theta}, \end{aligned}$$

$$lp(\alpha_1 | \mathbf{D}, M_2, \bar{\alpha}_1) = n \log \alpha_1 - \alpha_1 - S_2(\boldsymbol{\omega}) - n \log(1 - F_{M_2}(\mathbf{u})) + K_{\alpha_1},$$

$$lp(\alpha_2 | \mathbf{D}, M_2, \bar{\alpha}_2) = n \log \alpha_2 - \alpha_2 - S_2(\boldsymbol{\omega}) - n \log(1 - F_{M_2}(\mathbf{u})) + K_{\alpha_2},$$

$$\begin{aligned} lp(\alpha_3 | \mathbf{D}, M_2, \bar{\alpha}_3) &= n \log \alpha_3 - \frac{\alpha_3}{\xi_3} \sum_{i=1}^n \log x_{3,i} - \alpha_3 \\ &\quad - n \log(1 - F_{M_2}(\mathbf{u})) + K_{\alpha_3}, \end{aligned}$$

$$\begin{aligned} lp(\xi_1 | \mathbf{D}, M_2, \bar{\xi}_1) &= -n \log \xi_1 - \xi_1 + \frac{1}{\theta \xi_1} \sum_{i=1}^n \log x_{1,i} - S_2(\boldsymbol{\omega}) \\ &\quad - n \log(1 - F_{M_2}(\mathbf{u})) + K_{\xi_1}, \end{aligned}$$

$$\begin{aligned} lp(\xi_2 | \mathbf{D}, M_2, \bar{\xi}_2) &= -n \log \xi_2 - \xi_2 + \frac{1}{\theta \xi_2} \sum_{i=1}^n \log x_{2,i} - S_2(\boldsymbol{\omega}) \\ &\quad - n \log(1 - F_{M_2}(\mathbf{u})) + K_{\xi_2}, \end{aligned}$$

$$\begin{aligned} lp(\xi_3 | \mathbf{D}, M_2, \bar{\xi}_3) &= -n \log \xi_3 - \frac{\alpha_3}{\xi_3} \sum_{i=1}^n \log x_{3,i} - \xi_3 \\ &\quad - n \log(1 - F_{M_2}(\mathbf{u})) + K_{\xi_3}, \end{aligned}$$

where the  $K_{\omega_j}$  term represents the logarithm of the normalizing factor of the conditional posterior of  $\omega_j$ .

The results of  $M3(DD_{1,3})$  and  $M4(DD_{2,3})$  are similar to those showed in this section. Therefore, we move ahead to model  $M_5$ .

**Trivariate  $M5$ . TD.**

The joint distribution of  $M5$  satisfies the following:

$$\begin{aligned}
 1 - F_{M_5}(\mathbf{u}) &= \sum_{j=1}^3 \left( 1 + \alpha_j \left( u_{X_j}^{\frac{1}{\theta \xi_j}} - 1 \right) \right)^{-\theta} \\
 &\quad - \left( 1 + \alpha_1 \left( u_{X_1}^{\frac{1}{\theta \xi_1}} - 1 \right) + \alpha_2 \left( u_{X_2}^{\frac{1}{\theta \xi_2}} - 1 \right) \right)^{-\theta} \\
 &\quad - \left( 1 + \alpha_1 \left( u_{X_1}^{\frac{1}{\theta \xi_1}} - 1 \right) + \alpha_3 \left( u_{X_3}^{\frac{1}{\theta \xi_3}} - 1 \right) \right)^{-\theta} \\
 &\quad - \left( 1 + \alpha_2 \left( u_{X_2}^{\frac{1}{\theta \xi_2}} - 1 \right) + \alpha_3 \left( u_{X_3}^{\frac{1}{\theta \xi_3}} - 1 \right) \right)^{-\theta} \\
 &\quad + \left( 1 + \alpha_1 \left( u_{X_1}^{\frac{1}{\theta \xi_1}} - 1 \right) + \alpha_2 \left( u_{X_2}^{\frac{1}{\theta \xi_2}} - 1 \right) + \alpha_3 \left( u_{X_3}^{\frac{1}{\theta \xi_3}} - 1 \right) \right)^{-\theta}.
 \end{aligned}$$

Then, the last joint density we present is as follows:

$$\begin{aligned}
 f_{M_5}(\mathbf{x}|\boldsymbol{\omega}) &= \left( \frac{(\theta+2)(\theta+1)\alpha_1\alpha_2\alpha_3}{\theta^2\xi_1\xi_2\xi_3} \right) \frac{x_1^{\frac{1}{\theta\xi_1}-1} x_2^{\frac{1}{\theta\xi_2}-1} x_3^{\frac{1}{\theta\xi_3}-1}}{1 - F_{M_5}(\mathbf{u})} \\
 &\quad \times \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta\xi_1}} - 1 \right) + \alpha_2 \left( x_2^{\frac{1}{\theta\xi_2}} - 1 \right) + \alpha_3 \left( x_3^{\frac{1}{\theta\xi_3}} - 1 \right) \right)^{-\theta-3};
 \end{aligned}$$

Therefore, the log-posterior goes as follows:

$$\begin{aligned}
 lp(\boldsymbol{\omega}|\mathbf{D}, M_5) &= n \log \left( \frac{\alpha_1\alpha_2\alpha_3(\theta+1)(\theta+2)}{\xi_1\xi_2\xi_3\theta^2} \right) + \log \pi_{M_5}(\boldsymbol{\omega}) + \frac{1}{\theta\xi_1} \sum_{i=1}^n \log x_{1,i} \\
 &\quad + \frac{1}{\theta\xi_2} \sum_{i=1}^n \log x_{2,i} + \frac{1}{\theta\xi_3} \sum_{i=1}^n \log x_{3,i} - (\theta + 3) \times \\
 &\quad \sum_{i=1}^n \log \left( 1 + \alpha_1 \left( x_{1,i}^{\frac{1}{\theta\xi_1}} - 1 \right) + \alpha_2 \left( x_{2,i}^{\frac{1}{\theta\xi_2}} - 1 \right) + \alpha_3 \left( x_{3,i}^{\frac{1}{\theta\xi_3}} - 1 \right) \right) \\
 &\quad - n \log(1 - F_{M_5}(\mathbf{u})) + K_{M_5},
 \end{aligned}$$

where  $K_{M_5}$  is the logarithm of the normalizing constant of the posterior distribution.

If we define:

$$S(\boldsymbol{\omega}) = (\theta + 3) \sum_{i=1}^n \log \left( 1 + \alpha_1 \left( x_{1,i}^{\frac{1}{\theta\xi_1}} - 1 \right) + \alpha_2 \left( x_{2,i}^{\frac{1}{\theta\xi_2}} - 1 \right) + \alpha_3 \left( x_{3,i}^{\frac{1}{\theta\xi_3}} - 1 \right) \right).$$

Then, the conditional log-posterior distributions are expressed as:

$$\begin{aligned} lp(\theta|\mathbf{D}, M_5, \bar{\theta}) = & n \log \left( \frac{(\theta+1)(\theta+2)}{\theta^2} \right) + K_\theta - \log(3)\theta + \frac{1}{\theta\xi_1} \sum_{i=1}^n \log x_{1,i} \\ & + \frac{1}{\theta\xi_2} \sum_{i=1}^n \log x_{2,i} + \frac{1}{\theta\xi_3} \sum_{i=1}^n \log x_{3,i} \\ & - n \log(1 - F_{M_5}(\mathbf{u})) - S(\boldsymbol{\omega}), \end{aligned}$$

$$lp(\alpha_1|\mathbf{D}, M_5, \bar{\alpha}_1) = n \log \alpha_1 - \alpha_1 - n \log(1 - F_{M_5}(\mathbf{u})) + K_{\alpha_1} - S(\boldsymbol{\omega}),$$

$$lp(\alpha_2|\mathbf{D}, M_5, \bar{\alpha}_2) = n \log \alpha_2 - \alpha_2 - n \log(1 - F_{M_5}(\mathbf{u})) + K_{\alpha_2} - S(\boldsymbol{\omega}),$$

$$lp(\alpha_3|\mathbf{D}, M_5, \bar{\alpha}_3) = n \log \alpha_3 - \alpha_3 - n \log(1 - F_{M_5}(\mathbf{u})) + K_{\alpha_3} - S(\boldsymbol{\omega}),$$

$$\begin{aligned} lp(\xi_1|\mathbf{D}, M_5, \bar{\xi}_1) = & -n \log \xi_1 - \xi_1 + K_{\xi_1} + \frac{1}{\theta\xi_1} \sum_{i=1}^n \log x_{1,i} \\ & - n \log(1 - F_{M_5}(\mathbf{u})) - S(\boldsymbol{\omega}), \end{aligned}$$

$$\begin{aligned} lp(\xi_2|\mathbf{D}, M_5, \bar{\xi}_2) = & -n \log \xi_2 - \xi_2 + K_{\xi_2} + \frac{1}{\theta\xi_2} \sum_{i=1}^n \log x_{2,i} \\ & - n \log(1 - F_{M_5}(\mathbf{u})) - S(\boldsymbol{\omega}), \end{aligned}$$

$$\begin{aligned} lp(\xi_3|\mathbf{D}, M_5, \bar{\xi}_3) = & -n \log \xi_3 - \xi_3 + K_{\xi_3} + \frac{1}{\theta\xi_3} \sum_{i=1}^n \log x_{3,i} \\ & - n \log(1 - F_{M_5}(\mathbf{u})) - S(\boldsymbol{\omega}), \end{aligned}$$

where the  $K_{\omega_j}$  term represents the logarithm of the normalizing factor of the conditional posterior of  $\omega_j$ .

Although we are not analyzing the censored case, we note that the censoring becomes more cumbersome as the dimension increases, since the likelihood splits into more combinations of partial derivatives of the joint model, as in (5.19).

### 6.2.3.3 Reversible Jump MCMC

The aim of the Reversible Jump MCMC depends on the objective of the analysis: in the case the asymptotic dependence determination is required, then, it is necessary to sample from  $\omega$ , and  $\{w_{k,T}\}_{k=1}^5$ . In the case of prediction, the algorithm is quite similar, except for the fact that a sample of  $\mathbf{X}$  is drawn on each step, regardless whether the proposed jump is accepted or not.

The sampling for  $\mathbf{X}$  at step  $k(t)$  is as follows:

- If  $k(t) = 1$ , sample a uniform  $u_j$ . Then

$$X_j^S(t) = u_j^{-\frac{\xi_j(t)}{\alpha_j(t)}}, \quad (6.30)$$

for  $j = 1, 2, 3$ .

- If  $k(t) = 2$ , sample  $A$  from  $Gamma(\theta(t), 1)$ ,  $B_j$  from  $Exp(\alpha_j(t))$ , and a uniform  $u$ . Then

$$X_j^S(t) = (1 + B_j/A)^{\theta(t)\xi_j(t)} \text{ and } X_3^S(t) = u_j^{-\frac{\xi_3(t)}{\alpha_3(t)}},$$

for  $j = 1, 2$ .

- If  $k(t) = 5$ , sample  $A$  from  $Gamma(\theta(t), 1)$  and  $B_j$  from  $exp(\alpha_j(t))$ . Then

$$X_j^S(t) = (1 + B_j/A)^{\theta(t)\xi_j(t)},$$

for  $j = 1, 2, 3$ .

The cases where  $k(t) = 3$  and  $k(t) = 4$  are similar to  $k(t) = 2$ .

In any of the individual models, we sample  $\mathbf{X}(t)$  until  $\mathbf{X}(t) \in \{R_{000}\}^c$ .

In this way,  $\mathbf{X}^S$  forms a sample of the predictive distribution of  $\mathbf{X}$ . In the prediction approach, although the algorithm returns a sample of  $\omega$  and  $\{w_k\}_{k=1}^5$ , they are not the crucial information.

For the model selection approach, the most relevant information is in  $\{w_k\}_{k=1}^5$ . The iterations in the RJ are continued until the processes  $\{w_k\}_{k=1}^5$  stabilize. Assume the stabilization is achieved at  $t = T$ . Then, all the necessary information is contained in both  $\{w_k\}_{k=1}^5$  and  $\omega^S = \{\omega_1^S, \dots, \omega_T^S\}$ .

Finally, we close the discussion of the trivariate model with some simulation cases.

### 6.2.4 Trivariate Simulation

We made two simulation cases. The objective of the first case is to study the Model Selection approach. Therefore, we draw observations of the trivariate model, and then we are interested in how the Reversible Jump recognizes the dependence structure for which the data are simulated. The analysis will focus on the estimation of  $w_1, \dots, w_5$ , and of  $\theta$ .

For the second case, we are interested in the implementation of the Model Averaging approach. Similarly to Section 5.6.3 we draw simulations from a multivariate copula of GPD variables. The objective of this simulation is the prediction of the variables in the tails.

### 6.2.4.1 Simulations for Model Selection

We simulate observations for each model similarly to the simulations in Section 5.6.2. We simulate  $X_1, X_2$ , and  $X_3$  in  $\{(0, 1)\}^c$ , i.e., we are setting  $u_{X_1} = u_{X_2} = u_{X_3} = 1$  in the region (6.27).

Regarding the number of observations, the data is divided into small sample:  $n = 70$ , medium sample:  $n = 150$ , and large sample:  $n = 300$  (except for the  $M_1$  case, where a large sample is not needed). Regarding the heaviness of the tails, we proceed as in Model 2 in Section 5.6.3: a block of heavy-tailed and another of light-tailed models.

As the scale parameters do not have a great importance in the tail dependence structure, we fix them to 0.5, which means:

$$\alpha_1 = \alpha_2 = \alpha_3 = 0.5.$$

For the marginal shape parameters, we proceed as follows:

- For heavy-tailed models:  $\xi_1 = 2$ ,  $\xi_2 = 3$ , and  $\xi_3 = 2.5$ .
- For light-tailed models:  $\xi_1 = 0.1$ ,  $\xi_2 = 0.2$ , and  $\xi_3 = 0.15$ .

Regarding the priors, we fix the hyperparameters to 1, i.e.:

$$\delta_{\omega_j} = 1, \text{ for all } \omega_j \in \omega.$$

The sizes of the sampling for the Reversible Jump are set up to  $10^5$ , until a level which was appropriate to assume that the processes  $\{w_k\}_{k=1}^5$  were stable.

We plot in Figure D.8 two processes to show how they stabilize.

In Table C.4 we present the results when simulating from model  $M_1$ , i.e. we draw simulations of independent Paretos, as in (6.30). We can appreciate that practically most of the times, the jumping process stays in  $M_1$ .

The next case we want to expose is the model with triple dependence, i.e.  $M_5$ . We draw the samples via the representation in (6.3). In Table C.7 we present the results. For the small size sample, the method estimates accurately  $w_5$ , except when  $\theta \geq 3$  and when  $\theta \geq 4$  in light-tailed and heavy-tailed models, respectively. For the medium size sample, the difficulties arise when  $\theta \geq 4$  and when  $\theta \geq 4.5$  in light-tailed and heavy-tailed models, respectively. We can see that the inaccuracies are due to the lack of data. We drew a large sample with  $n = 300$  to illustrate this idea. We present the results in Table C.8. We can see that the difficulties arise, in both cases, only when  $\theta \geq 7$ .

Finally, to represent the models with double dependence, we simulate observations from  $M_2$ . The way to do so is by simulating  $X_1$  and  $X_2$  as in Model 2 (see Section 5.6.3), and  $X_3$ , an independent Pareto, as in  $M_1$ . The results for the small and medium size samples are showed in Table C.10. We can appreciate that the methodology can distinguish accurately the model  $M_2$  for  $\chi > 0.13$ , i.e., for  $\theta < 3$  in heavy tails and  $\theta \leq 3$ , for light-tails. For the medium size sample, the accuracy extends to  $\theta \leq 3.5$ , in both heavy and light tails. As a last step, we present the results of a large sample,  $n = 300$ , in Table C.9. We can notice that the accuracy is acceptable for  $\theta \leq 5$ , for both heavy and light tails.

We highlighted the cases where the identification presents some difficulties. However, as it can be seen in the figures, the difficulties appear when the value of  $\theta$  is considerably large. Therefore, from the discussion of the identifiability of Sections 6.2.3.1 and 5.6, the results are in accordance with what we expect.

#### 6.2.4.2 Simulations for Model Averaging

The simulations for Modeling Averaging are similar to those in Section 5.6.3, since we are not simulating from the trivariate model. We are aiming to fit the trivariate model to the region  $\{R_{000}\}^c$  (recall (6.25)) for different dependence structures.

For all the cases, the marginal distributions are GPD's with the following notation:

- Light tails:  $k_1 = 0.1$ ,  $k_2 = 0.15$ , and  $k_3 = 0.125$ ,
- Heavy tails:  $k_1 = 1$ ,  $k_2 = 1.5$ , and  $k_3 = 1.25$ ,

with  $\beta = 1$  for all of them (see (2.5) for the GPD representation).

In terms of joint structure, we simulated from four cases, which will be explained later:

- Triple medium tail dependence.
- Triple high tail dependence.
- Exact Independence.

- Double medium tail dependence.

We simulate 1500 observations in  $\mathbb{R}_+^3$ . Then, only the observations in  $\{R_{000}\}^c$  are taken into account. The threshold  $u_{X_j^*}$  is selected as the 0.95 quantile of  $X_j^*$ , for  $j = 1, 2, 3$ . Then, selected data is transformed via (6.26), this means, if the selected variables are  $X_1^*$ ,  $X_2^*$ , and  $X_3^*$ , then, the transformed variables  $X_1$ ,  $X_2$ , and  $X_3$  are:

$$X_1 = X_1^* + 1, \quad X_2 = X_2^* + 1, \quad \text{and} \quad X_3 = X_3^* + 1,$$

for  $(X_1^*, X_2^*, X_3^*) \in \{(0, u_{X_1^*}), (0, u_{X_2^*}), (0, u_{X_3^*})\}^c$ .

Therefore, the trivariate model (6.24) is fitted to  $X_1$ ,  $X_2$ , and  $X_3$ .

Recall the aim of Model Averaging approach is the prediction. Therefore, opposite to the previous simulation cases, we are not interested in estimating the parameters. The inference methodology followed is the one described in Section 6.2.3.3, where we asserted the crucial information of this approach consists in  $\mathbf{X}^S$ , the prediction sampling.

We use graphic tools to study how the prediction methodology works.

Recall the definition in (2.26), where for  $0 < u < 1$ :

$$\chi(u) = 2 - \frac{\log P(X_1 < F_{X_1}^+(u), X_2 < F_{X_2}^+(u))}{\log u},$$

where the inverse marginal distributions and the joint probability are estimated by its empirical versions.

In a similar way, we can define, for  $0 < u < 1$ :

$$\chi_3(u) = 2 - \frac{\log P(X_1 < F_{X_1}^+(u), X_2 < F_{X_2}^+(u), X_3 < F_{X_3}^+(u))}{\log u},$$

and we estimate it with the empirical distributions, as with  $\chi(u)$ .

Therefore, the graphic tools we use are the following:

- Bivariate marginal scatter plots: prediction sampling vs data.
- Sampling  $\chi(u)$  vs data  $\chi(u)$ .
- Sampling  $\chi_3(u)$  vs data  $\chi_3(u)$ , where applicable.

### Simulations for triple medium tail dependence

For the first case, we simulate  $X_1$ ,  $X_2$ , and  $X_3$  from a Clayton copula with  $\phi = 2$  (recall 5.12), and with light GPD marginals. This model implies  $\chi_3 = 1/3$ , and  $\chi = 0.5$ , for any pair.

We present in the first line of Figure D.9 the scatter plots for the three different pairs in log-scale, whereas in the bottom line, the predictive sampling is shown for each case. The thresholds are not as noticeable as in the bivariate case (see Figure D.6), since there are observations which are below the thresholds for the two variables in the graph, but above the threshold of the third variable. The predictive samplings give a similar behavior as the data. In fact, the three predictive samplings give a characteristic behavior of medium tail dependent variables.

On the upper line of Figure D.10 we show the graphics of the empirical  $\chi(u)$  for two different pairs of the variables, and the empirical  $\chi_3(u)$ . In the bottom line we present the same graphics for the predictive sampling. The three bottom plots look like a smooth version of the empirical graphics for

medium and large values of  $u$ , which suggests a good fit of the bivariate and trivariate tail dependence. Even more, these graphics show a considerable accurate estimation of the theoretical values of  $\chi$  and  $\chi_3$ , through the graphical method.

### **Simulations for triple high tail dependence**

For the triple high tail dependence case, we simulate  $X_1$ ,  $X_2$ , and  $X_3$  from a Clayton copula with  $\phi = 6$  and with light GPD marginals. This model implies  $\chi_3 = 0.802$ , and  $\chi = 0.871$ , for any pair.

The first line of Figure D.11 shows the scatter plots in log-scale for the pairs  $(X_1, X_2)$  and  $(X_1, X_3)$ , whereas in the bottom line the predictive sampling is shown for each case. The predictive samplings show the strong pair relation of data.

The upper line of Figure D.12 shows the graphics of the empirical  $\chi(u)$  for two different pairs of the variables, and the empirical  $\chi_3(u)$ . The bottom line presents the same graphics for the predictive sampling. The three bottom plots do not look as smooth as those in the medium tail dependence case. However, they look as a smoother version of the empirical graphics. As well, the estimated value of  $\chi_3$  suggested by the graphic is considerably accurate, whereas the estimation of  $\chi$  is overestimated, but in accordance with high tail dependence.

### **Simulations for exact independence**

In this case we simulated independently  $X_1$ ,  $X_2$ , and  $X_3$  from light GPD's.

The first line of Figure D.13 shows the scatter plots in log-scale for the pairs  $(X_1, X_2)$  and  $(X_1, X_3)$ , whereas the second line the predictive sampling is shown for each case. The predictive samplings show the same independent relation in the tails as the data.

The third line of Figure D.13 shows the graphics of the empirical  $\chi(u)$  for two different pairs of the variables. The bottom line presents the same graphics for the predictive sampling. The bottom plots seem to be a smooth version of the empirical graphics. The important part is that they do not show any tail dependence.

### Simulations for double medium tail dependence

For the triple high tail dependence case, we simulate  $X_1$  and  $X_2$  from a Clayton copula with  $\phi = 2$ , and with light GPD marginals, whereas  $X_3$  is simulated independently from a GPD with light tail. This model implies  $\chi = 0.5$  for  $X_1$  and  $X_2$ .

The first line of Figure D.14 shows the scatter plots in log-scale for the pairs  $(X_1, X_2)$  and  $(X_1, X_3)$ , whereas in the second line the predictive sampling is shown for each case. The predictive samplings show the same association in tails as the data: for the pair  $(X_1, X_2)$  there exists a strong relation in the tails, whereas for the pair  $(X_1, X_3)$ , there is no evident association.

The third line of Figure D.14 shows the graphics of the empirical  $\chi(u)$  for the pairs  $(X_1, X_2)$  and  $(X_1, X_3)$ . The bottom line presents the same graphics for the predictive sampling. The bottom plots seem to be a smooth version of

the empirical graphics. For the pair  $(X_1, X_2)$ , the graphic shows a tail dependence with  $\chi$  close to 0.5, whereas for the pair  $(X_1, X_3)$ , it shows independence.

Once we have shown the main features of the trivariate mixture model, we can set the multivariate generalization.

### 6.3 Multivariate Model

The generalization to the multivariate setting is quite straightforward, after the introduction of the trivariate model.

Let  $\mathbf{X} = (X_1, \dots, X_d)$  be a vector of r.v. in  $\{(1, \infty)^d\}$ .

We define latent variables that determine multivariate models with different tail dependence as follows:

- *TI*, if  $F_{\mathbf{X}}$  has  $d$ -tuple tail independence.
- $MD_{j^*_1, \dots, j^*_M}(\theta)$ , if  $F_{\mathbf{X}}$  has  $M$ -tuple tail dependence, and  $(X_{j^*_1}, \dots, X_{j^*_M})$  are  $d$ -dimensional tail dependent, for  $\{j^*_1, \dots, j^*_M\} \in \{1, \dots, d\}$ ,  $2 \leq M \leq d$ .

For example,  $2D_{3,4}(\theta)$  is the model with double tail dependence between the variables  $(X_3, X_4)$ .

We propose the following model averaging:

$$\begin{aligned} \bar{F}_{\mathbf{X}} = & w_1 \bar{F}_{\mathbf{X}|TI} + \sum_{k_2=1}^d \sum_{k_1=k_2+1}^d w_{g(k_1, k_2, 2)} \bar{F}_{\mathbf{X}|2D_{k_1, k_2}} + \dots + \\ & \sum_{k_M=1}^d \sum_{k_{M-1}=k_M+1}^d \dots \sum_{k_1=k_2+1}^d w_{g(k_1, \dots, k_M, M)} \bar{F}_{\mathbf{X}|MD_{k_1, \dots, k_M}}, \end{aligned} \quad (6.31)$$

where  $g(k_1, \dots, k_M, M) : \mathbb{N}^{M+1} \rightarrow \mathbb{N}$ , enumerates the weights, and  $0 \leq w_j \leq 1$ , for  $j = 1, \dots, d_T$  and  $\sum_{j=1}^{d_T} w_j = 1$ , where:

$$d_T = \binom{d}{0} + \binom{d}{2} + \binom{d}{3} + \dots + \binom{d}{d-1} + \binom{d}{0} = 2 + \binom{2d-3}{d-1}, \quad (6.32)$$

for  $d \geq 2$ , is the total number of individual models.

Similarly to the marginal bivariate and the trivariate models introduced in (6.6) and (6.19), respectively, we construct a model with dimensional coherence. Therefore, as a first attempt of a multivariate model, we propose the following parametric form for each individual model in (6.31), for  $\mathbf{x} = (x_1, \dots, x_d)$ :

$$\bar{F}_{\mathbf{x}|TI}^*(\mathbf{x}) = \prod_{j=1}^d \left( 1 + \alpha_j (x_j^{-\frac{1}{\theta \zeta_j}} - 1) \right)^{-\theta}. \quad (6.33)$$

$$\begin{aligned} \bar{F}_{\mathbf{x}|MD_{k_1, \dots, k_M}}^*(\mathbf{x}) &= \left( 1 + \alpha_{k_1} \left( x_{k_1}^{\frac{1}{\theta \zeta_{k_1}}} - 1 \right) + \dots + \alpha_{k_M} \left( x_{k_M}^{\frac{1}{\theta \zeta_{k_M}}} - 1 \right) \right)^{-\theta} \\ &\quad \times \prod_{j=k_1^*}^{k_{d-M}^*} \left( 1 + \alpha_j (x_j^{-\frac{1}{\theta \zeta_j}} - 1) \right)^{-\theta}, \end{aligned} \quad (6.34)$$

where  $\{k_1^*, \dots, k_{d-M}^*\}$  is the complement of  $\{k_1, \dots, k_M\}$ .

We define  $\bar{F}_{\mathbf{x}}^*$  with components represented by (6.33) and (6.34), as the multivariate mixture model with marginal coherence. We show the marginal coherence as follows:

Suppose we marginalize  $\bar{F}_{\mathbf{x}}^*$  with respect to  $X_i$ , for  $i \in \{1, \dots, d\}$ . Define  $\mathbf{z} = (x_1, \dots, x_i = 1, \dots, x_d)$ . Then the previous components are expressed as:

$$\bar{F}_{\mathbf{x}|TI}^*(\mathbf{z}) = \prod_{j=1, j \neq i}^d \left( 1 + \alpha_j (z_j^{-\frac{1}{\theta \zeta_j}} - 1) \right)^{-\theta}.$$

For each of the rest of the components, we have two options:

- if  $i \in \{k_1^*, \dots, k_{d-M}^*\}$ :

$$\begin{aligned} \bar{F}_{\mathbf{X}|MD_{k_1, \dots, k_M}}^* (\mathbf{z}) &= \left( 1 + \alpha_{k_1} \left( z_{k_1}^{\frac{1}{\theta \xi_{k_1}} - 1} \right) + \dots + \alpha_{k_M} \left( z_{k_M}^{\frac{1}{\theta \xi_{k_M}} - 1} \right) \right)^{-\theta} \\ &\quad \times \prod_{j=k_1^*, j \neq i}^{k_{d-M}^*} \left( 1 + \alpha_j \left( z_j^{-\frac{1}{\theta \xi_j}} - 1 \right) \right)^{-\theta}; \end{aligned}$$

- if  $i \in \{k_1, \dots, k_M\}$ , say  $i = k_i$ , then:

$$\begin{aligned} \bar{F}_{\mathbf{X}|MD_{k_1, \dots, k_M}}^* (\mathbf{z}) &= \left( 1 + \alpha_{k_1} \left( z_{k_1}^{\frac{1}{\theta \xi_{k_1}} - 1} \right) + \dots + \alpha_{k_{i-1}} \left( z_{k_{i-1}}^{\frac{1}{\theta \xi_{k_{i-1}}} - 1} \right) \right. \\ &\quad \left. + \alpha_{k_{i+1}} \left( z_{k_{i+1}}^{\frac{1}{\theta \xi_{k_{i+1}}} - 1} \right) + \dots + \alpha_{k_M} \left( z_{k_M}^{\frac{1}{\theta \xi_{k_M}} - 1} \right) \right)^{-\theta} \\ &\quad \times \prod_{j=k_1^*}^{k_{d-M}^*} \left( 1 + \alpha_j \left( z_j^{-\frac{1}{\theta \xi_j}} - 1 \right) \right)^{-\theta}. \end{aligned}$$

These components are identical as those in (6.31), but with one dimension less. For example, the second factor of (6.31) would be expressed as:

$$\sum_{k_2=1, k_2 \neq i}^d \sum_{k_1=k_2+1, k_1 \neq i}^d w_{g(k_1, k_2, 2)} \bar{F}_{\mathbf{X}|2D_{k_1, k_2}} (\mathbf{z}),$$

which has  $\binom{d-1}{2}$  components.

Hence, the marginal model has a total of  $2 + \binom{2(d-1)-3}{d-2}$ , for  $d > 2$ . Therefore, there exists marginal coherence in model (6.31).

**Theorem 6.3.1.** *Multivariate tail dependence structure*

Let  $\mathbf{X} = (X_1, \dots, X_d)$  be a random vector with joint survival function as in (6.31) with individual components as in (6.33) and (6.34).

Then, from Definition 6.1.4 the following holds:

- The tail dependence parameter  $\chi_3$  of the model is given by:

$$\chi_3^* = w_d d^{-\theta},$$

where  $w_d$  is the weight of the  $d$ -tuple model component.

- The pairwise tail dependence for the pair  $\{i, j\}$  is given by:

$$\chi_{\{i,j\}}^* = 2^{-\theta} \sum_{k=1}^{T_{i,j}} w_{\beta(i,j,M)}, \quad (6.35)$$

where  $\beta(i, j, M) : \mathbb{N}^2 \times \{1, \dots, T_{i,j}\} \rightarrow \mathbb{N}$  enumerates the weights where neither  $X_i$  nor  $X_j$  are exactly independent, and where:

$$T_{i,j} = 1 + (d-2) + \binom{d-2}{2} + \binom{d-3}{2} + \dots + \binom{3}{2} + \binom{2}{2},$$

for  $d > 3$ .

We show in Appendix B.10 the proof of the theorem.

Similarly to the bivariate and trivariate mixture models, we introduce a helpful modification of the multivariate mixture model with marginal coherence. This modification has the same representation as that in (6.31), but the individual components are represented as follows:

$$\begin{aligned} \bar{F}_{\mathbf{X}|TI}(\mathbf{x}) &= \prod_{j=1}^d x_j^{-\frac{\alpha_j}{\xi_j}}. \\ \bar{F}_{\mathbf{X}|MD_{k_1, \dots, k_M}}(\mathbf{x}) &= \left( 1 + \alpha_{k_1} \left( x_{k_1}^{\frac{1}{\theta \xi_{k_1}}} - 1 \right) + \dots + \alpha_{k_M} \left( x_{k_M}^{\frac{1}{\theta \xi_{k_M}}} - 1 \right) \right)^{-\theta} \\ &\quad \times \prod_{j=k_1^*}^{k_d^*-M} x_j^{-\frac{\alpha_j}{\xi_j}}, \end{aligned} \quad (6.36)$$

where  $\{k_1^*, \dots, k_{d-M}^*\}$  is the complement of  $\{k_1, \dots, k_M\}$ .

In this way, it is clear that this modification drops the marginal coherence. However, as we saw in the trivariate mixture model, the inference methodology is simplified if we express an independent model as a Pareto distribution.

The parameter space can be defined, similarly to the procedure in Section 6.2.3, as:

$$\omega = \{\theta, \alpha_1, \dots, \alpha_d, \xi_1, \dots, \xi_d\} \in \Omega = \mathbb{R}_+^{2d+1}.$$

Notice that we are adding the parameter  $\theta$  to the exact independent case, although it is not used in the individual model. Similarly to the trivariate model, the unification of parameter spaces will be useful when applying the Reversible Jump methodology.

The prior distributions are set as independent exponential with hyperparameters  $\delta_{\omega_k}$ , for  $\omega_k \in \omega$ .

As in the trivariate case, the tail dependence structure is the same as the dimensional coherence model in Theorem 6.3.1.

The next step is to select either model averaging or model selection as the best way to approach the problem.

Define the variable in study as  $\mathbf{X}^* = (X_1^*, \dots, X_d^*)$ . We transform the variables, similarly to (6.26), as  $X_j = X_j^* + 1$ . Therefore, the mixture model is applied to the region:

$$\{(1, u_{X_1}) \times \dots \times (1, u_{X_d})\}^c,$$

where  $u_{X_j} = u_{X_j^*} + 1$ , and where  $u_{X_j^*}$  is a positive high thresholds of  $X^*$ , for  $j = 1, \dots, d$ .

Finally, define the transformed data  $\mathbf{D}$  of size  $n$  as

$$\mathbf{D} = \{(x_{1,i}, \dots, x_{d,i}), i = 1, \dots, n\}.$$

Hence, we can reexpress the multivariate mixture model as:

$$\bar{F}_{\mathbf{X}} = w_1 \bar{F}_{M_1} + \dots + w_{d_T} \bar{F}_{M_{d_T}},$$

or, in terms of the posterior distribution:

$$p_{\mathbf{X}}(\boldsymbol{\omega}) = w_1 p(\boldsymbol{\omega} | \mathbf{D}, M_1) + \dots + w_{d_T} p(\boldsymbol{\omega} | \mathbf{D}, M_{d_T}),$$

where  $p(\boldsymbol{\omega} | \mathbf{D}, M_k)$  is the posterior distribution of individual model  $M_k$ , given the data  $\mathbf{D}$ , for  $k = 1, \dots, d_T$ .

It is important to notice that the density function of each model must be truncated individually as shown in Sections 5.5 and 6.2.3.2.

Then, the Reversible Jump methodology is applied: drawing samples of  $\boldsymbol{\omega}$  and of  $\{w_1, \dots, w_{d_T}\}$  for the model selection approach or sampling  $\mathbf{X}$  for the prediction or model averaging approach.

Finally, we need to discuss the extension of the multivariate model. This model is formed by a mixture of individual models which split the variables into two groups with respect to  $\theta$ : a group with tail dependence measured by  $\theta$  and an exact independent group. Therefore, these individual models cannot recover a dependence structure of more than two groups of dependence. For

example, in the 4-dimensional case, if  $X_1$  and  $X_2$  are medium tail dependent and  $X_3$  and  $X_4$  are high tail dependent, then, the individual models (except for the exact independent one) can only give one tail dependence structure, measured by  $\theta$ . In this sense, the model selection is not able to recover this structure.

However, the model averaging approach is flexible enough to give different tail dependencies to different pairs, as we saw in Theorems 6.2.2 and 6.3.1. Recall we interpret model averaging as the lack of certainty on which model is the correct one. In this way, we were spreading the uncertainty among different models. This spreading translates into flexibility of the model in the tail dependence structure: the tail dependence does not depend only on  $\theta$ .

Nevertheless, we need to stress that the objective of the model averaging approach is the prediction, and as such, the computation of the tail dependence parameters is not crucial on the analysis.

## 6.4 Our Multivariate Model and Literature

It is worth to discuss the works of Coles and Pauli (2002) and Apputhurai and Stephenson (2011), since they have some similarities with our Bayesian framework.

Coles and Pauli introduced a Bayesian model that includes both asymptotic dependence and independence. The model is constructed via the product of two bivariate survivor functions  $S_1$  and  $S_2$ , with uniform marginal distributions

as follows:

$$S(u_1, u_2) = (S_1(u_1, u_2))^q (S_2(u_1, u_2))^{1-q},$$

for  $0 \leq u_1, u_2 \leq 1$ .

If we define  $\chi = \lim_{u \rightarrow 1} \frac{S(u, u)}{1-u}$ , and similarly for  $\chi_1$  and  $\chi_2$ , then the tail dependence of the model would be given by:

$$\chi = \chi_1^q \chi_2^{1-q}.$$

Define, as well

$$\bar{\chi} = \lim_{u \rightarrow 1} \frac{2 \log(1-u)}{\log S(u, u)},$$

and similarly for  $\bar{\chi}_1$  and  $\bar{\chi}_2$ . Then, if  $S_1$  is chosen asymptotically dependent and  $S_2$ , asymptotically independent (meaning that  $\bar{\chi}_1 = 1$ , and  $\chi_2 = 0$ ), then  $\chi = 0$ , i.e., the model  $S$  is asymptotically independent, with

$$\bar{\chi} = \frac{q + (2-q)\bar{\chi}_2}{2-q+q\bar{\chi}_2},$$

unless  $q = 1$ , for which case,  $S$  is asymptotically dependent with  $\chi = \chi_1$ .

Coles and Pauli (2002) chose  $S_1$  and  $S_2$  as follows:

$$S_1(u_1, u_2) = ((1-u_1)^{1-k} + (1-u_2)^{1-k} - 1)^{\frac{1}{1-k}}, \quad k \geq 1,$$

$$S_2(u_1, u_2) = (1-u_1)(1-u_2).$$

Then, the estimation of the parameter  $q$  determines the tail dependence structure. This estimation is carried out via the Bayesian framework and MCMC methodology. Finally, they propose the following generalization:

$$S(\mathbf{u}) = \left( \sum_{j=1}^d (1-u_j)^{q_j(1-k)} + 1-d \right)^{\frac{1}{1-k}} \prod_{j=1}^d (1-u_j)^{1-q_j},$$

for  $\mathbf{u} = (u_1, \dots, u_d)$ , and  $0 \leq u_1, \dots, u_d \leq 1$ . Thus, the pair  $i$  and  $j$  are asymptotically dependent only when  $q_j = 1, \forall j$ . And, if that is the case and the pair  $i$  and  $m$  are as well asymptotically dependent, then it follows immediately that  $j$  and  $m$  are asymptotically dependent. This is not the case for our model in (6.31), since the weights of the factors involving the  $i, j$  and  $m$  elements can be 0, except for the following factors:

$$\left(1 + \alpha_i \left(x_i^{\frac{1}{\theta \xi_i}} - 1\right) + \alpha_j \left(x_j^{\frac{1}{\theta \xi_j}} - 1\right)\right)^{-\theta} \quad \text{and}$$

$$\left(1 + \alpha_i \left(x_i^{\frac{1}{\theta \xi_i}} - 1\right) + \alpha_j \left(x_m^{\frac{1}{\theta \xi_m}} - 1\right)\right)^{-\theta}.$$

We believe that the only similarity with our framework worth to discuss deeply is that of the bivariate case. Coles and Pauli (2002) managed to represent both asymptotically dependence and independence in one single model, via a parameter  $q$  that determines either form of tail dependence. The resulting model is the product of a tail dependent model (a Clayton one, which we know has a connection with our model, see Section 5.4) and a complete independent model, both of them to the  $q$ -th power. However, Coles and Pauli do not study the relation between the parameters  $q$  and  $k$ , which (as we know, since it is related to our parameter  $\theta$ ) determines as well the dependence structure.

Therefore, the option of representing both asymptotic dependences in one model is tempting. However, it involves the introduction of one more parameter which must be done more carefully, because of its relation with the parameter  $k$  (or, in our case,  $\theta$ ). In the case of our bivariate model, for example, the case where  $\theta \rightarrow \infty$  (exact independence), and  $q = 1$  cannot occur

simultaneously.

Hence, we opt to keep the dependence structure in one single parameter, with the consequence of representing asymptotic independence only by the complete tail independence.

On the other hand, Apputhurai and Stephenson (2011) introduced a model averaging framework for the bivariate case. Apputhurai and Stephenson selected an asymptotically dependent and two asymptotically independent models to average. The Reversible Jump methodology is followed for inference in a Bayesian framework.

A natural suggestion of the work of Apputhurai and Stephenson (2011) is to introduce an asymptotically independent model in our model averaging. However, introducing any asymptotic independent model would break the representation in (5.1) or (6.3), which are the basis of the construction of our model. Therefore, we opt to keep the dependence structure only on the parameter  $\theta$ .

## 6.5 Discussion of Multivariate Framework

In this section, we have presented the generalization of the bivariate model in (5.6). The resulting model is an averaging of models of variables represented via the common variable method, as in the bivariate case (see the expression (5.1), for example). The model is expressed in equation (6.31).

There are two ways to proceed under this model: the model selection and

the model averaging.

In the model selection approach, only one model is chosen, and this will determine the whole dependence structure. In the model averaging approach, prediction is the main objective of analysis. The main difference with the model selection is that the determination of the correct model is not crucial. Instead, we let all the combinations of different types of tail dependence to contribute to the joint probability.

We showed as well, that the Reversible Jump within a Bayesian framework is a reasonably accurate methodology to estimate the parameters and the weights. Therefore, we can proceed to real data applications.

# Chapter 7

## Illustrations

In this section we present real financial data analyses. In Section 7.1, we describe the data we are analyzing. We did two different types of analysis: the tail dependence determination analysis and the portfolio loss analysis.

The tail dependence determination analysis divides into two parts: a bivariate analysis presented in Section 7.2, and a trivariate analysis, presented in Section 7.2.2.

The portfolio analysis constitutes an application of prediction analysis. This analysis is presented in Section 7.3.

Finally, we close with the conclusions of the illustrations in 7.4.

### 7.1 Data Description

We carried out a financial data analysis similar as that used in Poon *et al.* (2004). The data consisted of daily return levels of five equity indices: French

CAC 40, German DAX, Japanese Nikkei 225, English FTSE 100, and US S&P 500. All of them run from April, 1993 until July, 2003.<sup>1</sup> The total number of observations is 2665. We plot in Figure D.15 an example of a series (CAC 40).

As in Poon *et al.* (2004), we took the previous day return for the S&P 500 observations, since the US market is the last one to close and any change on it will be more likely to influence on the very next day. In accordance with the work in Poon *et al.* (2004), we also applied the GJR-GARCH filter (see Glosten *et al.* (1993) for the explanation of it) in order to get rid of any high volatility clustering. Although we found that the application of this filter did not make a huge difference on the marginal tail behavior, we worked with the filtered data. In Figure D.16 we present the filtered and unfiltered data for the CAC 40 case.

In Section 2.1.3.1, we suggested the 0.96 quantile as a threshold for every variable. In Figure D.17, we presented the threshold selection tools for some of the indices.

Recall the threshold selection methods used the GPD as a univariate tail model. We assume this threshold sets a point where the extreme observations begin, rather than where a GPD must be fitted. In this sense, we use the same threshold selection, although the marginals are not GPD. In fact, we fix the threshold on the 0.95 quantile, to set it to the closest pragmatic value.

---

<sup>1</sup>They can be downloaded from MATLAB or directly from <http://finance.yahoo.com>.

The tickers symbols are: ^FCHI, ^GDAXI, ^N225, ^FTSE, and ^GSPC, respectively.

We split the data into negative and positive. For each part, we fixed the threshold as the 0.95 quantile of each Index. The region to study is:

$$RF = \{(0, u_{X_1^*}) \times \dots \times (0, u_{X_d^*})\}^c, \quad (7.1)$$

where  $u_{X_j^*}$  is the 0.95 quantile of the index  $X_j^*$ .

We did two types of analysis: the tail dependence determination and the portfolio analysis. On the last step before the fitting, we transformed the extreme observations as we stated in Section 5.5, for the bivariate case and in Section 6.2.3.2, for the trivariate case. The transformations are given by:

$$X_1 = X_1^* + 1, \dots, X_d = X_d^* + 1, \quad (7.2)$$

for  $(X_1^*, \dots, X_d^*) \in \{(0, u_{X_1^*}), \dots, (0, u_{X_d^*})\}^c$ , with  $d = 2$  in the bivariate case, and  $d = 3$  in the trivariate case.

## 7.2 Tail Dependence Determination Analysis

The aim of this section is to study the dependence among indices in the region  $RF$  in (7.1). This study splits into bivariate and trivariate analysis.

### 7.2.1 Bivariate Dependence Determination

We took all the different pairs of the filtered data for both the right and the left tail. We present the scatterplots of three pairs in Figure D.18. As the analysis will show further, and as the Figure D.19 can illustrate, the first two pairs represent medium tail dependence, whereas the last one, low tail dependence.

Inference was made via Metropolis-Hastings as described in Section 5.6.2, with the posterior as in (5.21). The prior distributions were unit exponential distributions, except for  $\theta$ , for which the prior was exponential distribution with hyperparameter  $\delta_\theta = \log(2)$ . We drew 20,000 posterior observations with a burn-in of 500. The estimation of each parameter was taken as the median of the posterior sampling. We present in Table C.11 the Bayesian estimations of the parameters for all the possible pairs of Indices. We give as well the 95% credibility intervals for both  $\theta$  and  $\chi$ .

Finally, to compare our results we did the following:

- We estimated  $\bar{\chi}$  with the methodology found in Ledford and Tawn (1996).
- We tested the asymptotic independence with Falk and Michel (2006) methodology, found in the *POT* package in *R*.
- When there is no evidence to accept asymptotic independence, we fit 5 models to data: the logistic, asymmetric logistic, negative logistic, asymmetric logistic, mixed and asymmetric mixed models. We used as well the *POT* package in *R*, which uses the MLE approach. We considered only the model with the lesser Akaike Information Criteria, and we give the estimate of  $\chi$  for this model.
- For those cases where the MLE presented problems to compute, we used the non-parametric estimator of  $\chi$  found in Poon *et al.* (2004).

As Table C.11 shows, we can classify the pairs of Indices regarding their

tail dependence structure as follows for both right and left tails:

- Medium tail dependence: France-UK, France-Germany, and Germany-UK;
- Low tail dependence: UK-US, France-US, Germany-US left tail, and Japan-US left tail;
- Presumably tail independence: Japan-UK, France-Japan, Germany-Japan, Germany-US right tail, and Japan-US right tail.

The results are in accordance with the estimates of  $\chi$  by Poon *et al.* (2004) methodology. In general, the censored approach estimates tend to underestimate the dependence. As well, the results agree regarding the economic view of the relationships among the countries involved. In fact, the results show a higher relation in the left tail (in comparison to the right tail) of the pairs Germany-US and Japan-US. This fact might show the effects of bad performance in the US index, in the German and Japanese markets. In fact, for all the cases but Japan-UK, the tail dependence in the left tail is larger than the tail dependence on the right tail, meaning that a bad performance is more likely to affect the rest of the countries, rather than a good performance.

## 7.2.2 Trivariate Dependence Determination

In this section, we analyze the same real data as in the previous section, but with the trivariate framework established in Section 6.2. We are interested

in which type of model is the one that has the heaviest weight on the model averaging. Therefore, the problem is treated as a model selection one.

The model selection is carried out via the Reversible Jump methodology as described in Section 6.2.3.3, which consisted in 10,000 jumps. The results are shown in Table C.12. As it can be seen, the following cases are in accordance with the bivariate results (compare it with Table C.11): France-Ger-Jap, France-Ger-US, France-Jap-UK, France-UK-US, Ger-Jap-UK, and Ger-UK-US.

The left tail of France-Jap-US is the only case to be chosen independent. On the other side, the right tail is selected as triple tail dependent. However, as the estimates of Table C.13 show, the tail dependence is low, as expected.

Finally, for the following cases, the selected model is the triple dependence model: France-Germany-UK, Ger-Jap-US, and Jap-UK-US. The former case is expected to be triple tail dependent. However, this was not the case for the two other cases. The estimates are shown in Figure C.13. This figure shows that the only group with medium triple tail dependence is the France-Germany-UK case, for both tails. For the rest of the cases, the estimates of  $\chi_3$  show a low triple tail dependence.

The analysis of this section show that the dependence estimated by the model is in accordance with a standard extreme methodology. As well, the results were in accordance with the economic interpretation of the data. Therefore, we can proceed to the next study, where we make a prediction analysis.

### 7.3 Portfolio Loss Analysis

This section deals with the minimization of the loss of an investment portfolio consisting of pairs of the indices data.

Suppose there is an investment portfolio  $L$  consisting of  $d$  assets returns, i.e. it is expressed as:

$$PS = \sum_{k=1}^d m_k Y_k^*,$$

where  $0 \leq m_k \leq 1$  is the portfolio allocation of asset  $k$ , i.e. the proportion invested in asset  $k$ , for  $k = 1, \dots, d$ , so that  $\sum_1^d m_k = 1$ . Then, the following issue might be of interest: given the historical data of the assets, what would be the asset allocation that minimizes the maximum loss of  $PS$ ?, or in other words, what is the asset allocation that minimizes the value at risk of the portfolio?

It is clear that we have to focus on the left tail of the variable  $PS$ , meaning, the negative changes on  $PS$ . Define the random variable  $L$  as the loss of the portfolio, i.e.  $L = -(PS)$ . Hence, the way to respond to this problem will be to find a combination of  $m_1, \dots, m_d$  such that  $P(L \geq u_L^*)$  is minimized, for a large value  $u_L^*$ . We will assume that a loss greater or equal to 90% of the portfolio is large enough to represent an extreme value.

For ease of explanation and without lose of generalization, we will assume the portfolio consists of two assets, so the loss  $L_m$  is given by:

$$L_m^* = mX_1^* + (1 - m)X_2^*,$$

for  $0 \leq m \leq 1$ , and where  $X_j^* = -Y_j^*$ , for  $j = 1, 2$ .

There are two regions to consider in this study. The first one is the extreme portfolio loss region  $RL_m^* = \{L_m \geq u_{L_m}^*\}$ , which will help us to define a high threshold. This region can be represented as:

$$RL_{m,u_{L_m}^*}^* = \{(X_1^*, X_2^*) : mX_1^* + (1-m)X_2^* \geq u_{L_m}^*\}, \quad (7.3)$$

where  $u_L^*$  is the 0.9 quantile of  $L$ . The form of  $RL_{m,u_{L_m}^*}^*$  is presented in Figure D.20, for a given  $m$ . The first aim of the analysis is to predict observations in  $RL_{m,u_{L_{0.5}}^*}^*$ .

Let the historical variable  $X$  be:

$$\mathbf{D}^* = \{(x_{1,i}^*, x_{2,i}^*), i = 1, \dots, n\}.$$

We assume the historical portfolio loss allocation has been fixed at  $m = 0.5$ , so that the historical loss  $\mathbf{L}_{0.5|\mathbf{D}^*}^*$  can be expressed as:

$$\mathbf{L}_{0.5|\mathbf{D}^*}^* = \left\{ 0.5 x_{1,i}^* + 0.5 x_{2,i}^* \right\}_{i=1}^n$$

for  $i = 1, \dots, n$ . In this way, we define the historical extreme portfolio loss region as:

$$RL_{0.5,u_{L_{0.5}}^*|\mathbf{D}^*}^* = \{(x_{1,i}^*, x_{2,i}^*) \in \mathbf{D}^* : 0.5 x_{1,i}^* + 0.5 x_{2,i}^* \geq u_{L_{0.5}}^*\}, \quad (7.4)$$

where  $u_{L_{0.5}}^*$  is the 0.9 quantile of  $\mathbf{L}_{0.5|\mathbf{D}^*}^*$ . Once the threshold has been set, the general extreme portfolio loss region is completely determined in (7.3) for a given  $m$ , and substituting  $u_{L_m}^*$  with  $u_{L_{0.5}}^*$ . Therefore, the next step is to predict observations in  $RL_{m,u_{L_{0.5}}^*}^*$ , for any  $m \in [0, 1]$ .

The second region to consider is where the data is taken from. We consider the observations inside the region:

$$RD_{u_{L_{0.5}}^*} = \{(0, u_{L_{0.5}}^*) \times (0, u_{L_{0.5}}^*)\}^c.$$

This means, the region the data is taken from is as in Figure D.1, with  $u_X = u_Y = u_{L_{0.5}}^*$ . This fact does not affect the inference, since  $RL_{m, u_{L_{0.5}}^*}^* \subset RD$ , for all  $m \in [0, 1]$ .

Since the aim of the analysis is the prediction, the data are transformed as in (7.2), i.e.:

$$x_{1,i} = x_{1,i}^* + 1, \text{ and } x_{2,i} = x_{2,i}^* + 1,$$

for  $(x_{1,i}^*, x_{2,i}^*) \in RD_{u_{L_{0.5}}^*}$ , for  $i = 1, \dots, n$ . Therefore, the transformed data we are modelling is given by:

$$\mathbf{D} = \left\{ (x_{1,i}, x_{2,i}) \in \{(1, u_{L_{0.5}}) \times (1, u_{L_{0.5}})\}^c \right\}_{i=1}^n.$$

where  $u_{L_{0.5}} = u_{L_{0.5}}^* + 1$ .

Then, the way to proceed is a modified version of the Reversible Jump methodology described in Section 6.2.3.3.

The joint model we will adjust to the transformed data will be that in (6.24), which reduces to the following expression for the bivariate case:

$$\begin{aligned} \bar{F}_X(\mathbf{x}) &= w_1 \left( x_1^{-\frac{\alpha_1}{\xi_1}} x_2^{-\frac{\alpha_2}{\xi_2}} \right) (1 - F_{M_1}(u_{L_{0.5}}))^{-1} \\ &\quad + w_2 (1 - F_{M_2}(u_{L_{0.5}}))^{-1} \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta \xi_1}} - 1 \right) + \alpha_2 \left( x_2^{\frac{1}{\theta \xi_2}} - 1 \right) \right)^{-\theta}, \end{aligned}$$

where

$$1 - F_{M_1}(u_{L_{0.5}}) = u_{L_{0.5}}^{-\frac{\alpha_1}{\xi_1}} + u_{L_{0.5}}^{-\frac{\alpha_2}{\xi_2}} - u_{L_{0.5}}^{-\frac{\alpha_1}{\xi_1}} u_{L_{0.5}}^{-\frac{\alpha_2}{\xi_2}},$$

and

$$\begin{aligned}
 1 - F_{M_2}(u_{L_{0.5}}) &= \left(1 + \alpha_1 \left(u_{L_{0.5}}^{\frac{1}{\theta \xi_1}} - 1\right)\right)^{-\theta} \\
 &\quad + \left(1 + \alpha_2 \left(u_{L_{0.5}}^{\frac{1}{\theta \xi_2}} - 1\right)\right)^{-\theta} \\
 &\quad - \left(1 + \alpha_1 \left(u_{L_{0.5}}^{\frac{1}{\theta \xi_1}} - 1\right) + \alpha_2 \left(u_{L_{0.5}}^{\frac{1}{\theta \xi_2}} - 1\right)\right)^{-\theta}.
 \end{aligned}$$

Since we have transformed the variables, we need to redefine the region we are sampling from as:

$$RL_{m,u_{L_{0.5}}} | \mathbf{D} = \{(x_{1,i}, x_{2,i}) \in \mathbf{D} : m x_{1,i} + (1 - m) x_{2,i} \geq u_{L_{0.5}}\}, \quad (7.5)$$

for a given  $m \in [0, 1]$ .

The modified Reversible Jump algorithm is as follows:

- Set a grid  $M$  in  $[0, 1]$ .
- At step  $t$ , propose a jump to model  $k^*$  (for  $k^* = 1, 2$ ).
- If the jump is rejected, sample  $X_{1,m}^S$  and  $X_{2,m}^S$  from model  $k_t$  in region  $RL_{m,u_{L_{0.5}}}$  (recall (7.5)) for every  $m \in M$ .
- If the jump is accepted, sample via Metropolis-Hastings a parameter vector  $\omega$  from model  $k^*$ , and then sample  $X_{1,m}^S$  and  $X_{2,m}^S$  from model  $k^*$  in region  $RL_{m,u_{L_{0.5}}}$ , for every  $m \in M$ , and set  $k_{t+1} = k^*$ .
- After a reasonable number of samplings  $(X_{1,m}^S, X_{2,m}^S)$  is achieved for every  $m \in M$ , assume it to converge to the predictive joint distribution of  $(X_1, X_2)$  in region  $RL_{m,u_{L_{0.5}}}$ , and define the sampling as  $\mathbf{X}_m^S = (X_{1,m}^S, X_{2,m}^S)$ .

- Once the sampling has been achieved, define the predictive loss  $L_m^S$  as follows:

$$L_m^S = m X_{1,m}^S + (1 - m)X_{2,m}^S, \quad (7.6)$$

for every  $m \in M$ .

- Select  $m^* \in M$ , where  $m^*$  minimizes the median of  $L_m^S$ .

We chose a sample size of  $2.5 \times 10^5$  and an interval width of 0.02 for the grid  $M$ . We present in Figure D.21 the graphic of the median of  $L_m^S$  for every  $m \in M$ , for each pair of indices. Although the graphics do not show a quite smooth behavior, the tendency is clear in all the cases. In order to get a smoother version, the sample must be increased and the width of the grid decreased. However, we did not consider an excessive smoothness an important feature, given the average computing time, which was 184.5 minutes per pair of Indices. In Table C.14 we present the portfolio allocation that minimizes the median of  $L_m^S$ .

In Table C.11, we showed the three strongest dependencies were given by the pairs: France-UK, France-Germany, and Germany-UK. In all of these pairs, the optimum allocation suggests a 100-0 allocation, except for the pair France-Germany, which was a 91-9 allocation. This fact makes economic sense, since if two assets have a strong dependence, then the minimum loss is achieved when allocating the most on the asset with less expected loss.

Consider the cases: France-UK, France-Germany, and Germany-UK. In

order to show which Index has a larger expected loss, we did the following: we computed the marginal predictive mean excess over the threshold  $u_{L_{m^*}}$  for each Index of each pair. This means,  $u_{L_{m^*}}$  varies from pair to pair. Therefore, the marginal predictive mean excess is different for the same Index from pair to pair. Finally, the results are transformed to the original change of the Indices. We show in Table C.15 the results for these cases. The results confirm that indeed the Index with the lesser predictive excess loss is the one that takes the higher allocation.

On the other hand, for those cases where the tail dependence was determined as low in the previous section, the optimum allocation is split not as extremely as in the medium tail dependent pairs.

In order to verify the fit of the predictions, we made a comparison of histograms. Suppose  $m^*$  is the optimum allocation. Then, on one hand, we present the empirical histogram of the observations in region  $RL_{m^*, u_{L_{0.5}^*}} | \mathcal{D}$  (see (7.4)) for the transformed data. This means, the historical transformed observations for which the optimum allocation portfolio exceeds  $u_{L_{0.5}^*}$ . On the other hand, we present the histogram of the samplings inside the region  $RL_{m^*, u_{L_{0.5}^*}}^S = \{L_{m^*}^S > u_{L_{0.5}^*}^*\}$  (recall expression (7.6)).

The histograms are shown in Figure D.22 for four pairs of Indices. The predictive histograms show quite similar shapes and levels to the historical histograms. This fact suggests a good fit of the predictions.

Therefore, we can proceed to define the generalization of the portfolio al-

location. Assume the analysis consists of  $d$  assets returns , which are:

$$(Y_1^*, \dots, Y_d^* \in \mathbb{R}^d).$$

We only consider the negative returns of the assets, and define the variables

$$X_j^* = -Y_j^*, \text{ for } j = 1, \dots, d.$$

The general loss function is expressed as:

$$L_{\mathbf{m}}^* = \sum_{j=1}^d m_j X_j^*,$$

for  $\mathbf{m} = (m_1, \dots, m_d)$ , and where  $0 \leq m_j \leq 1$ , for  $j = 1, \dots, d$ , and  $m_d = 1 - \sum_1^d m_j$ .

Let the historical variable  $X$  be:

$$\mathbf{D}^* = \{(x_{1,i}^*, \dots, x_{d,i}^*), i = 1, \dots, n\}.$$

The, the historical loss function for a given allocation  $\mathbf{m}$  is given by:

$$L_{\mathbf{m}|\mathbf{D}^*}^* = \left\{ \sum_{j=1}^d m_j x_{j,i}^* \right\}_{i=1}^n,$$

for  $\mathbf{m} = (m_1, \dots, m_d)$ , and where  $0 \leq m_j \leq 1$ , for  $j = 1, \dots, d$ , and  $m_d = 1 - \sum_1^d m_j$ .

Assume a fixed historical portfolio loss allocation at  $\mathbf{m}_0 = (m_{1_0}, \dots, m_{d_0})$ ,

with  $0 \leq m_{j_0} \leq 1$ , for  $j = 1, \dots, d$ , and  $m_{d_0} = 1 - \sum_1^d m_{j_0}$ .

Then, the extreme portfolio region is defined as:

$$RL_{\mathbf{m}_0, u_{L_{\mathbf{m}_0}^*}}^* | \mathbf{D}^* = \left\{ (x_1^*, \dots, x_d^*) : m_{1_0} x_1^* + \dots + m_{d_0} x_d^* \geq u_{L_{\mathbf{m}_0}^*}^* \right\},$$

where  $u_{L_{m_0}}^*$  is a high quantile of  $L_{m_0}$ , say 0.9. The aim is to predict observations in  $RL_{m, u_{L_{m_0}}^*}$ , for a given allocation  $m$ .

The data is taken from the following region:

$$RD_{u_{L_{m_0}}^*} = \left\{ \left( 0, u_{L_{m_0}}^* \right) \times \dots \times \left( 0, u_{L_{m_0}}^* \right) \right\}^c.$$

The next step is to transform the variables as follows:

$$x_{1,i} = x_{1,i}^* + 1, \dots, x_{d,i} = x_{d,i}^* + 1,$$

for  $(x_{1,i}^*, \dots, x_{d,i}^*) \in RD_{u_{L_{m_0}}^*}$ . In this way, the region we are modelling the transformed data is given by:

$$RD_{u_{L_{m_0}}} = \left\{ \left( 1, u_{L_{m_0}} \right) \times \dots \times \left( 1, u_{L_{m_0}} \right) \right\}^c.$$

On the other hand, the region we are sampling from is given by:

$$RL_{m, u_{L_{m_0}}} | D = \left\{ (x_1, \dots, x_d) : m_1 x_1 + \dots + m_d x_d \geq u_{L_{m_0}} \right\}.$$

Then, the way to proceed is a modified version of the Reversible Jump as in the bivariate portfolio case. The joint model that underpins the algorithm is the multivariate model averaging procedure presented in Section 6.3. The following step is to construct a joint predictive sampling distribution from this multivariate joint model.

The modified Reversible Jump algorithm is as follows:

- Set a grid  $M$  in  $[0, 1]^d$ .
- At step  $t$ , propose a jump to model  $k^*$  (for  $k^* = 1, \dots, d_T$ , where  $d_T$  is as in (6.32)).

- If the jump is rejected, sample  $X_{1,m}^S, \dots, X_{d,m}^S$  from model  $k_t$  in region  $RL_{m,u_{L_{m_0}}}$ , for every  $m \in M$ .
- If the jump is accepted, sample via Metropolis-Hastings a parameter vector  $\omega$  from model  $k^*$ , and then sample  $X_{1,m}^S, \dots, X_{d,m}^S$  from model  $k^*$  in region  $RL_{m,u_{L_{m_0}}}$ , for every  $m \in M$ , and set  $k_{t+1} = k^*$ .
- After a reasonable number of samplings  $(X_{1,m}^S, \dots, X_{d,m}^S)$  is achieved for every  $m \in M$ , assume it to converge to the predictive joint distribution of  $(X_1, \dots, X_d)$  in region  $RL_{m,u_{L_{m_0}}}$ , and define the sampling as  $\mathbf{X}_m^S = (X_{1,m}^S, \dots, X_{d,m}^S)$ .
- Once the sampling has been achieved, define the predictive loss  $L_m^S$  as follows:

$$L_m^S = m_1 X_{1,m}^S + \dots + m_d X_{d,m}^S,$$

for every  $m \in M$ .

- Select  $m^* \in M$ , where  $m^*$  minimizes the median of  $L_m^S$ .

## 7.4 Conclusions

This section presented some applications of the models developed in Chapters 5 and 6.

The first illustration was a bivariate application. We fitted the bivariate model to financial data, in order to determine the tail dependence. In general

terms, the estimates looked in accordance with extremes methodologies, as that in Poon *et al.* (2004).

The multivariate model we are proposing is a mixture model. The application of this model depends on the problem, whether it is a model selection or model averaging. We viewed the model selection as a tail dependence determination.

For the trivariate case, we did a model selection application, which is translated as a tail dependence structure determination. The results of this analysis showed a behavior in accordance with was found in the bivariate analysis.

Both the bivariate and trivariate analysis showed acceptable results from an economical perspective.

We did a prediction application in the portfolio analysis section. We found convenient to choose the model averaging interpretation of the mixture model, since it spreads the uncertainty among different models, rather than selecting only one of them. The methodology we proposed was the Reversible Jump with a Metropolis-Hasting MCMC sampling. We proposed an algorithm to sample predictions of each plausible model. Although the prediction application was limited to the bivariate case, we presented the straightforward generalization of the algorithm. The predictions of the mixture model helped us to solve a portfolio analysis problem. The predictions were joined to describe the behavior of the portfolio loss variable. We showed the fit of this procedure was accurate, in accordance with the data. Therefore, the model and methodolo-

gies we presented seem to be flexible enough to transcend to other type of applications via the prediction.

Hence, we find that the main purpose of the mixture model we present is the prediction, since it is on that when the model becomes more flexible, rather than selecting one model. Beside, the majority of applications or problems can be attacked, from a Bayesian perspective, with prediction.

# Chapter 8

## Conclusions

### 8.1 Final Discussion

The aim of the dissertation was to present a multivariate model for extremes.

We opted for a mixture of models with different asymptotic dependencies.

The first objective was to construct a parametric family of models with the following features:

- Interpretable parametric form.
- Sufficient asymptotic dependence structure.
- Fitting a realistic range of tails.

In Chapter 4, we introduced some models constructed via the Common Variable method. Studying their properties, we found they cannot fit all of the four requirements.

In Chapter 5, we presented a modification of the models, where all the requirements were achieved. The marginal distributions are quasi Burr-XII type. An important feature was that the tail dependence relies only on one single parameter,  $\theta$ . The parametric form of the bivariate model is simple and inference was developed with Bayesian methods. The tail dependence structure is as follows, for  $X_1$  and  $X_2$ :

- if  $\theta \rightarrow \infty$ , then  $X_1$  and  $X_2$  are tail independent, in fact, they are exactly tail independent.
- if  $\theta \rightarrow 0$ , then  $X_1$  and  $X_2$  are completely tail dependent.

Anywhere else, the tail dependence measure  $\chi$  has the following form:  $\chi = 2^{-\theta}$ . In the exact independent case, the model factorizes into the product of Pareto marginals.

We want to make clear that we are aware that our model only fits tail dependence and exact independence, and no other type of tail independence is covered. However, we believe this model is a valid first attempt of an individual component of the mixture model.

Another important feature of the bivariate model is the simplistic construction of dependent variables. This feature made possible the introduction of the triple dependent model, and in general, the  $m$ -tuple dependent model. As in the bivariate case, in every  $m$ -tuple dependent model, the asymptotic dependence is governed exclusively by  $\theta$ .

We introduce the multivariate model in Chapter 6, i.e. the mixture of models with different possible combinations of dependencies. In principle, this might seem to drop the simplicity of the bivariate model. However, each individual model of the averaging constitutes a simple version of a parametric model constructed via the common variable method, as in the bivariate case.

Each individual model splits into two parts: the tail dependence and the exact independence parts. We presented a model where the tail dependence part is an  $m$ -tuple dependent model constructed as in the bivariate case. On the other hand, the exact independent part consists of the product of univariate quasi Burr-XII type models. In this way, this model has marginal coherence.

Every mixture model can be interpreted in two ways: as a model selection or as a model averaging problem.

The model selection approach assumes only one model is correct. Therefore, its objective is to estimate the model that fits the best the data, i.e. only one model describes the data. This turns out to be quite restrictive, since there is only one measure of asymptotic dependence for all the asymptotic dependent variables. Hence, it is not possible to model different strengths of asymptotic dependence among different subsets of variables.

On the other hand, the model averaging approach recognizes the uncertainty of the problem and spreads it over the individual components. Therefore, no individual model is assumed to be the only one to describe data. The

asymptotic dependence structure of this approach results quite flexible, since for different subsets of the vector of variables, different asymptotic dependencies can be achieved.

Hence, we strongly opt for the model averaging approach. The only reason why we used the model selection approach was to compare our results in Chapter 7 with standard procedures in extremes.

When coming to inference, we used the reversible jump methodology within the Bayesian paradigm. MCMC methods were the basic tool for parameter estimations. Within this inference procedure, we found that a modification of the model facilitates the whole procedure. Instead of the product of quasi Burr-XII type marginals, we replaced it by the product of Pareto marginals. From Chapter 5, we know these two models are equivalent marginal models in the independent case. Therefore, the asymptotic dependence of the mixture model is respected.

Finally, in Chapter 7, we presented some applications in the financial area. We performed an analysis of tail dependences among some indices. The model seems to fit well enough to the returns, even though their magnitude is small (recall the tails we are aiming are non-light tails). The dependence structure measured by  $\theta$ , seems to identify the relation in tails of the variables, for both the bivariate and the trivariate case.

In the last part of the applications, we presented a methodology to minimize the loss of a portfolio, using the multivariate model of Chapter 6. Although

the results are shown only for two assets portfolios, it can be easily generalized to any dimension.

## 8.2 Future Work

We finish presenting the potential topics that could follow this work.

Firstly, a generalization of the marginals would be needed to embrace light tails. This means, variables for which all the moments are finite.

A second issue is how to get all the forms of tail independence inside the mixture model, rather than only the exact independence.

The next topic is the application of the multivariate model to a large set of variables (say more than five). This study would get closer to a real portfolio application. The algorithm we introduced in Section 7.3 turned out to be time consuming. The issue of the declaration of a grid for the allocation is potentially a problem as the dimension increases. Therefore, the special task on this issue would be the way to minimize the time of the computations.

We saw in Chapter 6 that the asymptotic dependence information is given by both the weights and  $\theta$ . Therefore, an important topic is to study the relation between both.

Finally, it would be interesting to relate the Common Variable construction with some special topics found in the literature: different tail independence structures, concomitant tail behavior, spatial extreme statistics, and the study of negatively associated variables.

# Appendix A

## Regular Variation

Regular variation constitutes a wide field and its applications embrace a lot of areas. Since we are dealing with extremes, there is a natural connection between the maximum and the tail of the distribution. Information on the tail of a distribution implies information on the maximum and vice versa. Regular variation is a property of some functions and its rate of decay. We will only present the concepts that will help us to explain the extremes theory. The works in Bingham *et.al.* (1987) and de Haan (1970) treat extensively this topic.

**Definition A.0.1.** *Regular Variation*

1. A measurable function  $L : \mathbb{R} \rightarrow \mathbb{R}_+$  is said to be slow varying at  $\infty$ , i.e.

$L \in RV_0$ , if for  $x > 0$ :

$$\lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1.$$

2. A measurable function  $H : \mathbb{R} \rightarrow \mathbb{R}_+$  is said to be regular varying at  $\infty$  with index  $\tau$ , i.e.  $H \in RV_\tau$ , if for  $x > 0$ :

$$\lim_{t \rightarrow \infty} \frac{H(tx)}{H(t)} = x^\tau. \quad (\text{A.1})$$

We present some examples as relevant illustrations of tail behaviors.

**Example A.0.1.** Let  $H(t) = \log(t^c)$ , for  $c \in \mathbb{R}$ . Then, for  $x > 0$ :

$$\lim_{t \rightarrow \infty} \frac{H(tx)}{H(t)} = \lim_{t \rightarrow \infty} \frac{\log(t) + \log(x)}{\log(t)} = 1.$$

Therefore,  $H(x) \in RV_0$ .

**Example A.0.2.** Let  $H(t) = \exp(-t^c)$ , for  $c \in \mathbb{R}$ . Then, for  $x > 0$ :

$$\lim_{t \rightarrow \infty} \frac{\exp(-t^c x^c)}{\exp(-t^c)} = \lim_{t \rightarrow \infty} \exp(-t^c(x^c - 1)) = \begin{cases} 0 & \text{if } x > 1 \\ \infty & \text{if } 0 < x < 1. \end{cases}$$

Therefore,  $H(x)$  is not a regularly varying function.

**Example A.0.3.** Let  $H(t) = t^c \log(t)$ , for  $c \in \mathbb{R}$ . Then, for  $x > 0$ :

$$\lim_{t \rightarrow \infty} \frac{t^c x^c (\log(t) + \log(x))}{t^c \log(t)} = x^c.$$

Therefore,  $H(x) \in RV_c$ .

Regularly varying functions have a special representation (see Resnick (1987) and Bingham *et.al.* (1987)). If we define a slow varying function  $L(x)$ , then, any regularly varying function  $H$  can be written as:

$$H(x) = x^\tau L(x), \quad (\text{A.2})$$

for  $\tau \in \mathbb{R}$ .

# Appendix B

## Proofs

### B.1 Proof of Theorem 4.1.1

We seek to find an asymptotic expression of the joint survival (4.8). In order to do so, we need to find an expression of the inverse of the marginal distributions, so that we can compute the function  $G(s)$  as in (4.7). However, it is not possible to have a close form for the inverse, so we use the asymptotic form of the marginal in (4.6).

$$F_X(x) = \begin{cases} 1 - \left(\frac{\alpha}{\alpha-1}\right) e^{-x} & \text{for } \alpha > 1 \\ 1 - (x+1)e^{-x} & \text{for } \alpha = 1 \\ 1 - \left(\frac{1}{1-\alpha}\right) e^{-\alpha x} & \text{for } 0 < \alpha < 1. \end{cases} \quad \text{as } x \rightarrow \infty.$$

We will focus our attention to the  $\alpha \neq 1$  cases, since subsequently we will

be able to distinguish the  $\alpha = 1$  case. With this representation, we can write down the asymptotic form of the inverse:

$$F_X^+(x) = \begin{cases} -\log\left(\left(\frac{\alpha-1}{\alpha}\right)(1-x)\right), & \alpha > 1 \\ -\frac{1}{\alpha}\log((\alpha-1)(1-x)), & 0 < \alpha < 1 \end{cases} \quad \text{as } x \rightarrow \infty.$$

With this result, it is possible to compute the asymptotic form of the  $G$  function as follows :

$$G(s) = \begin{cases} -\log\left(\left(\frac{\alpha-1}{\alpha}\right)(1-e^{-1/s})\right), & \alpha > 1 \\ -\frac{1}{\alpha}\log((\alpha-1)(1-e^{-1/s})), & 0 < \alpha < 1 \end{cases} \quad \text{as } s \rightarrow \infty.$$

However, we know that

$$(1 - \exp(-1/s)) = s^{-1} + o(s^{-1}), \text{ as } s \rightarrow \infty, \tag{B.1}$$

then, the  $G$  function can be expressed as:

$$G(s) = \begin{cases} -\log\left(\left(\frac{\alpha-1}{\alpha}\right)s^{-1}\right), & \alpha > 1 \\ -\frac{1}{\alpha}\log((\alpha-1)s^{-1}), & 0 < \alpha < 1 \end{cases} \quad \text{as } s \rightarrow \infty. \tag{B.2}$$

Since we have different representations for both the  $G(s)$  and the joint survival function, we will write down the asymptotic form of the latter in separated cases. In all of them we will recall (B.2) and the joint distribution in (4.8).

For the  $0 < \alpha < 1$  case we have the following result:

$$\begin{aligned}
 P(S > s, T > s) &= \frac{\alpha}{\alpha - 2} e^{-2\left(-\frac{1}{\alpha} \log((1-\alpha)s^{-1})\right)} - \frac{2}{\alpha - 2} e^{-\alpha\left(-\frac{1}{\alpha} \log((1-\alpha)s^{-1})\right)} \\
 &= \frac{\alpha}{\alpha - 2} (1 - \alpha)^{\frac{2}{\alpha}} (s^{-1})^{\frac{2}{\alpha}} - \frac{2}{\alpha - 2} (1 - \alpha) s^{-1} \\
 &= \frac{\alpha}{\alpha - 2} (1 - \alpha)^{\frac{2}{\alpha}} s^{-\frac{2}{\alpha}} - \frac{2}{\alpha - 2} (1 - \alpha) s^{-1} \\
 &= 2 \frac{(1 - \alpha)}{(2 - \alpha)} s^{-1}, \text{ as } s \rightarrow \infty.
 \end{aligned} \tag{B.3}$$

Hence, we can get the expression for the tail dependence:

$$\chi = \frac{2(1 - \alpha)}{(2 - \alpha)}.$$

It is worth to note that in (B.3), we can distinguish the coefficient of tail dependence  $\eta$ , (recall (2.28)), which is equal to one. If we find a case where  $\eta < 1$ , then, we can assure that  $\chi = 0$ .

In the  $\alpha \geq 1$  and  $\alpha \neq 2$  case, we have the joint survival function:

$$\begin{aligned}
 P(S > s, T > s) &= \frac{\alpha}{\alpha - 2} e^{-2\left(-\log\left(\frac{\alpha-1}{\alpha} s^{-1}\right)\right)} - \frac{2}{\alpha - 2} e^{-\alpha\left(-\log\left(\frac{\alpha-1}{\alpha} s^{-1}\right)\right)} \\
 &= \frac{\alpha}{\alpha - 2} \left(\frac{\alpha - 1}{\alpha}\right)^2 s^{-2} - \frac{2}{\alpha - 2} \left(\frac{\alpha - 1}{\alpha}\right)^\alpha s^{-\alpha} \\
 &= \frac{(\alpha - 1)^2}{(\alpha - 2)\alpha} s^{-2} - \frac{2}{\alpha - 2} \left(\frac{\alpha - 1}{\alpha}\right)^\alpha s^{-\alpha}, \text{ as } s \rightarrow \infty.
 \end{aligned}$$

We need to look at the asymptotic dominance of any of these two factors.

For the  $1 \leq \alpha < 2$  case the joint survival reduces to:

$$P(S > s, T > s) = \frac{2}{2 - \alpha} \left(\frac{\alpha - 1}{\alpha}\right)^\alpha s^{-\alpha}, \text{ as } s \rightarrow \infty.$$

Which leads to a tail coefficient  $\eta > 1$  and the following expression for the tail dependence:

$$\chi = \lim_{s \rightarrow \infty} \frac{2}{2 - \alpha} \left( \frac{\alpha - 1}{\alpha} \right)^\alpha s^{-\alpha+1} = 0,$$

The next case is when  $\alpha > 2$ :

$$P(S > s, T > s) = \frac{(\alpha - 1)^2}{(\alpha - 2)\alpha} s^{-2}, \text{ as } s \rightarrow \infty.$$

which leads the same  $\chi$  as in the previous case, i.e.

$$\chi = \lim_{s \rightarrow \infty} \frac{(\alpha - 1)^2}{(\alpha - 2)\alpha} s^{-1} = 0.$$

Finally, when  $\alpha = 2$  (recalling (B.2) and (4.8)):

$$\begin{aligned} P(S > s, T > s) &= -2 \log \left( \frac{\alpha - 1}{\alpha} s^{-1} \right) e^{-2(-\log(\frac{\alpha-1}{\alpha} s^{-1}))} \\ &= -2 \log \left( \frac{\alpha - 1}{\alpha} s^{-1} \right) \left( \frac{\alpha-1}{\alpha} \right)^2 s^{-2}, \text{ as } s \rightarrow \infty, \end{aligned}$$

then:

$$\chi = 0.$$

Therefore, the relation in (4.9) holds.

## B.2 Proof of Theorem 4.1.2

As in the previous proof, we seek to find an asymptotic form for  $P(S > s, T > s)$ . Given the form of the marginals in (4.11), it is not possible to have a close

form for the inverse functions, then we have to use the asymptotic form of these functions.

$$F_X(x) \approx \begin{cases} 1 + \frac{\beta}{\alpha - \beta} e^{-\alpha x} & 0 < \alpha < \beta \\ 1 - \alpha x e^{-\alpha x} & \alpha = \beta \\ 1 - \frac{\alpha}{\alpha - \beta} e^{-\beta x} & \alpha > \beta \end{cases} \quad \text{as } x \rightarrow \infty.$$

$$F_Y(y) \approx \begin{cases} 1 + \frac{\gamma}{\alpha - \gamma} e^{-\alpha y} & 0 < \alpha < \gamma \\ 1 - \alpha y e^{-\alpha y} & \alpha = \gamma \\ 1 - \frac{\alpha}{\alpha - \gamma} e^{-\gamma y} & \alpha > \gamma \end{cases} \quad \text{as } y \rightarrow \infty.$$

Similarly to the simple exponential case, we will focus on the  $\alpha \neq \beta, \gamma$  cases, since later on we will be able to distinguish the  $\alpha = \beta$  and  $\alpha = \gamma$  results. We write the inverse functions down:

$$F_X^{\leftarrow}(x) \approx \begin{cases} -\frac{1}{\alpha} \log \left( \frac{\alpha - \beta}{\beta} (x - 1) \right), & 0 < \alpha < \beta \\ -\frac{1}{\beta} \log \left( \frac{\alpha - \beta}{\alpha} (1 - x) \right), & \alpha > \beta \end{cases} \quad \text{as } x \rightarrow \infty.$$

$$F_Y^{\leftarrow}(y) \approx \begin{cases} -\frac{1}{\alpha} \log \left( \frac{\alpha - \gamma}{\gamma} (y - 1) \right), & 0 < \alpha < \gamma \\ -\frac{1}{\gamma} \log \left( \frac{\alpha - \gamma}{\alpha} (1 - y) \right), & \alpha > \gamma \end{cases} \quad \text{as } y \rightarrow \infty.$$

Thus, we can define the  $G$  functions using (B.1).

$$G_X(s) \approx \begin{cases} -\frac{1}{\alpha} \log\left(\frac{\beta - \alpha}{\beta} s^{-1}\right), & 0 < \alpha < \beta \\ -\frac{1}{\beta} \log\left(\frac{\alpha - \beta}{\alpha} s^{-1}\right), & \alpha > \beta \end{cases} \quad \text{as } s \rightarrow \infty. \quad (\text{B.4})$$

$$G_Y(s) \approx \begin{cases} -\frac{1}{\alpha} \log\left(\frac{\gamma - \alpha}{\gamma} s^{-1}\right), & 0 < \alpha < \gamma \\ -\frac{1}{\gamma} \log\left(\frac{\alpha - \gamma}{\alpha} s^{-1}\right), & \alpha > \gamma \end{cases} \quad \text{as } s \rightarrow \infty. \quad (\text{B.5})$$

In order to compute the joint survival function, we need to show that  $G_X(s) < G_Y(s)$ , when  $s \rightarrow \infty$ . Recall that  $\gamma < \beta$ , then, the first case is when  $\alpha < \gamma < \beta$ :

$$\begin{aligned} G_X(s) < G_Y(s) &\iff -\frac{1}{\alpha} \log\left(\frac{\beta - \alpha}{\beta} s^{-1}\right) < -\frac{1}{\alpha} \log\left(\frac{\gamma - \alpha}{\gamma} s^{-1}\right) \\ &\iff 1 - \frac{\alpha}{\beta} > 1 - \frac{\alpha}{\gamma} \\ &\iff \beta > \gamma. \end{aligned}$$

When  $\gamma < \beta < \alpha$ :

$$\begin{aligned} G_X(s) < G_Y(s) &\iff -\frac{1}{\beta} \log\left(1 - \frac{\beta}{\alpha}\right) + \frac{1}{\beta} \log(s) < -\frac{1}{\gamma} \log\left(1 - \frac{\gamma}{\alpha}\right) + \frac{1}{\gamma} \log(s) \\ &\iff \frac{1}{\beta} \log(s) < \frac{1}{\gamma} \log(s) \\ &\iff \beta > \gamma; \end{aligned}$$

as  $s \rightarrow \infty$ , whereas, when  $\gamma < \alpha < \beta$ .

$$\begin{aligned} G_X(s) < G_Y(s) &\iff -\frac{1}{\alpha} \log\left(1 - \frac{\alpha}{\beta}\right) + \frac{1}{\alpha} \log(s) < -\frac{1}{\gamma} \log\left(1 - \frac{\gamma}{\alpha}\right) + \frac{1}{\gamma} \log(s) \\ &\iff \frac{1}{\alpha} \log(s) < \frac{1}{\gamma} \log(s) \end{aligned}$$

as  $s \rightarrow \infty$ .

Therefore:

$$G_X(s) < G_Y(s), \text{ as } s \rightarrow \infty.$$

Hence, it is possible to express the joint survival function as follows:

$$\begin{aligned} P(S > s, T > s) &= P(B > G_X(s) - A, C > G_Y(s) - A) \\ &= \int P(B > G(s) - a, C > G(s) - a | A) f_A(a) da \\ &= \int_0^{G_X(s)} e^{-\beta(G_X(s)-a)} e^{-\gamma(G_Y(s)-a)} \alpha e^{-\alpha a} da \\ &\quad + \int_{G_X(s)}^{G_Y(s)} e^{-\gamma(G_Y(s)-a)} \alpha e^{-\alpha a} da + \int_{G_Y(s)}^{\infty} \alpha e^{-\alpha a} da \\ &= \alpha e^{-\beta G_X(s) - \gamma G_Y(s)} \left( \frac{1}{\alpha - \beta - \gamma} \right) (1 - e^{-(\alpha - \beta - \gamma)G_X(s)}) \\ &\quad + \alpha e^{-\gamma G_Y(s)} \left( \frac{1}{\alpha - \gamma} \right) (e^{-(\alpha - \gamma)G_X(s)} - e^{-(\alpha - \gamma)G_Y(s)}) + e^{\alpha G_Y(s)} \\ &= \frac{\alpha}{\alpha - \beta - \gamma} e^{-\beta G_X(s) - \gamma G_Y(s)} - \frac{\gamma}{\alpha - \gamma} e^{\alpha G_Y(s)} \\ &\quad - \frac{\beta \alpha}{(\alpha - \gamma)(\alpha - \beta - \gamma)} e^{-\gamma G_Y(s) - (\alpha - \gamma)G_X(s)}, \end{aligned}$$

for  $\alpha \neq \gamma, \alpha \neq \gamma + \beta$ .

We divide again the study of the joint distribution in cases. Those cases where  $\eta > 1$  will lead to  $\chi = 0$ , so that we will compute  $\chi$  only when  $\eta = 1$ .

When  $\alpha < \gamma < \beta$ , the  $G$  functions are given by:

$$G_X(s) \approx -\frac{1}{\alpha} \log \left( 1 - \frac{\alpha}{\beta} \right) + \frac{1}{\alpha} \log(s), \text{ and } G_Y(s) \approx -\frac{1}{\alpha} \log \left( 1 - \frac{\alpha}{\gamma} \right) + \frac{1}{\alpha} \log(s)$$

as  $s \rightarrow \infty$ , thus, the survival function is given by:

$$\begin{aligned} P(S > s, T > s) &\approx \frac{\alpha}{\alpha-\beta-\gamma} \left(1 - \frac{\alpha}{\beta}\right)^{\beta/\alpha} s^{-\beta/\alpha} \left(1 - \frac{\alpha}{\gamma}\right)^{\gamma/\alpha} s^{-\gamma/\alpha} - \frac{\gamma}{\alpha-\gamma} \left(\frac{\gamma-\alpha}{\gamma}\right) s^{-1} \\ &\quad - \frac{\beta\alpha}{(\alpha-\gamma)(\alpha-\beta-\gamma)} \left(1 - \frac{\alpha}{\gamma}\right)^{\gamma/\alpha} s^{-\gamma/\alpha} \left(1 - \frac{\alpha}{\beta}\right)^{1-\gamma/\alpha} s^{\gamma/\alpha-1}, \\ &\approx K s^{-\frac{\beta+\gamma}{\alpha}} + s^{-1} - \frac{\beta\alpha}{(\alpha-\gamma)(\alpha-\beta-\gamma)} \left(1 - \frac{\alpha}{\gamma}\right)^{\gamma/\alpha} \left(1 - \frac{\alpha}{\beta}\right)^{1-\gamma/\alpha} s^{-1}, \end{aligned}$$

as  $s \rightarrow \infty$ , for a constant  $K$ , but, since  $\beta + \gamma > \alpha$ , the expression reduces to:

$$P(S > s, T > s) \approx \left[1 - \frac{\beta\alpha}{(\gamma-\alpha)(\beta+\gamma-\alpha)} \left(1 - \frac{\alpha}{\gamma}\right)^{\gamma/\alpha} \left(1 - \frac{\alpha}{\beta}\right)^{1-\gamma/\alpha}\right] s^{-1},$$

as  $s \rightarrow \infty$ , so that  $\eta = 1$  and the tail dependence is

$$\chi = 1 - \frac{\beta\alpha}{(\gamma-\alpha)(\beta+\gamma-\alpha)} \left(1 - \frac{\alpha}{\gamma}\right)^{\gamma/\alpha} \left(1 - \frac{\alpha}{\beta}\right)^{1-\gamma/\alpha},$$

then  $0 < \chi < 1$ , as required.

The following case is when  $\gamma < \beta < \alpha$ , where the  $G$  functions are:

$$G_X(s) \approx -\frac{1}{\beta} \log\left(1 - \frac{\beta}{\alpha}\right) + \frac{1}{\beta} \log(s), \text{ and } G_Y(s) \approx -\frac{1}{\gamma} \log\left(1 - \frac{\gamma}{\alpha}\right) + \frac{1}{\gamma} \log(s)$$

as  $s \rightarrow \infty$ , thus, the survival function is given by:

$$\begin{aligned} P(S > s, T > s) &\approx \frac{\alpha}{\alpha-\beta-\gamma} \left(1 - \frac{\beta}{\alpha}\right) s^{-1} \left(1 - \frac{\gamma}{\alpha}\right) s^{-1} - \frac{\gamma}{\alpha-\gamma} \left(1 - \frac{\gamma}{\alpha}\right)^{\alpha/\gamma} s^{-\alpha/\gamma} \\ &\quad - \frac{\beta\alpha}{(\alpha-\gamma)(\alpha-\beta-\gamma)} \left(1 - \frac{\gamma}{\alpha}\right) s^{-1} \left(1 - \frac{\beta}{\alpha}\right)^{(\alpha-\gamma)/\beta} s^{-(\alpha-\gamma)/\beta}, \\ &\approx K_1 s^{-2} + K_2 s^{-\frac{\alpha}{\gamma}} + K_3 s^{\frac{\gamma-\alpha}{\beta}-1}, \end{aligned}$$

as  $s \rightarrow \infty$ , for constants  $K_1, K_2$ , and  $K_3$ , but, all the exponents lead to the  $\eta > 1$  case, thus:

$$\chi = 0.$$

The last case to consider is when  $\gamma < \alpha < \beta$ . As in the previous cases, we present the  $G$  functions

$$G_X(s) \approx -\frac{1}{\alpha} \log\left(1 - \frac{\alpha}{\beta}\right) + \frac{1}{\alpha} \log(s), \text{ and } G_Y(s) \approx -\frac{1}{\gamma} \log\left(1 - \frac{\gamma}{\alpha}\right) + \frac{1}{\gamma} \log(s)$$

as  $s \rightarrow \infty$ , then, the survival function is:

$$\begin{aligned} P(S > s, T > s) &\approx \frac{\alpha}{\alpha - \beta - \gamma} \left(1 - \frac{\alpha}{\beta}\right)^{\beta/\alpha} s^{-\beta/\alpha} \left(1 - \frac{\gamma}{\alpha}\right) s^{-1} - \frac{\gamma}{\alpha - \gamma} \left(1 - \frac{\gamma}{\alpha}\right)^{\alpha/\gamma} s^{-\alpha/\gamma} \\ &\quad - \frac{\beta\alpha}{(\alpha - \gamma)(\alpha - \beta - \gamma)} \left(1 - \frac{\gamma}{\alpha}\right) s^{-1} \left(1 - \frac{\alpha}{\beta}\right)^{1 - \gamma/\alpha} s^{-1 + \gamma/\alpha}, \\ &\approx K_1 s^{-1 - \beta/\alpha} + K_2 s^{-\alpha/\gamma} + K_3 s^{-2 + \gamma/\alpha}, \end{aligned}$$

as  $s \rightarrow \infty$ , for constants  $K_1, K_2$ , and  $K_3$ , where again, all the exponents lead to the  $\eta > 1$  case, thus:

$$\chi = 0.$$

Therefore, (4.12) holds.

### B.3 Proof of Theorem 4.2.1

The asymptotic form of  $P(S > s, T > s)$  is the target as in the previous proofs.

The first step is to find expressions for the marginal inverse functions. Given the marginals of  $X$  and  $Y$  in (4.17), we get the following expressions:

$$F_X^{\leftarrow}(x) \approx \left[ \alpha_1^{-1} \left( (1 - x)^{-\frac{1}{\theta}} - 1 \right) \right]^{\psi_1}, \text{ as } x \rightarrow \infty,$$

and

$$F_Y^{\leftarrow}(y) \approx \left[ \alpha_2^{-1} \left( (1 - y)^{-\frac{1}{\theta}} - 1 \right) \right]^{\psi_2}, \text{ as } y \rightarrow \infty,$$

then

$$G_X(s) = F_X^{\leftarrow}(e^{-1/s}) = \alpha_1^{-\psi_1} \left( (1 - e^{-1/s})^{-\frac{1}{\theta}} - 1 \right)^{\psi_1} \approx \alpha_1^{-\psi_1} s^{\frac{\psi_1}{\theta}}, \text{ as } s \rightarrow \infty, \tag{B.6}$$

and

$$G_Y(s) \approx \alpha_2^{-\psi_2} s^{\frac{\psi_2}{\theta}}, \text{ as } s \rightarrow \infty. \tag{B.7}$$

Hence, the joint survival distribution is expressed as follows:

$$\begin{aligned} P(S > s, T > s) &= P(B > G_X(s)^{\frac{1}{\psi_1}} A, C > G_Y(s)^{\frac{1}{\psi_2}} A) \\ &= \frac{1}{\Gamma(\theta)} \int_0^\infty e^{-a \left( \alpha_1 G_X(s)^{\frac{1}{\psi_1}} + \alpha_2 G_Y(s)^{\frac{1}{\psi_2}} \right)} e^{-a} a^{\theta-1} da \\ &= \left( 1 + \alpha_1 G_X(s)^{\frac{1}{\psi_1}} + \alpha_2 G_Y(s)^{\frac{1}{\psi_2}} \right)^{-\theta}, \end{aligned}$$

and using (B.6) and (B.7), the following holds:

$$P(S > s, T > s) \approx 2^{-\theta} s^{-1}, \text{ as } s \rightarrow \infty.$$

Hence,

$$\chi = \lim_{s \rightarrow \infty} P(S > s | T > s) = 2^{-\theta}.$$

Therefore, (4.19) holds.

## B.4 Conditions of Theorem 4.2.2

We show the conditions for which (4.23) holds. We need to introduce the following theorem, where the conditions on the density guarantee that a MLE is consistent. See Newey and McFadden (1994).

**Theorem B.4.1. MLE Consistency Conditions**

Suppose  $\mathbf{z} = (z_1, z_2, \dots)$  are i.i.d. with density  $f(z_i|\theta_0)$  and the following hold:

1. *Identification:* if  $\theta \neq \theta_0$ , then  $f(z_i|\theta) \neq f(z_i|\theta_0)$ ,  $i = 1, 2, \dots$
2. *Compactness:*  $\theta_0 \in \Theta$ , which is compact or the log-likelihood function  $l(\theta|\mathbf{z})$  is concave.
3. *Continuity:*  $\log f(z_i|\theta)$  is continuous at each  $\theta \in \Theta$ .
4.  $E(\sup_{\theta \in \Theta} |\log f(z_i|\theta)|) < \infty$ .

Then the MLE  $\hat{\theta}$  is consistent at  $\theta_0$ , i.e.

$$\hat{\theta} \xrightarrow{P} \theta_0.$$

Regarding our case:

Condition 1 holds if  $l(\theta|\mathbf{D})$  has a unique global maximum. But we showed in (4.23) that this is the case.

Condition 2 holds since  $\frac{d^2}{d\theta^2} l(\theta|\mathbf{D}) < 0$  for all  $\theta \in \Theta$ .

Condition 3 clearly holds.

For Condition 4, it suffices to show that

$$\sup_{\theta \in \Theta} \log f(X, Y|\theta) < \infty$$

with:

$$\begin{aligned} \log f(X, Y|\theta) &= \log \left( \frac{\alpha_1 \alpha_2}{\psi_1 \psi_2} X^{\frac{1}{\psi_1}-1} Y^{\frac{1}{\psi_2}-1} \right) + \log(\theta) + \log(\theta + 1) \\ &\quad - (\theta + 2) \log \left( 1 + \alpha_1 X^{\frac{1}{\psi_1}-1} + \alpha_2 Y^{\frac{1}{\psi_2}-1} \right) \end{aligned}$$

which we can write down as

$$\log f(X, Y | \theta) = k + \log(\theta) + \log(\theta + 1) - (\theta + 2)m,$$

with

$$k = \log \left( \frac{\alpha_1 \alpha_2}{\psi_1 \psi_2} X^{\frac{1}{\psi_1} - 1} Y^{\frac{1}{\psi_2} - 1} \right) \quad \text{and} \quad m = \log \left( 1 + \alpha_1 X^{\frac{1}{\psi_1} - 1} + \alpha_2 Y^{\frac{1}{\psi_2} - 1} \right).$$

Then

$$\log f(X, Y | \theta) < k + 2 \log(\theta + 2) - (\theta + 2)m.$$

However,  $2 \log(\theta + 2) - (\theta + 2)m$  has a finite maximum at  $\theta^* = \frac{2}{m} - 2$ , so

$$\log f(X, Y | \theta) < k + 2 \log \left( \frac{2}{m} \right) - 2,$$

which means Condition 4 holds.

Therefore, the four conditions of Theorem B.4 are satisfied. Hence, (4.25) holds.

## B.5 Proof of Theorem 4.2.3

We need to show the asymptotic normality of  $\hat{\theta}$ , as in (4.23). The following theorem sets the conditions for asymptotic normality to hold. See Schervish (1995).

### Theorem B.5.1. *Asymptotic Normality of MLEs*

*Suppose  $z_1, z_2, \dots$  are i.i.d. with density  $f(z_i | \theta)$ , where  $\theta \in \Theta$ , the parameter space. Let  $\hat{\theta}$  be an MLE. Assume  $\hat{\theta}$  is consistent at  $\theta$  and the following conditions hold:*

1.  $f(z_i|\theta)$  has continuous second derivative with respect to  $\theta$ .
2. Differentiation can be passed under the integral sign.
3. There exists  $H_r(z, \theta)$  such that for each  $\theta_0 \in \Theta$ :

$$\sup_{|\theta - \theta_0| \leq r} \left| \frac{d^2}{d\theta^2} \log f(z|\theta_0) - \frac{d^2}{d\theta^2} \log f(z|\theta) \right| \leq H_r(z, \theta_0),$$

with  $\lim_{r \rightarrow 0} E(H_r(z, \theta_0)) = 0$ .

4. The Fisher information matrix  $I(\theta)$  is finite and non-singular.

Then,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0)).$$

Applying this theorem to our model, we know that  $\hat{\theta}$  is consistent and:

Condition 1 holds, as the joint density can be expressed as follows; see (4.21):

$$f(x, y|\theta) \propto \theta(\theta + 1) K^{-\theta-2},$$

where  $K = 1 + \alpha_1 x^{\frac{1}{\psi_1}} + \alpha_2 y^{\frac{1}{\psi_2}}$  and  $K^{-\theta-2}$  is continuously differentiable with respect to  $\theta$ .

Condition 2 holds, as the support of  $X$  and  $Y$  does not depend on  $\theta$ .

For Condition 3, it suffices to show that  $\frac{d^2}{d\theta^2} \log f(x, y|\theta)$  is differentiable and has finite mean. However, by (4.24):

$$\frac{d^2}{d\theta^2} \log f(x, y|\theta) = -\frac{1}{\theta^2} - \frac{1}{(\theta + 1)^2},$$

then, it is differentiable and it has a finite mean. Thus, Condition 3 holds.

Condition 4 holds, since:

$$I(\theta) = -E \left( \frac{d^2}{d\theta^2} \log f(x, y | \theta) \right) = \frac{1}{\theta^2} + \frac{1}{(\theta + 1)^2} > 0,$$

Hence,  $\hat{\theta}$  is asymptotically normal, i.e. (4.26) holds.

## B.6 Proof of Theorem 5.3.1

Part 1. Consider the expansion in (5.5), which sets that, for large  $\theta$ , we have  $z^{1/\theta} - 1 \approx \frac{\log(z)}{\theta}$ . Then, we can write down the limit of the joint survival as follows:

$$\begin{aligned} P(X > x, Y > y) &= \left( 1 + \alpha_1(x^{\frac{1}{\theta\xi_1}} - 1) + \alpha_2(y^{\frac{1}{\theta\xi_2}} - 1) \right)^{-\theta} \\ &\approx \left( 1 + \frac{\alpha_1 \log(x^{\frac{1}{\xi_1}}) + \alpha_2 \log(y^{\frac{1}{\xi_2}})}{\theta} \right)^{-\theta} \\ &\approx \exp(-\alpha_1 \log(x^{1/\xi_1})) \exp(-\alpha_2 \log(y^{1/\xi_2})) \\ &= x^{-\alpha_1/\xi_1} y^{-\alpha_2/\xi_2}, \\ &\text{as } \theta \rightarrow \infty. \end{aligned}$$

which means that the joint density can be expressed as:

$$\lim_{\theta \rightarrow \infty} f_{X,Y}(x, y) = \frac{\alpha_1}{\xi_1} x^{-\frac{\alpha_1}{\xi_1} - 1} \frac{\alpha_2}{\xi_2} y^{-\frac{\alpha_2}{\xi_2} - 1}.$$

Consequently,

$$\chi = \lim_{u \rightarrow \infty} P(Y > u | X > u) = P(Y > u),$$

proving the independence in the tail.

Part 2. Define  $G_X(s)$  as in (B.8), i.e.:

$$G_X(s) \approx \left( \alpha_1^{-1} \left( s^{\frac{1}{\theta}} - 1 \right) + 1 \right)^{\theta \xi_1}, \text{ as } s \rightarrow \infty,$$

and similarly for  $G_Y(s)$ . Then, the joint survival function has the following expression:

$$\begin{aligned} P(X > x, Y > y) &= P(S > s, T > t) \\ &= \left( 1 + \alpha_1(G_X(s))^{1/\theta \xi_1} - 1 + \alpha_2(G_Y(s))^{1/\theta \xi_2} - 1 \right)^{-\theta} \\ &\approx \left( 1 + (s^{\frac{1}{\theta}} - 1) + (t^{\frac{1}{\theta}} - 1) \right)^{-\theta}, \end{aligned}$$

for large  $x$  and  $y$ .

Hence,

$$P(S > s, T > t) \approx s^{-1},$$

for large  $s$ . Therefore, as  $P(S > s) \approx s^{-1}$ , for large  $s$ , then the complete tail dependence is held.

## B.7 Proof of Theorem 5.3.2

The asymptotic form of  $P(S > s, T > s)$  is the target as in the previous proofs.

Given the marginal model in (5.2), we get the following expressions:

$$F_X^{\leftarrow}(x) = \left( \alpha_1^{-1} \left( (1-x)^{-\frac{1}{\theta}} - 1 \right) + 1 \right)^{\theta \xi_1},$$

and

$$F_Y^{\leftarrow}(y) = \left( \alpha_2^{-1} \left( (1-y)^{-\frac{1}{\theta}} - 1 \right) + 1 \right)^{\theta \xi_2}.$$

Then, the  $G$  functions are calculated as:

$$G_X(s) = F_X^+(e^{-1/s}) \approx \left( \alpha_1^{-1} \left( s^{\frac{1}{\theta}} - 1 \right) + 1 \right)^{\theta \xi_1}, \text{ as } s \rightarrow \infty, \quad (\text{B.8})$$

and

$$G_Y(s) \approx \left( \alpha_2^{-1} \left( s^{\frac{1}{\theta}} - 1 \right) + 1 \right)^{\theta \xi_2}, \text{ as } s \rightarrow \infty.$$

Hence, the joint survival distribution is expressed as follows:

$$\begin{aligned} P(S > s, T > s) &= \Pr(B > (G_X(s)^{1/\theta \xi_1} - 1)A, C > (G_Y(s)^{1/\theta \xi_2} - 1)A) \\ &= \frac{1}{\Gamma(\theta)} \int_0^\infty \exp(-a(1 + \alpha_1 G_X(s)^{1/\theta \xi_1} - \alpha_1 + \alpha_2 G_Y(s)^{1/\theta \xi_2} - \alpha_2)) a^{\theta-1} da \\ &= (1 + \alpha_1 G_X(s)^{1/\theta \xi_1} - \alpha_1 + \alpha_2 G_Y(s)^{1/\theta \xi_2} - \alpha_2)^{-\theta} \\ &\approx \left( \alpha_1 \left( \alpha_1^{-1} \left( s^{\frac{1}{\theta}} - 1 \right) + 1 \right) - \alpha_1 + \alpha_2 \left( \alpha_2^{-1} \left( s^{\frac{1}{\theta}} - 1 \right) + 1 \right) - \alpha_2 \right)^{-\theta} \\ &\approx \left( 2s^{\frac{1}{\theta}} - 2 \right)^{-\theta} \\ &\approx 2^{-\theta} s^{-1} \\ &\text{as } s \rightarrow \infty. \end{aligned}$$

Hence,

$$\chi = \lim_{s \rightarrow \infty} P(S > s | T > s) = 2^{-\theta}.$$

Therefore, (5.10) holds.

## B.8 Proof of Theorem 6.2.1

We know, from expression (6.4), that the marginals have the following form:

$$\bar{F}_{X_1}^*(x) = \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta \xi_1}} - 1 \right) \right)^{-\theta},$$

and similarly for  $\bar{F}_{X_2}^*$ . In this way, the  $G$  functions are expressed as:

$$\begin{aligned} G_{X_1}^*(s) &= (F_{X_1}^*)^\leftarrow(e^{-1/s}) \\ &= \left(1 + \alpha_1^{-1} \left((1 - e^{-1/s})^{-1/\theta} - 1\right)\right)^{\theta\xi_1} \\ &\approx \left(1 + \alpha_1^{-1} (s^{1/\theta} - 1)\right)^{\theta\xi_1} \end{aligned}$$

as  $s \rightarrow \infty$ , and

$$G_{X_2}^*(s) \approx \left(1 + \alpha_2^{-1} (s^{1/\theta} - 1)\right)^{\theta\xi_2}$$

Therefore, from the joint survival distribution in (6.9), the survival of  $S$  and  $T$  transformations (see (2.30)) is given by:

$$\begin{aligned} P(S^* > s, T^* > s) &= \bar{F}_{X_1, X_2}^* \left( G_{X_1}^*(s), G_{X_2}^*(s) \right) \\ &\approx w_1 \left(1 + (s^{1/\theta} - 1)\right)^{-\theta} + w_2 s^{-2} \\ &\approx w_1 2^{-\theta} s^{-1} + w_2 s^{-2}, \end{aligned}$$

as  $s \rightarrow \infty$ .

We can see the first term is dominant for  $w_1 \neq 0$  for large  $s$ . Therefore, the tail dependence parameter  $\chi$  is given by:

$$\chi = \lim_{s \rightarrow \infty} P(S^* > s | T^* > s) = w_1 2^{-\theta}.$$

Hence, (6.10) holds.

## B.9 Proof of Theorem 6.2.2

We know, from the proof in Appendix B.8, the  $G$  functions are expressed as:

$$G_{X_j}^*(s) \approx \left(1 + \alpha_j^{-1} (s^{1/\theta} - 1)\right)^{\theta\xi_j}, \quad (\text{B.9})$$

as  $s \rightarrow \infty$ , for  $j = 1, 2, 3$ .

Therefore, from (6.19), the joint survival of  $\mathbf{S}^* = (S_1^*, S_2^*, S_3^*)$  (see (6.17)) is given by:

$$\begin{aligned} \bar{F}_{\mathbf{S}^*}(s, s, s) &= \bar{F}_{\mathbf{X}}^* \left( G_{X_1}^*(s), G_{X_2}^*(s), G_{X_3}^*(s) \right) \\ &\approx w_1 s^{-3} + (w_2 + w_3 + w_4) \left( 1 + 2s^{\frac{1}{\theta}} \right)^{-\theta} s^{-1} + w_5 3^{-\theta} s^{-1} \\ &\approx w_1 s^{-3} + (w_2 + w_3 + w_4) 2^{-\theta} s^{-2} + w_5 3^{-\theta} s^{-1} \\ &\approx \begin{cases} w_5 3^{-\theta} s^{-1} & \text{for } w_5 \neq 0 \\ (w_2 + w_3 + w_4) 2^{-\theta} s^{-2} & \text{for } w_5 = 0, \text{ and } \max(w_2, w_3, w_4) \neq 0 \\ w_1 s^{-3} & \text{for } w_1 = 1, \end{cases} \end{aligned} \tag{B.10}$$

as  $s \rightarrow \infty$ .

Hence,

$$\chi_3^* = \lim_{s \rightarrow \infty} P(S_2^* > s, S_3^* > s | S_1^* > s) = w_5 3^{-\theta}.$$

Thus, (6.21) holds.

For the second part of the proof, we will prove (6.22) for  $\chi_{\{1,2\}}^*$ , since for the rest of the cases, the result follows immediatly.

By (6.20), we know the bivariate marginal is given by:

$$\begin{aligned} \bar{F}_{X_1, X_2}^*(x_1, x_2) &= w_* \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta \xi_1}} - 1 \right) \right)^{-\theta} \left( 1 + \alpha_2 \left( x_2^{\frac{1}{\theta \xi_2}} - 1 \right) \right)^{-\theta} \\ &\quad + (1 - w_*) \left( 1 + \alpha_1 \left( x_1^{\frac{1}{\theta \xi_1}} - 1 \right) + \alpha_2 \left( x_2^{\frac{1}{\theta \xi_2}} - 1 \right) \right)^{-\theta}, \end{aligned}$$

where  $w_* = w_1 + w_3 + w_4$ . Therefore, from (B.9):

$$P(S_1^* > s, S_2^* > s) \approx w_* s^{-2} + (1 - w_*) 2^{-\theta} s^{-1}, \quad (\text{B.11})$$

as  $s \rightarrow \infty$ . Hence

$$\chi_{\{1,2\}}^* = \lim_{s \rightarrow \infty} P(S_2^* > s | S_1^* > s) = (w_2 + w_5) 2^{-\theta}.$$

For the last part of the proof, we use (B.10) and (B.11) to get the following:

$$\begin{aligned} \chi_{\{[3],[1,2]\}}^* &= \lim_{s \rightarrow \infty} P(S_3^* > s | S_1^* > s, S_2^* > s) \\ &= \lim_{s \rightarrow \infty} \frac{P(S_1^* > s, S_2^* > s, S_3^* > s)}{P(S_1^* > s, S_2^* > s)} \\ &= \left( \frac{w_5}{w_2 + w_5} \right) \left( \frac{2}{3} \right)^\theta. \end{aligned}$$

Therefore, it is straightforward to see that (6.22) and (6.23) hold for all the cases.

## B.10 Proof of Theorem 6.3.1

We know, similarly to Proofs B.8 and B.9, the  $G$  functions for the univariate marginal of model (6.31) which holds (6.33) and (6.34), are given by:

$$G_{X_j}^*(s) \approx \left( 1 + \alpha_j^{-1} (s^{1/\theta} - 1) \right)^{\theta \xi_j} \quad (\text{B.12})$$

as  $s \rightarrow \infty$ , for  $j = 1, \dots, d$ . As well, we know this representation leads to the following asymptotic relation:

$$\left( 1 + \alpha_j \left( \left( G_{X_j}^*(s) \right)^{\frac{1}{\theta \xi_j}} - 1 \right) \right)^{-\theta} \approx s^{-1},$$

as  $s \rightarrow \infty$ , for  $j = 1, \dots, d$ . This means that for the  $d$ -dimensional case of the averaging, the dominant component of the probability:

$$P(S_1 > s, \dots, S_d > s)$$

is the model without exact independence variables for large  $s$ . In this way, if we set  $w_d$  as the weight of this model, then:

$$\begin{aligned} P(S_1 > s, \dots, S_d > s) &\approx w_d \left( 1 + \alpha_1 \left( G_{X_1}^*(s)^{\frac{1}{\theta \epsilon_1}} - 1 \right) + \dots \right. \\ &\quad \left. + \alpha_d \left( G_{X_d}^*(s)^{\frac{1}{\theta \epsilon_d}} - 1 \right) \right)^{-\theta} \\ &\approx w_d d^{-\theta} s^{-1}, \end{aligned}$$

as  $s \rightarrow \infty$ . Therefore:

$$\chi_d = \lim_{s \rightarrow \infty} P(S_2 > s, \dots, S_d > s | S_1 > s) = w_d d^{-\theta}.$$

For the second part of the proof, we know there exists dimensional coherence for model (6.31) when holds (6.33) and (6.34).

We divide the models in the averaging model in two: those where  $X_i$  and  $X_j$  are tail dependent and those where it is not the case.

Let  $\beta_1$  be the model in the averaging where  $X_i$  and  $X_j$  are tail dependent, such that it can be written as:

$$w_{\beta_1} \left( 1 + \alpha_i \left( X_i^{\frac{1}{\theta \epsilon_i}} - 1 \right) + \alpha_j \left( X_j^{\frac{1}{\theta \epsilon_j}} - 1 \right) + \dots \right)^{-\theta} \times \Pi_{\beta_1},$$

where  $\Pi_{\beta_1}$  is the product of the exact independent variables of model  $\beta_1$ .

Then, by dimensional coherence, if we marginalize with respect to the rest

of the variables in model  $\beta_1$ , the marginal expression of this model would be given by:

$$w_{\beta_1} \left( 1 + \alpha_i \left( X_i^{\frac{1}{\theta \xi_i}} - 1 \right) + \alpha_j \left( X_j^{\frac{1}{\theta \xi_j}} - 1 \right) \right)^{-\theta}.$$

Hence, by (B.12) the contribution of model  $\beta_1$  to  $P(S_1 > s, S_2 > s)$ , as  $s \rightarrow \infty$ , would be:

$$w_{\beta_1} 2^{-\theta} s^{-1}.$$

For the case where  $X_i$  and  $X_j$  are not tail dependent, let it be  $\beta_2$ , it is clear that by dimensional coherence, the bivariate marginal component of this model is expressed as:

$$w_{\beta_2} \left( 1 + \alpha_i \left( X_i^{\frac{1}{\theta \xi_i}} - 1 \right) \right)^{-\theta} \left( 1 + \alpha_j \left( X_j^{\frac{1}{\theta \xi_j}} - 1 \right) \right)^{-\theta}.$$

Therefore, the contribution of model  $\beta_2$  to  $P(S_1 > s, S_2 > s)$  would be:

$$w_{\beta_2} s^{-2},$$

as  $s \rightarrow \infty$ .

Hence, the dominant components, as  $s \rightarrow \infty$ , in  $P(S_1 > s, S_2 > s)$  are those where  $X_i$  and  $X_j$  are tail dependent. This means, the following holds:

$$P(S_1 > s, S_2 > s) \approx \sum_{k=1}^{T_{i,j}} w_{\beta_k} 2^{-\theta} s^{-1},$$

as  $s \rightarrow \infty$ , where  $T_{i,j}$  is the number of models where  $X_i$  and  $X_j$  are tail dependent.

Finally, we need to count how many models contain  $X_i$  and  $X_j$  as tail dependent. The easiest way to do it is by counting per type of model as follows:

- For  $d$ -tuple tail dependent model: 1 case.
- For  $(d-1)$ -tuple tail dependent models, only two models have either  $X_i$  or  $X_j$  as exact tail independent. Then, the number in this case is  $d-1$ .
- For  $(d-2)$ -tuple tail dependent models we can count the different ways we can choose the 2 exact independent models. Therefore, the number is:  $\binom{d-2}{2}$ .
- The last result holds for  $d-3, \dots, 2$ .

Therefore, the total of components where  $X_i$  and  $X_j$  are tail dependent is:

$$T_{i,j} = 1 + (d-2) + \binom{d-2}{2} + \binom{d-3}{2} + \dots + \binom{3}{2} + \binom{2}{2},$$

for  $d > 3$ .

Hence, (6.35) holds.

# Appendix C

## Tables

### C.1 Tables

LOW TAIL DEPENDENCE. LIGHT TAILS							LOW TAIL DEPENDENCE. HEAVY TAILS							
$\theta$	$\chi$	$\psi_1$	$\psi_2$	$\alpha_1$	$\alpha_2$		$\theta$	$\chi$	$\psi_1$	$\psi_2$	$\alpha_1$	$\alpha_2$		
5	0.031	1.25	1	1	1		5	0.031	5	7.5	1	1		
	$\theta$	$2^{-\theta}$	$\theta_{2.5\%}$	$\theta_{97.5\%}$	$\chi_{2.5\%}$	$\chi_{97.5\%}$		$\theta$	$2^{-\theta}$	$\theta_{2.5\%}$	$\theta_{97.5\%}$	$\chi_{2.5\%}$	$\chi_{97.5\%}$	
N=500	4.99	0.031	4.66	5.34	0.025	0.039	N=500	5.05	0.030	4.72	5.33	0.025	0.038	
	(0.152)	0.003	0.143	(0.162)	(0.003)	(0.004)		(0.162)	(0.004)	(0.152)	(0.234)	(0.004)	(0.004)	
SEMI LOW TAIL DEPENDENCE. LIGHT TAILS							SEMI LOW TAIL DEPENDENCE. HEAVY TAILS							
$\theta$	$\chi$	$\psi_1$	$\psi_2$	$\alpha_1$	$\alpha_2$		$\theta$	$\chi$	$\psi_1$	$\psi_2$	$\alpha_1$	$\alpha_2$		
3	0.125	0.75	0.6	1	1		3	0.125	3	4.5	1	1		
	$\theta$	$2^{-\theta}$	$\theta_{2.5\%}$	$\theta_{97.5\%}$	$\chi_{2.5\%}$	$\chi_{97.5\%}$		$\theta$	$2^{-\theta}$	$\theta_{2.5\%}$	$\theta_{97.5\%}$	$\chi_{2.5\%}$	$\chi_{97.5\%}$	
N=500	2.93	0.13	2.73	3.15	0.113	0.150	N=500	3.02	0.12	2.84	3.24	0.106	0.140	
	(0.085)	(0.008)	(0.079)	(0.007)	(0.007)	(0.008)		(0.088)	(0.008)	(0.114)	(0.094)	(0.007)	(0.011)	
MEDIUM TAIL DEPENDENCE. LIGHT TAILS							MEDIUM TAIL DEPENDENCE. HEAVY TAILS							
$\theta$	$\chi$	$\psi_1$	$\psi_2$	$\alpha_1$	$\alpha_2$		$\theta$	$\chi$	$\psi_1$	$\psi_2$	$\alpha_1$	$\alpha_2$		
1	0.5	0.25	0.2	1	1		1	0.5	1	1.5	1	1		
	$\theta$	$2^{-\theta}$	$\theta_{2.5\%}$	$\theta_{97.5\%}$	$\chi_{2.5\%}$	$\chi_{97.5\%}$		$\theta$	$2^{-\theta}$	$\theta_{2.5\%}$	$\theta_{97.5\%}$	$\chi_{2.5\%}$	$\chi_{97.5\%}$	
N=500	1.01	0.497	0.93	1.08	0.475	0.524	N=500	1.02	0.493	0.94	1.10	0.465	0.519	
	(0.043)	(0.015)	(0.040)	(0.069)	(0.023)	(0.015)		(0.043)	(0.014)	(0.040)	(0.046)	(0.015)	(0.014)	
SEMI HIGH TAIL DEPENDENCE. LIGHT TAILS							SEMI HIGH TAIL DEPENDENCE. HEAVY TAILS							
$\theta$	$\chi$	$\psi_1$	$\psi_2$	$\alpha_1$	$\alpha_2$		$\theta$	$\chi$	$\psi_1$	$\psi_2$	$\alpha_1$	$\alpha_2$		
0.333	0.794	0.083	0.133	1	1		0.333	0.794	0.333	0.5	1	1		
	$\theta$	$2^{-\theta}$	$\theta_{2.5\%}$	$\theta_{97.5\%}$	$\chi_{2.5\%}$	$\chi_{97.5\%}$		$\theta$	$2^{-\theta}$	$\theta_{2.5\%}$	$\theta_{97.5\%}$	$\chi_{2.5\%}$	$\chi_{97.5\%}$	
N=500	0.341	0.789	0.31	0.37	0.773	0.805	N=500	0.331	0.795	0.30	0.36	0.779	0.810	
	(0.021)	(0.012)	(0.020)	(0.023)	(0.012)	(0.011)		(0.012)	(0.007)	(0.011)	(0.013)	(0.007)	(0.006)	
HIGH TAIL DEPENDENCE. LIGHT TAILS							HIGH TAIL DEPENDENCE. HEAVY TAILS							
$\theta$	$\chi$	$\psi_1$	$\psi_2$	$\alpha_1$	$\alpha_2$		$\theta$	$\chi$	$\psi_1$	$\psi_2$	$\alpha_1$	$\alpha_2$		
0.2	0.871	0.050	0.080	1	1		0.2	0.871	0.2	0.3	1	1		
	$\theta$	$2^{-\theta}$	$\theta_{2.5\%}$	$\theta_{97.5\%}$	$\chi_{2.5\%}$	$\chi_{97.5\%}$		$\theta$	$2^{-\theta}$	$\theta_{2.5\%}$	$\theta_{97.5\%}$	$\chi_{2.5\%}$	$\chi_{97.5\%}$	
N=500	0.196	0.873	0.18	0.21	0.861	0.883	N=500	0.195	0.873	0.18	0.21	0.863	0.882	
	(0.009)	(0.006)	(0.008)	(0.010)	(0.006)	(0.005)		(0.009)	(0.005)	(0.010)	(0.010)	(0.006)	(0.006)	

Table C.1: MLE in Bivariate Model 1: estimating  $\theta$ ,  $n=500$  (standard error of replications is given in brackets).

LOW DEPENDENCE AND LIGHT TAILS										LOW DEPENDENCE AND HEAVY TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$	$\chi$					$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$	$\chi$					
5	1	1	0.1	0.15	$\chi$ 0.031					5	1	1	1	1.5	$\chi$ 0.031					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		
X	67.1	3.51	0.3	63 · 10 <sup>-21</sup>	0.88	0.543				X	164	2.05	1.03	48 · 10 <sup>-80</sup>	1.15	0.450				
Y	(121)	2.68	0.37	(0.128)	(0.58)	(0.250)				Y	(263)	3.34	4.13	(0.012)	(1.02)	(0.373)				
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		
X	15.4	3.57	0.34	24 · 10 <sup>-8</sup>	4	0.092				X	17.4	1.93	1.88	5.8 · 10 <sup>-6</sup>	3.41	0.094				
Y	(22.3)	3.82	0.34	(0.048)	(2.02)	(0.094)				Y	(30.0)	1.63	2.41	(0.038)	(1.30)	(0.034)				
SEMI LOW DEPENDENCE AND LIGHT TAILS										SEMI LOW DEPENDENCE AND HEAVY TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$	$\chi$					$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$	$\chi$					
3	1	1	0.1	0.15	$\chi$ 0.125					3	1	1	1	1.5	$\chi$ 0.125					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		
X	49.0	2.49	0.25	13 · 10 <sup>-18</sup>	1.10	0.468				X	3.38	1.47	1.45	0.096	1.37	0.388				
Y	(102)	1.19	0.18	(0.135)	(0.60)	(0.264)				Y	(2.17)	1.11	1.57	(0.133)	(0.66)	(0.130)				
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		
X	5.25	1.60	0.17	0.026	2.67	0.157				X	3.28	1.04	1.01	0.103	2.31	0.201				
Y	(6.40)	1.46	0.21	(0.042)	(1.31)	(0.067)				Y	(0.38)	0.90	1.43	(0.029)	(0.22)	(0.031)				
MEDIUM DEPENDENCE AND LIGHT TAILS										MEDIUM DEPENDENCE AND HEAVY TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$	$\chi$					$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$	$\chi$					
1	1	1	0.1	0.15	$\chi$ 0.5					1	1	1	1	1.5	$\chi$ 0.5					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		
X	0.91	1.18	0.16	0.534	0.54	1.53	0.346	0.680		X	1.08	0.80	0.80	0.472	0.62	1.90	0.267	0.651		
Y	(0.25)	0.96	0.15	(0.085)	(0.12)	(0.52)	(0.106)	(0.053)		Y	(0.18)	0.93	1.4	(0.059)	(0.08)	(0.12)	(0.072)	(0.037)		
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		
X	0.97	1.03	0.10	0.510	0.70	1.19	0.438	0.376		X	0.90	1.00	0.98	0.503	0.81	1.21	0.431	0.571		
Y	(0.12)	1.06	0.15	(0.043)	(0.60)	(0.16)	(0.050)	(0.037)		Y	(0.00)	1.02	1.48	(0.032)	(0.07)	(0.12)	(0.036)	(0.026)		
SEMI HIGH DEPENDENCE AND LIGHT TAILS										SEMI HIGH DEPENDENCE AND HEAVY TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$	$\chi$					$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$	$\chi$					
1/3	1	1	0.1	0.15	$\chi$ 0.794					1/3	1	1	1	1.5	$\chi$ 0.794					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		
X	0.31	1.39	0.11	0.807	0.21	0.46	0.730	0.862		X	0.33	1.16	1.01	0.793	0.23	0.40	0.713	0.854		
Y	(0.06)	1.29	0.16	(0.036)	(0.04)	(0.10)	(0.051)	(0.024)		Y	(0.07)	1.07	1.51	(0.034)	(0.04)	(0.11)	(0.058)	(0.028)		
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		
X	0.34	1.02	0.10	0.790	0.20	0.39	0.762	0.816		X	0.32	0.90	1.02	0.801	0.20	0.38	0.781	0.818		
Y	(0.03)	1.00	0.15	(0.018)	(0.03)	(0.04)	(0.022)	(0.016)		Y	(0.03)	0.96	1.52	(0.018)	(0.06)	(0.03)	(0.015)	(0.033)		
HIGH DEPENDENCE AND LIGHT TAILS										HIGH DEPENDENCE AND HEAVY TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$	$\chi$					$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$	$\chi$					
0.2	1	1	0.1	0.15	$\chi$ 0.871					0.2	1	1	1	1.5	$\chi$ 0.871					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		
X	0.20	1.05	0.10	0.871	0.14	0.28	0.823	0.907		X	0.20	1.11	1.00	0.868	0.14	0.26	0.817	0.905		
Y	(0.04)	1.35	0.15	(0.022)	(0.02)	(0.05)	(0.031)	(0.015)		Y	(0.05)	1.17	1.48	(0.028)	(0.03)	(0.07)	(0.040)	(0.019)		
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{1\%}$	$\chi_{5\%}$		
X	0.21	0.90	0.10	0.866	0.18	0.24	0.848	0.881		X	0.20	1.04	1.00	0.869	0.18	0.23	0.852	0.884		
Y	(0.02)	0.90	0.15	(0.006)	(0.01)	(0.02)	(0.010)	(0.008)		Y	(0.01)	1.02	1.50	(0.005)	(0.01)	(0.01)	(0.006)	(0.004)		

Table C.2: MLE estimates in Bivariate Model 2. Standard errors are given in brackets. Average computing time: 10.9 and 2.9 seconds for  $N = 500$  and  $N = 70$ , respectively.

LOW DEPENDENCE AND LIGHT TAILS							LOW DEPENDENCE AND HEAVY TAILS						
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$			$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$		
5	1	1	0.1	0.15	$\chi$ 0.031		5	1	1	1	1.5	$\chi$ 0.031	
N=1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{5\%}$	$\chi_{95\%}$	N=1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{5\%}$	$\chi_{95\%}$
X	5.29	1.801	0.175	0.026	3.74	0.075	X	4.85	1.078	1.06	0.035	3.80	0.072
Y	(1.92)	1.251	0.186	(0.035)	(1.11)	(0.051)	Y	(0.92)	1.038	1.544	(0.035)	(1.38)	(0.050)

Table C.3: MLE in Bivariate Model 2 in low tail dependence with large sample size.

M <sub>1</sub> . HEAVY TAILS						M <sub>1</sub> . LIGHT TAILS					
$\alpha_1$	$\alpha_2$	$\alpha_3$	$\xi_1$	$\xi_2$	$\xi_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\xi_1$	$\xi_2$	$\xi_3$
0.5	0.5	0.5	2	3	2.5	0.5	0.5	0.5	0.1	0.2	0.15
N	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	N	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$
70	0.995	0.002	0.001	0.002	0.001	70	0.982	0.002	0.006	0.010	0
150	0.999	0	0	0	0.001	150	0.968	0.014	0.007	0.007	0.003

Table C.4: Trivariate independent simulations and Reversible Jump estimations.

LOW DEPENDENCE AND LIGHT TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$						
5	1	1	0.1	0.15	X 0.031					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$X_{95\%}$				
	X	3.21	1.42	0.12	0.108	1.70	0.307			
	Y	(0.84)	1.12	0.16	(0.062)	(0.44)	(0.107)			
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$X_{95\%}$				
	X	4.75	1.36	0.13	0.037	3.24	0.106			
	Y	(1.25)	1.21	0.18	(0.038)	(0.75)	(0.067)			
SEMI LOW DEPENDENCE AND LIGHT TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$						
3	1	1	0.1	0.15	X 0.125					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$X_{95\%}$				
	X	2.95	1.34	0.14	0.129	1.57	0.336			
	Y	(0.79)	1.18	0.17	(0.089)	(0.37)	(0.083)			
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$X_{95\%}$				
	X	3.50	1.00	0.10	0.088	2.50	0.178			
	Y	(0.97)	1.23	0.17	(0.040)	(0.37)	(0.053)			
MEDIUM DEPENDENCE AND LIGHT TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$						
1	1	1	0.1	0.15	X 0.5					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$X_{2.5\%}$	$X_{97.5\%}$		
	X	0.94	1.22	0.11	0.520	0.55	1.87	0.274	0.681	
	Y	(0.23)	1.05	0.15	(0.077)	(0.11)	(0.70)	(0.109)	(0.048)	
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$X_{2.5\%}$	$X_{97.5\%}$		
	X	0.98	1.04	0.10	0.508	0.79	1.23	0.425	0.575	
	Y	(0.13)	1.06	0.15	(0.044)	(0.00)	(0.18)	(0.052)	(0.038)	
SEMI HIGH DEPENDENCE AND LIGHT TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$						
1/3	1	1	0.1	0.15	X 0.794					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$X_{2.5\%}$	$X_{97.5\%}$		
	X	0.32	1.31	0.11	0.801	0.22	0.48	0.717	0.860	
	Y	(0.07)	1.23	0.16	(0.037)	(0.04)	(0.11)	(0.056)	(0.024)	
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$X_{2.5\%}$	$X_{97.5\%}$		
	X	0.34	1.02	0.10	0.790	0.20	0.40	0.759	0.818	
	Y	(0.03)	1.00	0.15	(0.019)	(0.03)	(0.04)	(0.022)	(0.018)	
HIGH DEPENDENCE AND LIGHT TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$						
0.2	1	1	0.1	0.15	X 0.871					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$X_{2.5\%}$	$X_{97.5\%}$		
	X	0.20	1.03	0.1	0.870	0.14	0.20	0.817	0.908	
	Y	(0.03)	1.27	0.15	(0.019)	(0.02)	(0.05)	(0.029)	(0.014)	
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$X_{2.5\%}$	$X_{97.5\%}$		
	X	0.21	0.91	0.10	0.866	0.18	0.24	0.848	0.882	
	Y	(0.01)	0.91	0.15	(0.009)	(0.01)	(0.02)	(0.010)	(0.008)	

LOW DEPENDENCE AND HEAVY TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$						
5	1	1	1	1.5	X 0.031					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$X_{95\%}$				
	X	4.29	0.83	0.86	0.051	2.29	0.204			
	Y	(0.56)	0.76	1.11	(0.022)	(0.30)	(0.043)			
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$X_{95\%}$				
	X	4.81	1.01	0.87	0.035	3.32	0.100			
	Y	(1.25)	0.98	1.33	(0.030)	(0.76)	(0.040)			
SEMI LOW DEPENDENCE AND HEAVY TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$						
3	1	1	1	1.5	X 0.125					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$X_{95\%}$				
	X	2.82	0.87	0.92	0.133	1.59	0.331			
	Y	(0.85)	0.86	1.33	(0.089)	(0.42)	(0.060)			
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$X_{95\%}$				
	X	3.38	0.99	0.97	0.098	2.45	0.183			
	Y	(0.37)	0.86	1.36	(0.026)	(0.22)	(0.029)			
MEDIUM DEPENDENCE AND HEAVY TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$						
1	1	1	1	1.5	X 0.5					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$X_{2.5\%}$	$X_{97.5\%}$		
	X	1.27	0.82	0.83	0.414	0.73	2.63	0.163	0.603	
	Y	(0.21)	0.8	1.27	(0.060)	(0.10)	(0.52)	(0.063)	(0.043)	
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$X_{2.5\%}$	$X_{97.5\%}$		
	X	1.02	0.98	0.97	0.463	0.83	1.29	0.490	0.564	
	Y	(0.00)	0.90	1.45	(0.033)	(0.07)	(0.13)	(0.030)	(0.028)	
SEMI HIGH DEPENDENCE AND HEAVY TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$						
1/3	1	1	1	1.5	X 0.794					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$X_{2.5\%}$	$X_{97.5\%}$		
	X	0.37	1.02	0.95	0.774	0.25	0.57	0.875	0.840	
	Y	(0.07)	0.94	1.4	(0.036)	(0.04)	(0.12)	(0.057)	(0.025)	
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$X_{2.5\%}$	$X_{97.5\%}$		
	X	0.32	0.97	1.01	0.797	0.28	0.38	0.768	0.823	
	Y	(0.03)	0.94	1.50	(0.019)	(0.03)	(0.04)	(0.022)	(0.018)	
HIGH DEPENDENCE AND HEAVY TAILS										
$\theta$	$\alpha_1$	$\alpha_2$	$\xi_1$	$\xi_2$						
0.2	1	1	1	1.5	X 0.871					
N=70	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$X_{2.5\%}$	$X_{97.5\%}$		
	X	0.22	1	0.94	0.857	0.16	0.32	0.800	0.897	
	Y	(0.05)	1.05	1.30	(0.030)	(0.03)	(0.08)	(0.048)	(0.020)	
N=500	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$X_{2.5\%}$	$X_{97.5\%}$		
	X	0.21	1.01	0.98	0.866	0.18	0.24	0.848	0.882	
	Y	(0.01)	0.90	1.47	(0.005)	(0.01)	(0.01)	(0.006)	(0.006)	

Table C.5: Bayesian estimations for Model 2 simulations. Standard errors are given in brackets. Average computing time: 76.8 and 21.5 seconds for  $N = 500$  and  $N = 70$ , respectively.

EXACT INDEPENDENCE. LIGHT GPD's										EXACT INDEPENDENCE. HEAVY GPD's									
Marg.		$k_1$	$\sigma_1$	$k_2$	$\sigma_2$					Marg.		$k_1$	$\sigma_1$	$k_2$	$\sigma_2$				
Params.		0.1	1	0.15	1					Params.		1	1	1.5	1				
N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.	N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.
X	4.91	0.94	0.25	0.033	3.71	0.076	0.046			X	8.25	1.41	1.30	0.003	5.30	0.024	0.048		
Y	(0.462)	1.07	0.28	(0.102)	(0.407)	(0.021)				Y	(0.723)	1.47	1.21	(0.002)	(0.488)	(0.007)			
LOW TAIL DEPENDENCE. LIGHT GPD's										LOW TAIL DEPENDENCE. HEAVY GPD's									
Clayton		$\phi$	$\chi$							Clayton		$\phi$	$\chi$						
Params.		0.20	0.031							Params.		0.20	0.031						
N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.	N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.
X	2.95	0.87	0.26	0.130	2.37	0.194	0.176			X	5.22	0.91	0.93	0.027	3.43	0.093	0.168		
Y	(0.295)	1.14	0.20	(0.076)	(0.216)	(0.256)				Y	(0.832)	1.03	1.45	(0.015)	(0.438)	(0.028)			
Clayton		$\phi$	$\chi$							Clayton		$\phi$	$\chi$						
Params.		1.33	0.125							Params.		1.33	0.125						
N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.	N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.
X	2.22	0.90	0.27	0.214	1.73	3.01	0.124	0.301	0.248	X	3.16	0.88	0.94	0.112	2.00	1.78	0.010	0.235	0.226
Y	(0.110)	0.94	0.28	(0.017)	(0.099)	(0.217)	(0.017)	(0.021)		Y	(0.421)	0.98	1.35	(0.031)	(0.186)	(0.900)	(0.012)	(0.030)	
MEDIUM TAIL DEPENDENCE. LIGHT GPD's										MEDIUM TAIL DEPENDENCE. HEAVY GPD's									
Clayton		$\phi$	$\chi$							Clayton		$\phi$	$\chi$						
Params.		2	0.5							Params.		2	0.5						
N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.	N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.
X	1.10	0.64	0.26	0.487	0.85	1.41	0.377	0.557	0.502	X	1.11	0.61	0.92	0.465	0.83	1.61	0.328	0.583	0.502
Y	(0.180)	0.88	0.30	(0.059)	(0.151)	(0.220)	(0.050)	(0.050)		Y	(0.180)	0.88	1.42	(0.061)	(0.130)	(0.263)	(0.066)	(0.054)	
Clayton		$\phi$	$\chi$							Clayton		$\phi$	$\chi$						
Params.		4	0.794							Params.		4	0.794						
N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.	N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.
X	0.339	0.56	0.28	0.790	0.118	0.462	0.726	0.921	0.794	X	0.306	0.50	0.92	0.776	0.266	0.491	0.711	0.830	0.794
Y	(0.056)	0.97	0.30	(0.031)	(0.052)	(0.067)	(0.034)	(0.032)		Y	(0.050)	0.94	1.36	(0.027)	(0.037)	(0.071)	(0.035)	(0.022)	
Clayton		$\phi$	$\chi$							Clayton		$\phi$	$\chi$						
Params.		6	0.871							Params.		6	0.871						
N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.	N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.
X	0.171	0.50	0.34	0.888	0.031	0.265	0.832	0.970	0.870	X	0.216	0.52	0.96	0.800	0.127	0.266	0.614	0.916	0.870
Y	(0.044)	0.90	0.37	(0.027)	(0.005)	(0.062)	(0.036)	(0.003)		Y	(0.031)	1.07	1.40	(0.019)	(0.030)	(0.049)	(0.027)	(0.019)	
TAIL INDEPENDENCE. LIGHT GPD's										TAIL INDEPENDENCE. HEAVY GPD's									
Gaussian		$\rho$	$\chi$							Gaussian		$\rho$	$\chi$						
Params.		0.5	0							Params.		0.5	0						
N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.	N = 1000	$\theta$	$\alpha$	$\xi$	$2^{-\theta}$	$\theta_{1\%}$	$\theta_{5\%}$	$\chi_{2.5\%}$	$\chi_{7.5\%}$	$\chi$ Emp.
X	1.90	0.73	0.27	0.267	1.48	2.70	0.153	0.350	0.246	X	2.72	0.70	0.91	0.151	1.79	3.08	0.030	0.200	0.246
Y	(0.180)	0.73	0.28	(0.032)	(0.143)	(0.347)	(0.037)	(0.037)		Y	(0.370)	0.76	1.32	(0.039)	(0.167)	(1.086)	(0.025)	(0.037)	

Table C.6: Bayesian estimations for data simulations. Standard errors are given in brackets. Average computing time: 11.9 seconds per replication.

M <sub>5</sub> (TD). HEAVY TAILS							M <sub>5</sub> (TD). LIGHT TAILS						
N	α <sub>1</sub>	α <sub>2</sub>	α <sub>3</sub>	ξ <sub>1</sub>	ξ <sub>2</sub>	ξ <sub>3</sub>	N	α <sub>1</sub>	α <sub>2</sub>	α <sub>3</sub>	ξ <sub>1</sub>	ξ <sub>2</sub>	ξ <sub>3</sub>
70	0.5	0.5	0.5	2	3	2.5	70	0.5	0.5	0.5	0.1	0.2	0.15
θ	3 <sup>-θ</sup>	w <sub>1</sub>	w <sub>2</sub>	w <sub>3</sub>	w <sub>4</sub>	w <sub>5</sub>	θ	3 <sup>-θ</sup>	w <sub>1</sub>	w <sub>2</sub>	w <sub>3</sub>	w <sub>4</sub>	w <sub>5</sub>
0.1	0.90	0.00	0.00	0.00	0.00	1.00	0.1	0.90	0.00	0.00	0.00	0.00	1.00
0.5	0.58	0.00	0.00	0.00	0.00	1.00	0.5	0.58	0.00	0.00	0.00	0.00	1.00
1	0.33	0.00	0.00	0.01	0.00	0.99	1	0.33	0.02	0.00	0.00	0.00	0.97
2	0.11	0.00	0.00	0.00	0.00	0.99	2	0.11	0.01	0.00	0.00	0.00	0.99
3	0.04	0.00	0.00	0.05	0.00	0.95	3	0.04	0.03	0.13	0.26	0.01	0.57
3.5	0.02	0.00	0.00	0.03	0.00	0.97	3.5	0.02	0.05	0.22	0.01	0.04	0.68
4	0.01	0.23	0.08	0.08	0.28	0.33	4	0.01	0.01	0.01	0.01	0.04	0.93
4.5	0.01	0.00	0.33	0.04	0.00	0.63	4.5	0.01	0.15	0.01	0.38	0.25	0.21
5	0.00	0.11	0.07	0.12	0.01	0.69	5	0.00	0.01	0.04	0.00	0.00	0.29
M <sub>5</sub> (TD). HEAVY TAILS							M <sub>5</sub> (TD). LIGHT TAILS						
N	α <sub>1</sub>	α <sub>2</sub>	α <sub>3</sub>	ξ <sub>1</sub>	ξ <sub>2</sub>	ξ <sub>3</sub>	N	α <sub>1</sub>	α <sub>2</sub>	α <sub>3</sub>	ξ <sub>1</sub>	ξ <sub>2</sub>	ξ <sub>3</sub>
150	0.5	0.5	0.5	2	3	2.5	150	0.5	0.5	0.5	0.1	0.2	0.15
θ	3 <sup>-θ</sup>	w <sub>1</sub>	w <sub>2</sub>	w <sub>3</sub>	w <sub>4</sub>	w <sub>5</sub>	θ	3 <sup>-θ</sup>	w <sub>1</sub>	w <sub>2</sub>	w <sub>3</sub>	w <sub>4</sub>	w <sub>5</sub>
0.1	0.90	0.00	0.00	0.00	0.00	1.00	0.1	0.90	0.00	0.00	0.00	0.00	1.00
0.5	0.58	0.00	0.00	0.00	0.00	1.00	0.5	0.58	0.00	0.00	0.00	0.00	1.00
1	0.33	0.00	0.00	0.00	0.00	1.00	1	0.33	0.04	0.00	0.00	0.00	0.96
2	0.11	0.00	0.00	0.00	0.00	1.00	2	0.11	0.00	0.00	0.00	0.00	1.00
3	0.04	0.00	0.01	0.00	0.00	0.99	3	0.04	0.04	0.00	0.00	0.00	0.96
4	0.01	0.01	0.04	0.01	0.00	0.93	3.5	0.02	0.01	0.00	0.00	0.00	0.99
4.5	0.01	0.00	0.00	0.45	0.02	0.53	4	0.01	0.08	0.13	0.01	0.13	0.65
5	0.00	0.00	0.00	0.00	0.00	1.00	4.5	0.01	0.02	0.00	0.18	0.00	0.80
							5	0.00	0.16	0.00	0.29	0.07	0.48

Table C.7: Trivariate M<sub>5</sub> simulations and R. J. estimations: small and medium size.

$M_5$ (TD). HEAVY TAILS							$M_5$ (TD). LIGHT TAILS						
N	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\xi_1$	$\xi_2$	$\xi_3$	N	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\xi_1$	$\xi_2$	$\xi_3$
300	0.5	0.5	0.5	2	3	2.5	300	0.5	0.5	0.5	0.1	0.2	0.15
$\theta$	$3^{-\theta}$	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$\theta$	$3^{-\theta}$	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$
4.5	0.01	0.01	0.05	0.00	0.00	0.95	4.5	0.01	0.01	0.00	0.00	0.00	0.98
5	0.00	0.00	0.00	0.00	0.00	1.00	5	0.00	0.03	0.00	0.00	0.03	0.93
6	0.00	0.01	0.05	0.00	0.01	0.94	6	0.00	0.08	0.00	0.02	0.00	0.90
6.5	0.00	0.00	0.00	0.00	0.00	1.00	6.5	0.00	0.02	0.00	0.00	0.00	0.98
7	0.00	0.00	0.00	0.00	0.92	0.08	7	0.00	0.65	0.01	0.00	0.32	0.01
7.5	0.00	0.06	0.00	0.00	0.94	0.01	7.5	0.00	0.79	0.14	0.01	0.02	0.04
8	0.00	0.00	0.19	0.00	0.00	0.80	8	0.00	0.02	0.00	0.00	0.00	0.98
9	0.00	0.00	0.00	0.02	0.00	0.98	9	0.00	0.72	0.03	0.03	0.02	0.20

Table C.8: Trivariate  $M_5$  simulations and R. J. estimations: large size.

$M_2$ (DD <sub>1,2</sub> ). HEAVY TAILS							$M_2$ (DD <sub>1,2</sub> ). LIGHT TAILS						
N	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\xi_1$	$\xi_2$	$\xi_3$	N	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\xi_1$	$\xi_2$	$\xi_3$
300	0.5	0.5	0.5	2	3	2.5	300	0.5	0.5	0.5	0.1	0.2	0.15
$\theta$	$2^{-\theta}$	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$\theta$	$2^{-\theta}$	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$
4	0.06	0.00	1.00	0.00	0.00	0.00	4	0.06	0.14	0.86	0.00	0.00	0.00
4.5	0.04	0.00	1.00	0.00	0.00	0.00	4.5	0.04	0.06	0.94	0.00	0.00	0.00
5	0.03	0.18	0.80	0.01	0.00	0.01	5	0.03	0.25	0.75	0.00	0.00	0.00
6	0.02	0.80	0.20	0.00	0.00	0.00	6	0.02	0.65	0.33	0.00	0.00	0.01
7	0.01	0.99	0.01	0.00	0.00	0.00	7	0.01	1.00	0.00	0.00	0.00	0.00

Table C.9: Trivariate  $M_2$  simulations and R. J. estimations: large size.

$M_2$ (DD <sub>1,2</sub> ). HEAVY TAILS							$M_2$ (DD <sub>1,2</sub> ). LIGHT TAILS						
N	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\xi_1$	$\xi_2$	$\xi_3$	N	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\xi_1$	$\xi_2$	$\xi_3$
70	0.5	0.5	0.5	2	3	2.5	70	0.5	0.5	0.5	0.1	0.2	0.15
$\theta$	$2^{-\theta}$	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$\theta$	$2^{-\theta}$	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$
0.1	0.93	0.00	1.00	0.00	0.00	0.00	0.1	0.93	0.00	1.00	0.00	0.00	0.00
0.5	0.71	0.00	1.00	0.00	0.00	0.00	0.5	0.71	0.01	0.97	0.00	0.00	0.02
1	0.50	0.00	0.99	0.00	0.00	0.01	1	0.50	0.07	0.93	0.00	0.00	0.00
2	0.25	0.28	0.71	0.00	0.00	0.00	2	0.25	0.03	0.97	0.00	0.00	0.00
3	0.13	0.51	0.14	0.09	0.03	0.24	3	0.13	0.23	0.73	0.01	0.00	0.03
3.5	0.09	0.12	0.87	0.00	0.00	0.00	3.5	0.09	0.46	0.51	0.02	0.00	0.01
4	0.06	0.00	0.96	0.00	0.00	0.03	4	0.06	0.86	0.06	0.01	0.05	0.01
5	0.03	0.25	0.75	0.00	0.00	0.00	5	0.03	0.60	0.30	0.09	0.00	0.00
$M_2$ (DD <sub>1,2</sub> ). HEAVY TAILS							$M_2$ (DD <sub>1,2</sub> ). LIGHT TAILS						
N	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\xi_1$	$\xi_2$	$\xi_3$	N	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\xi_1$	$\xi_2$	$\xi_3$
150	0.5	0.5	0.5	2	3	2.5	150	0.5	0.5	0.5	0.1	0.2	0.15
$\theta$	$2^{-\theta}$	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$\theta$	$2^{-\theta}$	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$
2	0.25	0.00	1.00	0.00	0.00	0.00	2	0.25	0.01	0.99	0.00	0.00	0.00
3	0.13	0.00	1.00	0.00	0.00	0.00	3	0.13	0.01	0.98	0.00	0.00	0.01
3.5	0.09	0.01	0.99	0.00	0.00	0.00	3.5	0.09	0.06	0.94	0.00	0.00	0.00
4	0.06	0.92	0.08	0.00	0.00	0.00	4	0.06	0.73	0.27	0.00	0.00	0.00
5	0.03	0.20	0.79	0.00	0.00	0.00	5	0.03	0.23	0.53	0.18	0.00	0.05
6	0.02	0.96	0.01	0.00	0.03	0.00	6	0.02	0.94	0.04	0.00	0.01	0.00

Table C.10: Trivariate  $M_2$  simulations and R. J. estimations: small and medium size.

FRANCE-UK LEFT TAIL										-FRANCE-UK RIGHT TAIL											
N = 70	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town	N = 67	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town
FRAN	0.73	0.59	0.01	0.60	0.54	1.00	0.50	0.69	0.57	0.96	FRAN	0.81	1.01	0.01	0.57	0.59	1.15	0.45	0.67	0.51	1.16
UK	1.27	0.01									UK	0.72	0.01								
FRANCE-GERMANY LEFT TAIL										FRANCE-GERMANY RIGHT TAIL											
N = 75	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Logistic	$\bar{\chi}$ Ledford Town	N = 67	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Asym-Mix	$\bar{\chi}$ Ledford Town
FRAN	0.77	1.03	0.01	0.58	0.57	1.10	0.47	0.67	0.57	0.88	FRAN	0.79	0.95	0.01	0.58	0.58	1.08	0.47	0.67	0.38	0.97
GER	1.04	0.01									GER	0.86	0.01								
GERMANY-UK LEFT TAIL										GERMANY-UK RIGHT TAIL											
N = 75	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town	N = 65	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town
GER	0.94	0.53	0.01	0.52	0.97	1.34	0.39	0.51	0.46	0.90	GER	1.10	0.99	0.01	0.47	0.82	1.53	0.33	0.57	0.49	0.85
UK	0.97	0.01									UK	0.78	0.01								
UK-US LEFT TAIL										UK-US RIGHT TAIL											
N = 67	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town	N = 70	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town
UK	2.29	1.08	0.01	0.20	1.61	3.77	0.07	0.33	0.28	0.73	UK	2.54	0.01	0.01	0.17	1.70	4.25	0.05	0.31	0	0.35
US	1.31	0.01									US	1.42	0.01								
GERMANY-US LEFT TAIL										GERMANY-US RIGHT TAIL											
N = 72	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town	N = 68	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town
GER	2.54	0.91	0.01	0.17	1.77	4.39	0.05	0.29	0	0.27	GER	3.80	1.38	0.01	0.07	2.43	0.19			0	0.45
US	1.71	0.01									US	1.18	0.01								
FRANCE-US LEFT TAIL										FRANCE-US RIGHT TAIL											
N = 67	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town	N = 68	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town
FRAN	2.43	1.05	0.01	0.19	1.75	3.66	0.08	0.30	0.26	0.95	FRAN	2.42	1.61	0.01	0.19	1.27	4.38	0.05	0.42	0	0.57
US	1.30	0.01									US	1.27	0.01								
JAPAN-US LEFT TAIL										JAPAN-US RIGHT TAIL											
N = 71	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town	N = 70	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town
JAP	2.29	1.24	0.01	0.20	1.63	3.71	0.08	0.32	0	0.23	JAP	4.01	1.16	0.01	0.06	2.64	0.16			0	0.51
US	1.19	0.01									US	1.43	0.01								
JAPAN-UK LEFT TAIL										JAPAN-UK RIGHT TAIL											
N = 70	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town	N = 68	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town
JAP	3.83	1.25	0.01	0.07	2.62	0.16			0	0.27	JAP	3.00	1.20	0.01	0.12	2.02	0.25			0	0.25
UK	1.30	0.01									UK	1.17	0.01								
FRANCE-JAPAN LEFT TAIL										FRANCE-JAPAN RIGHT TAIL											
N = 70	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town	N = 68	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town
FRAN	3.44	1.27	0.01	0.09	2.37	0.19			0	0.49	FRAN	3.65	2.00	0.02	0.08	2.45	0.18			0	0.37
JAP	0.52	0.01									JAP	0.89	0.01								
GERMANY-JAPAN LEFT TAIL										GERMANY-JAPAN RIGHT TAIL											
N = 72	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town	N = 68	$\theta$	$\alpha$	$\xi$	$2^\theta$	$\theta_{15\%}$	$\theta_{5\%}$	$\chi_{15\%}$	$\chi_{5\%}$	$\chi$ Poon et al.	$\bar{\chi}$ Ledford Town
GER	4.15	0.58	0.01	0.06	2.00	0.16			0	0.31	GER	4.00	1.81	0.02	0.06	2.64	0.16			0	0.23
JAP	1.20	0.01									JAP	1.31	0.01								

Table C.11: Bayesian estimations for pair Indices and 95% Credibility Intervals for  $\theta$ , and other extreme estimations (average computing time: 21.8s).

FRANCE-GERMANY-UK. Negative								FRANCE-GERMANY-UK. Positive							
N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION	N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION
67	18	0.954	0	0	0	0.046	TRIPLE DEP	73	13	0.943	0	0	0	0.057	TRIPLE DEP
GERMANY-JAPAN-US. Negative								GERMANY-JAPAN-US. Positive							
N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION	N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION
61	1	0.942	0.007	0	0	0.051	TRIPLE DEP	61	5	0.820	0.006	0.006	0.003	0.166	TRIPLE DEP
JAPAN-UK-US. Negative								JAPAN-UK-US. Positive							
N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION	N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION
61	1	0.932	0.001	0	0	0.067	TRIPLE DEP	60	1	0.930	0.027	0	0.001	0.042	TRIPLE DEP
FRANCE-JAPAN-US. Negative								FRANCE-JAPAN-US. Positive							
N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION	N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION
62	1	0.001	0	0	0	0.999	INDEPENDENCE	58	2	0.837	0	0.069	0	0.093	TRIPLE DEP
FRANCE-GERMANY-JAP. Negative								FRANCE-GERMANY-JAP. Positive							
N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION	N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION
69	3	0.002	0	0	0.999	0.029	JAP INDEP	63	1	0.001	0	0	0.990	0.040	JAP INDEP
FRANCE-GERMANY-US. Negative								FRANCE-GERMANY-US. Positive							
N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION	N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION
61	5	0.002	0	0	0.924	0.074	US INDEP	65	4	0.004	0	0	0.932	0.064	US INDEP
FRANCE-JAPAN-UK. Negative								FRANCE-JAPAN-UK. Positive							
N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION	N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION
62	1	0.002	0	0.958	0	0.040	JAP INDEP	65	2	0.001	0	0.968	0	0.031	JAP INDEP
FRANCE-UK-US. Negative								FRANCE-UK-US. Positive							
N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION	N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION
68	5	0.013	0	0	0.904	0.083	US INDEP	62	8	0.011	0	0	0.809	0.090	US INDEP
GERMANY-JAPAN-UK. Negative								GERMANY-JAPAN-UK. Positive							
N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION	N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION
62	1	0.001	0	0.934	0	0.065	JAP INDEP	67	1	0.001	0	0.942	0	0.057	JAP INDEP
GERMANY-UK-US. Negative								GERMANY-UK-US. Positive							
N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION	N	N ∈ R <sub>111</sub>	M <sub>5</sub>	M <sub>4</sub>	M <sub>3</sub>	M <sub>2</sub>	M <sub>1</sub>	SELECTION
62	5	0.005	0	0.001	0.957	0.037	US INDEP	65	3	0.018	0	0	0.958	0.023	US INDEP

Table C.12: Bayesian Model Selection for trivariate Indices (average computing time: 22.2s).

FRANCE-GERMANY-UK, Negative								FRANCE-GERMANY-UK, Positive									
Parameter Estimation, $M_3$								Parameter Estimation, $M_3$									
	$\theta$	$\alpha$	$\xi$	$\chi_3$	$\theta_{(1\%)}$	$\theta_{(5\%)}$	$\chi_{3(10\%)}$	$\chi_{3(5\%)}$		$\theta$	$\alpha$	$\xi$	$\chi_3$	$\theta_{(1\%)}$	$\theta_{(5\%)}$	$\chi_{3(10\%)}$	$\chi_{3(5\%)}$
FRAN	0.88	0.54	0.01	0.38	0.07	1.10	0.27	0.48	FRAN	0.90	0.78	0.01	0.37	0.07	1.23	0.26	0.48
GER		0.57	0.01						GER		0.71	0.01					
UK		1.44	0.01						UK		0.88	0.01					
GERMANY-JAPAN-US, Negative								GERMANY-JAPAN-US, Positive									
Parameter Estimation, $M_3$								Parameter Estimation, $M_3$									
	$\theta$	$\alpha$	$\xi$	$\chi_3$	$\theta_{(1\%)}$	$\chi_{3(5\%)}$			$\theta$	$\alpha$	$\xi$	$\chi_3$	$\theta_{(1\%)}$	$\chi_{3(5\%)}$			
GER	2.38	0.88	0.01	0.07	1.70	0.14		GER	3.33	1.49	0.02	0.03	2.28	0.08			
JAP		1.26	0.01					JAP		0.70	0.01						
US		1.15	0.01					US		2.21	0.02						
JAP-UK-US, Negative								JAP-UK-US, Positive									
Parameter Estimation, $M_3$								Parameter Estimation, $M_3$									
	$\theta$	$\alpha$	$\xi$	$\chi_3$	$\theta_{(1\%)}$	$\chi_{3(5\%)}$			$\theta$	$\alpha$	$\xi$	$\chi_3$	$\theta_{(1\%)}$	$\chi_{3(5\%)}$			
JAP	2.39	1.23	0.01	0.07	1.70	0.15		JAP	2.54	1.51	0.02	0.06	1.85	0.13			
UK		1.14	0.01					UK		1.10	0.01						
US		1.18	0.01					US		1.24	0.01						
FRANCE-JAPAN-US, Positive																	
Parameter Estimation, $M_3$																	
	$\theta$	$\alpha$	$\xi$	$\chi_3$	$\theta_{(1\%)}$	$\chi_{3(5\%)}$											
FRAN	2.46	1.58	0.02	0.07	1.77	0.14											
JAP		1.17	0.01														
US		1.36	0.01														

Table C.13: Bayesian Trivariate estimates of triple tail dependence models.

OPTIMALLY ALLOCATED PORTFOLIO	
FRANCE	GERMANY
0.91	0.09
FRANCE	JAPAN
0.49	0.51
FRANCE	UK
0	1
FRANCE	US
0.40	0.60
GERMANY	JAPAN
0.34	0.66
GERMANY	UK
0	1
GERMANY	US
0.34	0.66
JAPAN	UK
0.43	0.57
JAPAN	US
0.40	0.60
UK	US
0.52	0.48

Table C.14: Optimum allocations of different pairs of Global Indices.

PREDICTIVE MEAN DAILY EXCESS LOSS 10% TAIL		
FRANCE	GERMANY	$u_L$
3.81%	3.87%	2.66%
FRANCE	UK	$u_L$
3.27%	3.06%	2.21%
GERMANY	UK	$u_L$
3.68%	3.23%	2.35%

Table C.15: Predictive mean daily excess loss over 90% loss threshold of original pairs of Global Indices.

# Appendix D

## Figures

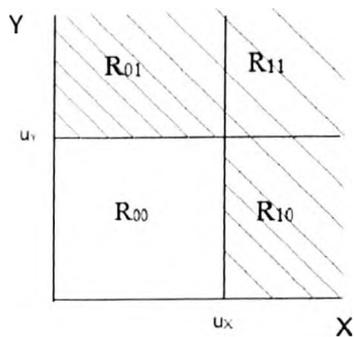


Figure D.1: Bivariate regions:  $R_{01}$ ,  $R_{10}$ , and  $R_{11}$ .

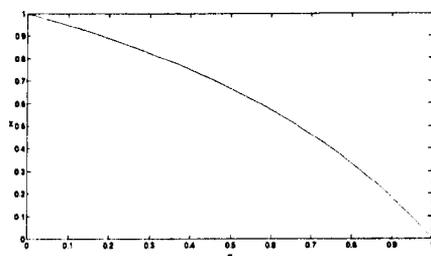
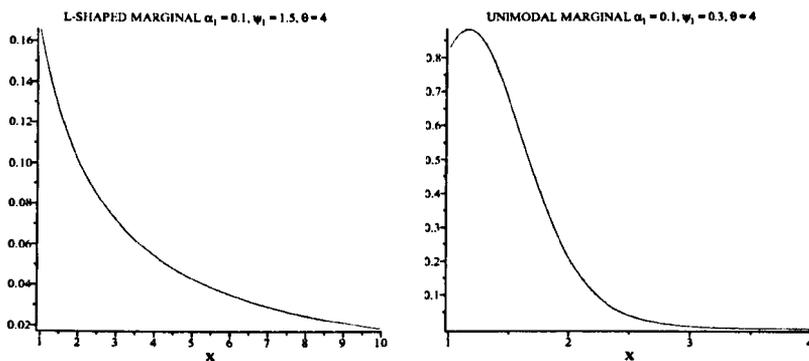
Figure D.2:  $\chi$  in the simple exponential case.

Figure D.3: Examples of Model 1 Marginals.

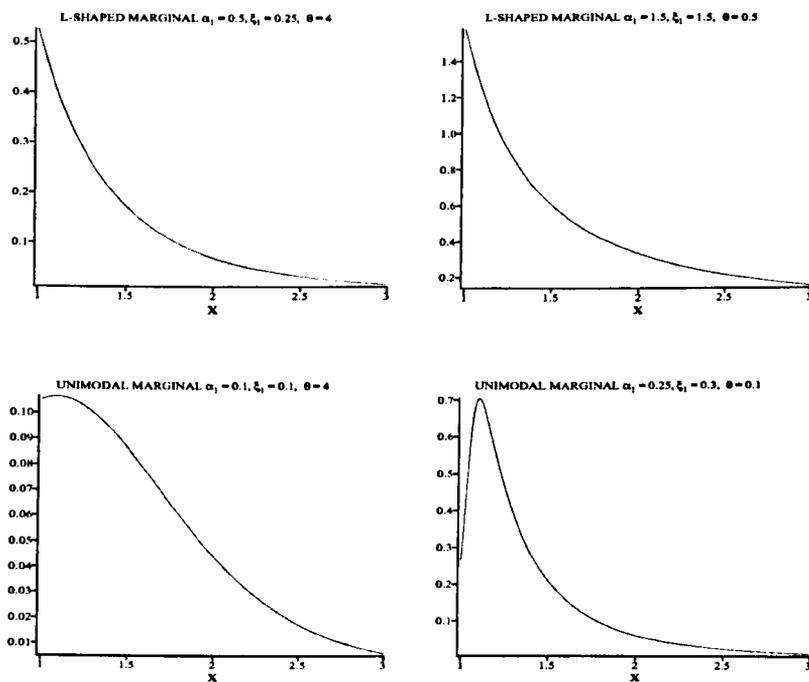
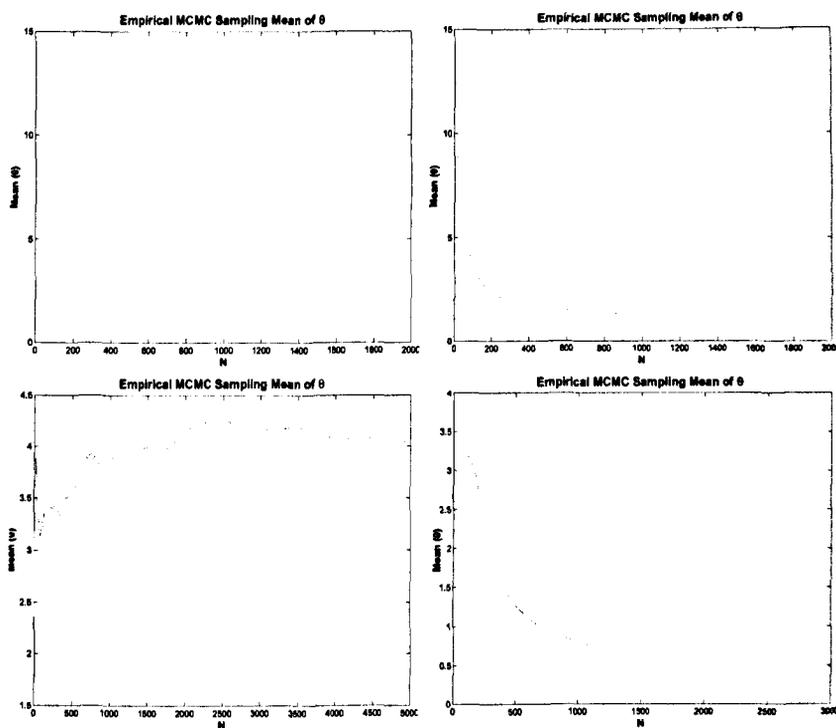


Figure D.4: Examples of Model 2 Marginals.

Figure D.5: MCMC sample means for  $\theta$ .

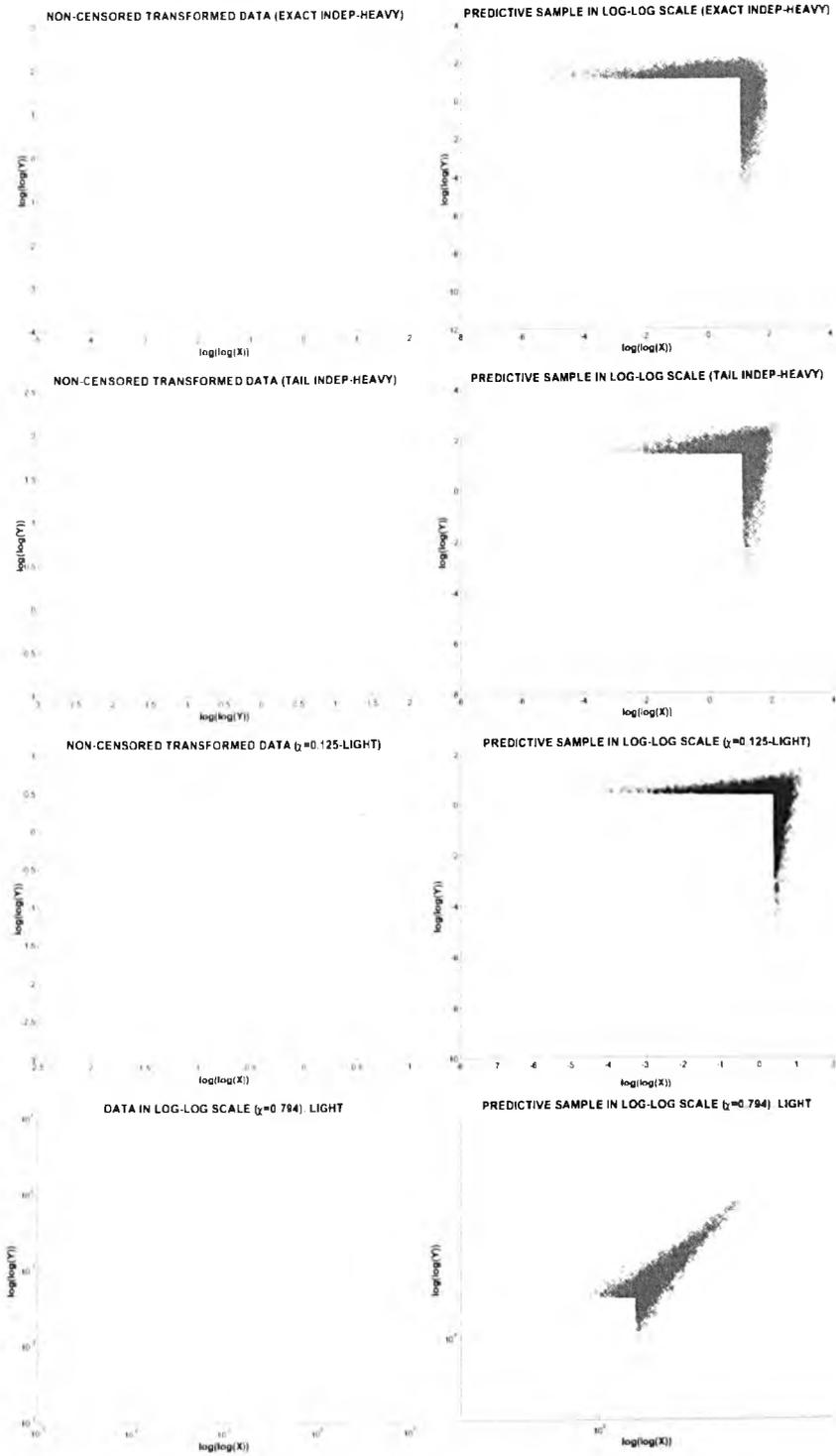


Figure D.6: Bayesian predictions for data simulations. Predictions vs data in log-log scale.

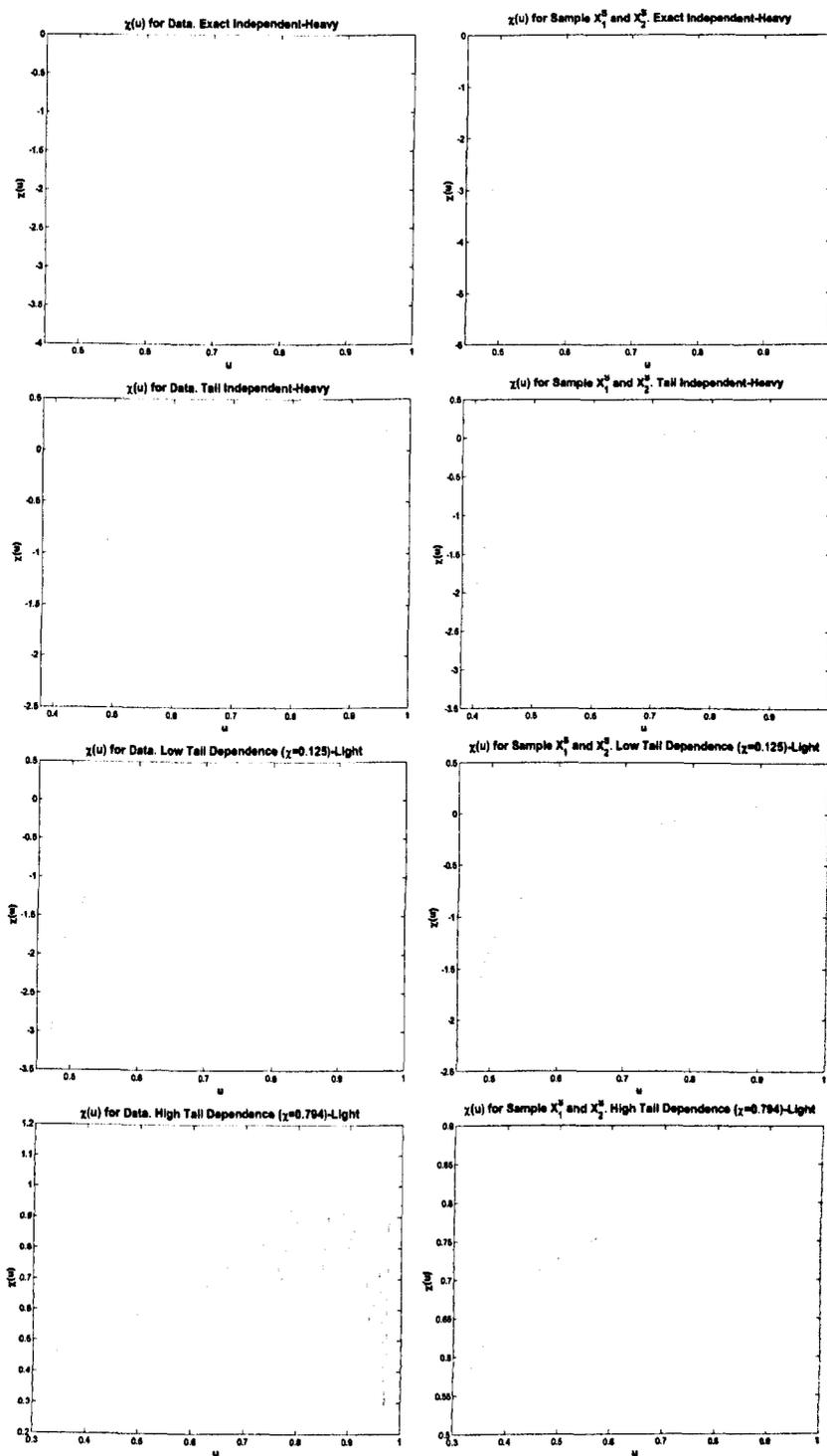


Figure D.7: Bayesian predictions for  $\chi(u)$  vs empirical  $\chi(u)$ .

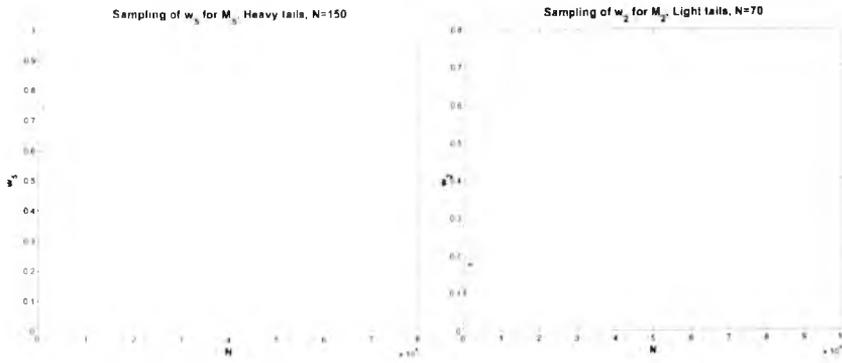


Figure D.8: Sampling processes of different weights for simulations of the trivariate model.

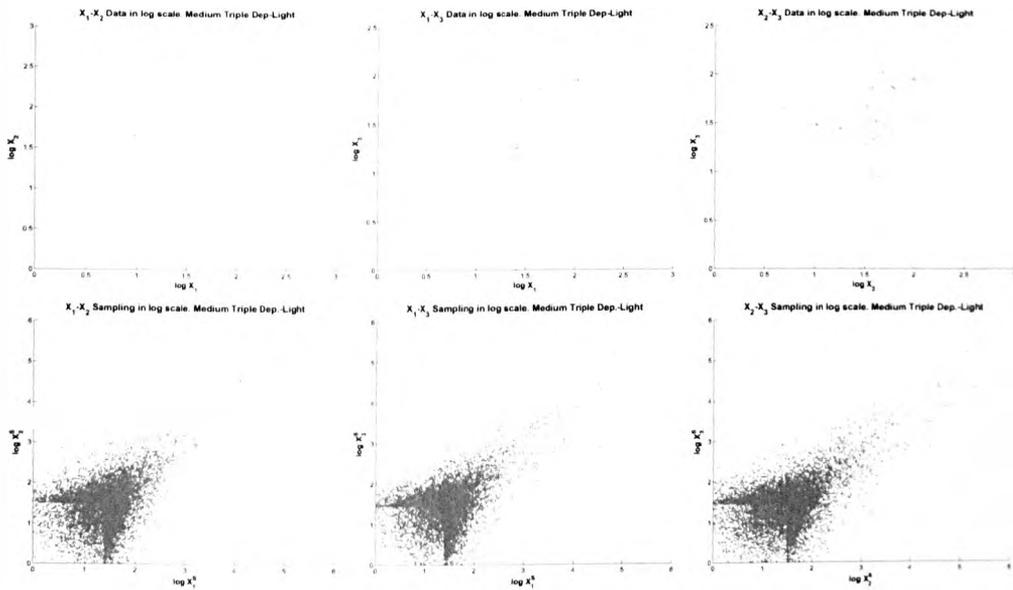


Figure D.9: Simulated data and predictive sampling for trivariate simulations: medium triple tail dependence.

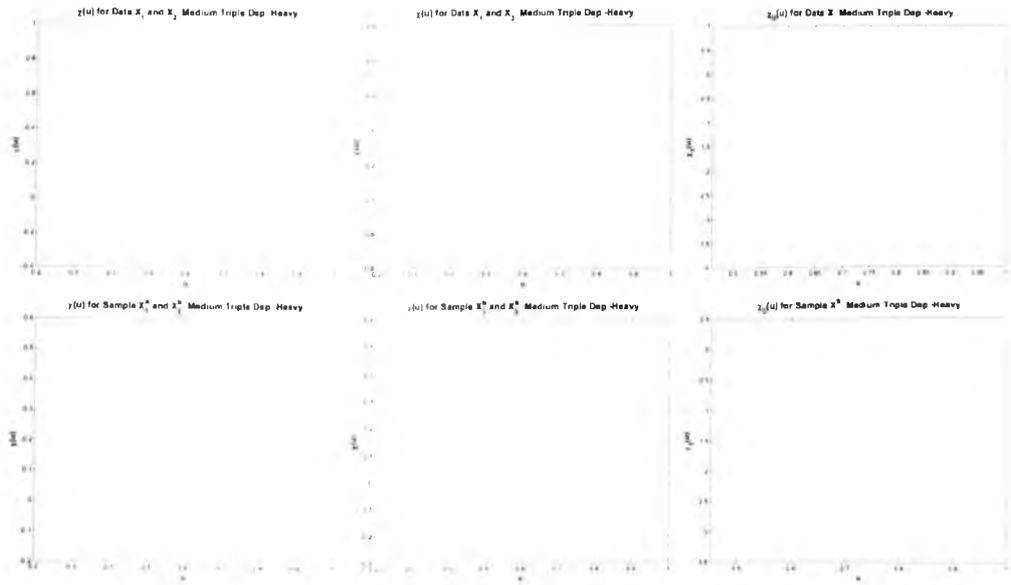


Figure D.10: Simulated and predictive  $\lambda(u)$  and  $\lambda_3(u)$  for trivariate simulations: triple medium tail dependence.

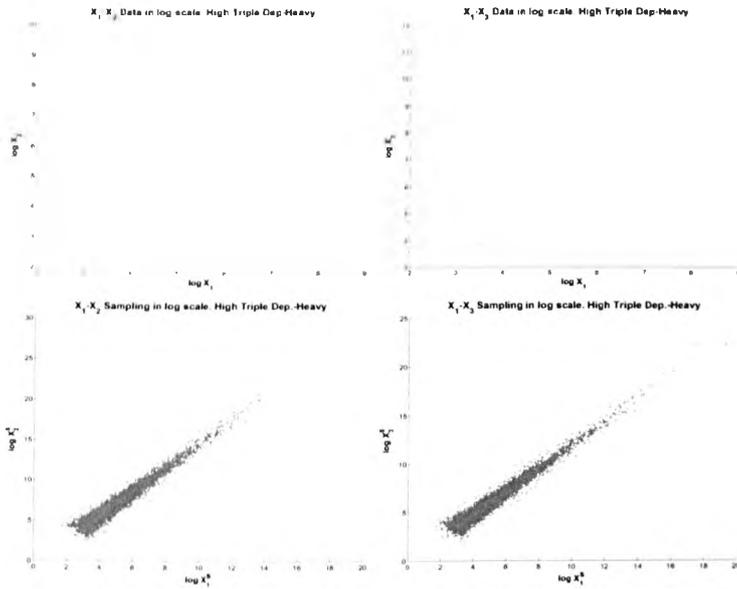


Figure D.11: Simulated data and predictive sampling for trivariate simulations: triple high tail dependence

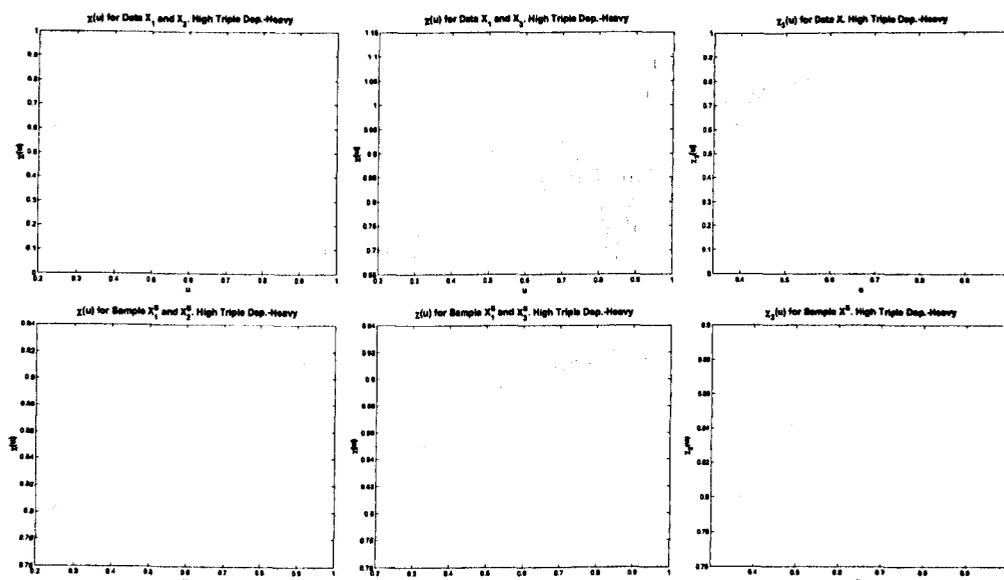


Figure D.12: Simulated and predictive  $\chi(u)$  and  $\chi_3(u)$  for trivariate simulations: triple high tail dependence.

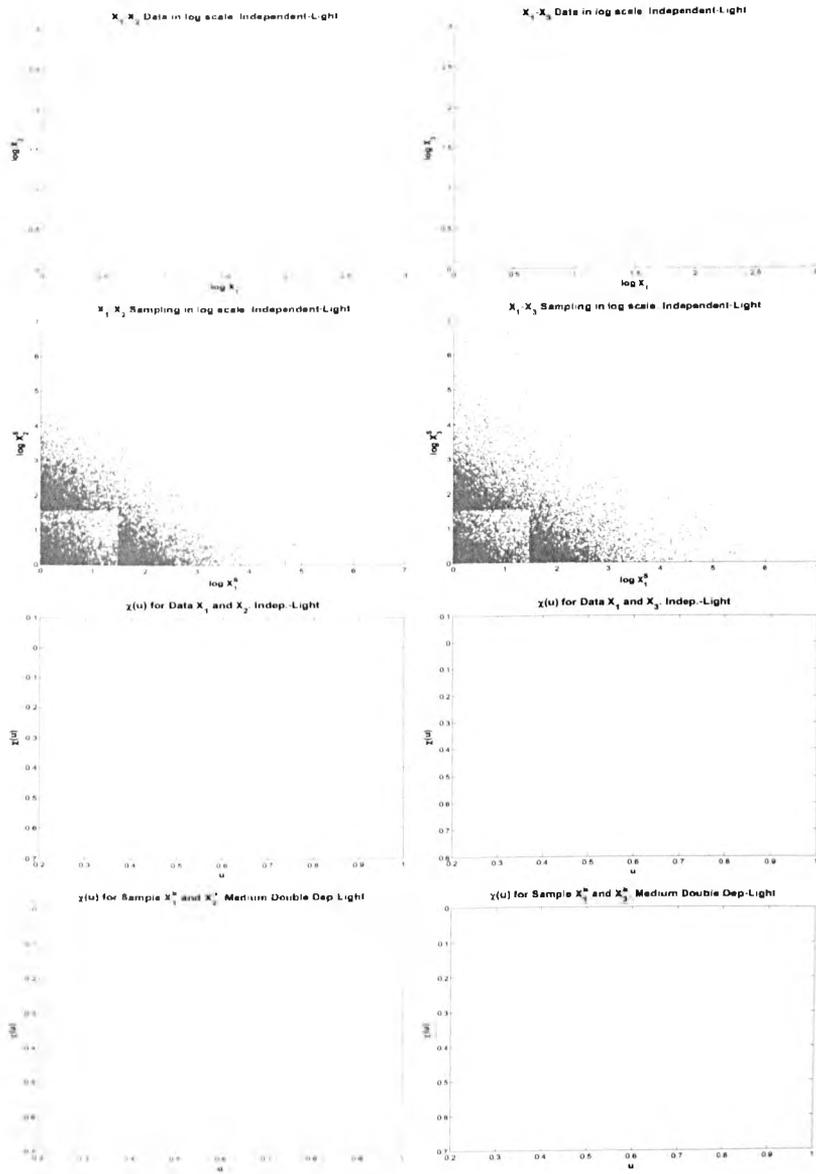


Figure D.13: Simulated data and predictive sampling related graphics for trivariate simulations: exact independence

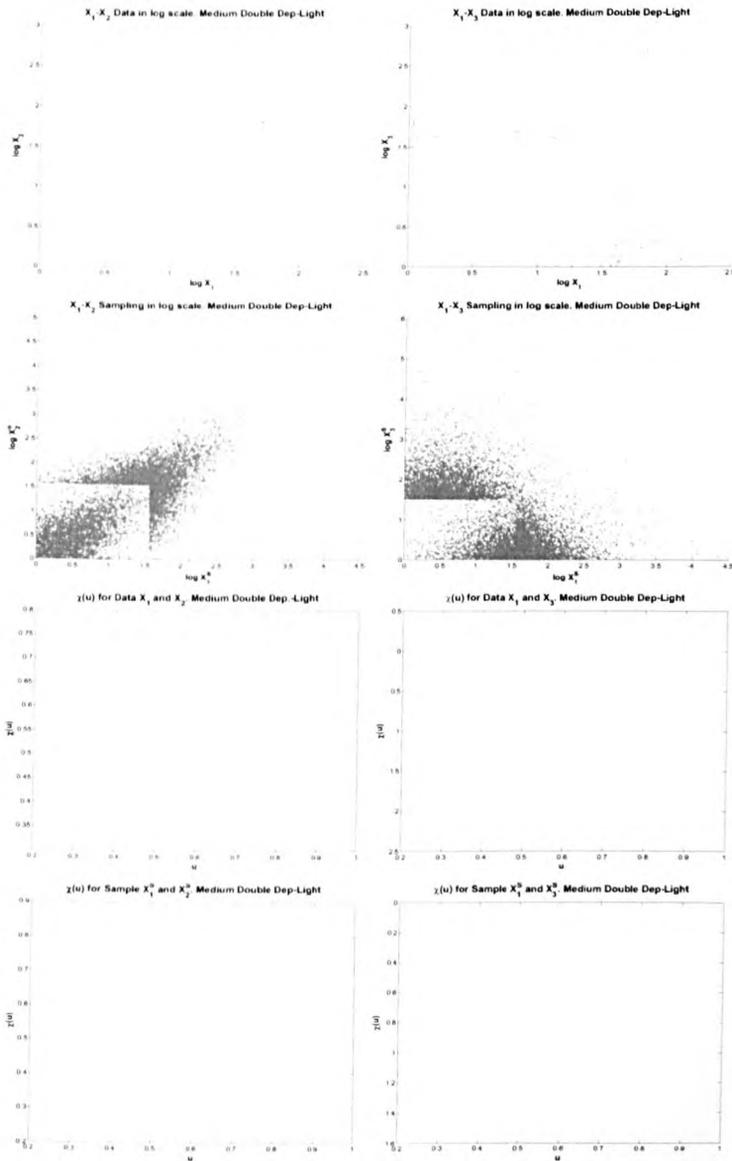


Figure D.14: Simulated data and predictive sampling related graphics for trivariate simulations: medium double tail dependence ( $X_3$  exact independent).

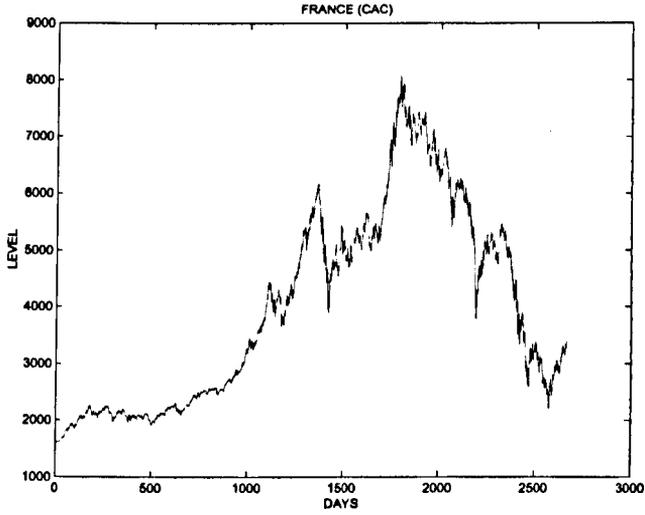


Figure D.15: France CAC 40.

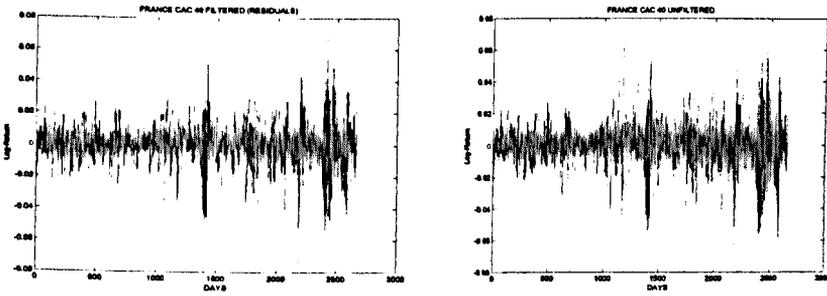


Figure D.16: France CAC 40 filtered and unfiltered series.

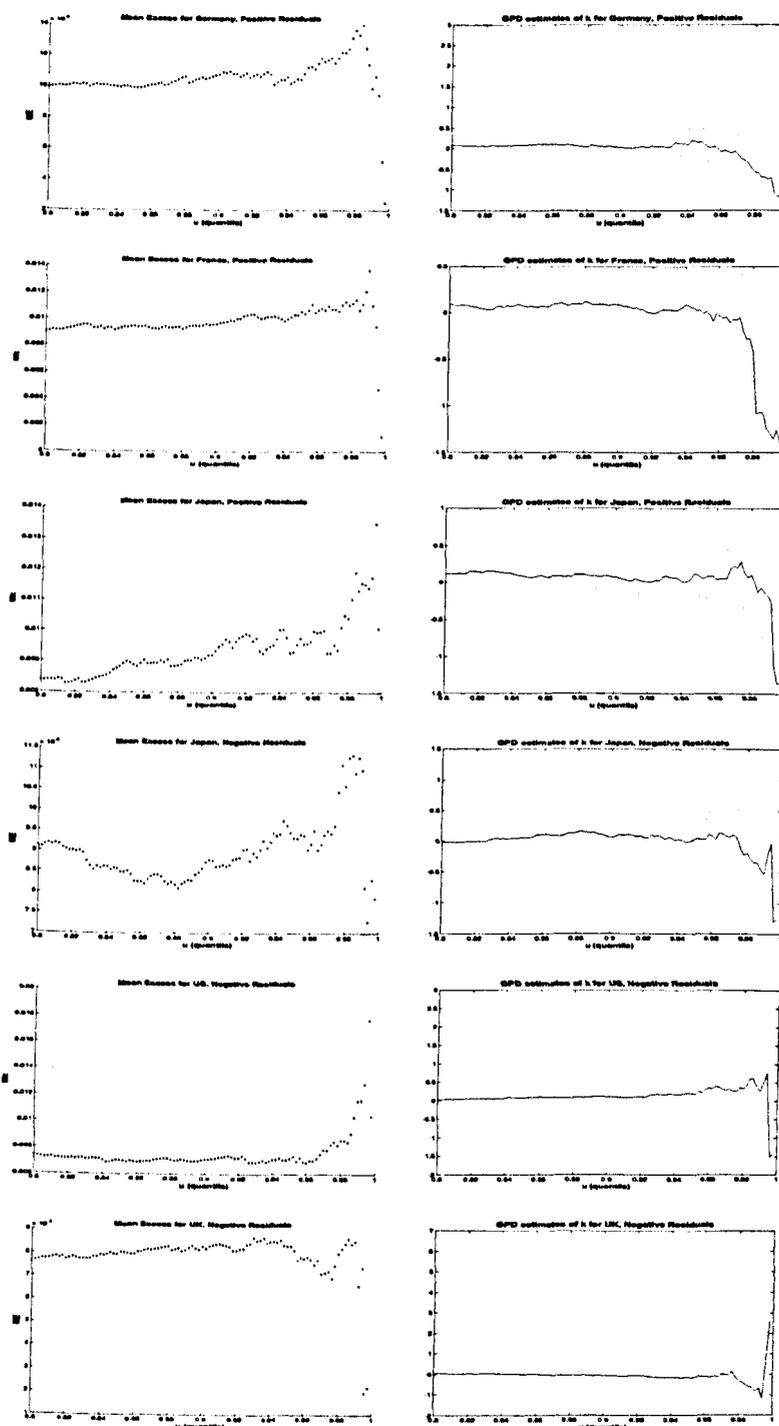


Figure D.17: Threshold selection methods. On the left side, the Mean Excess. On the right side, estimations of  $k$  and its 95% confidence intervals.

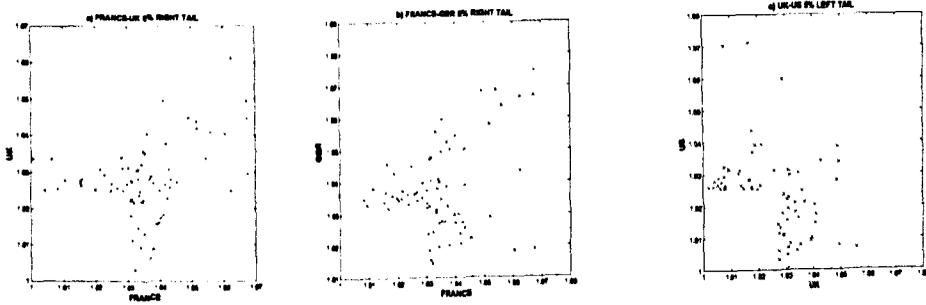


Figure D.18: Scatterplot of excesses over the 0.95 quantile of the filtered pairs: a) France-UK, b) France-Germany, and c) UK-US.

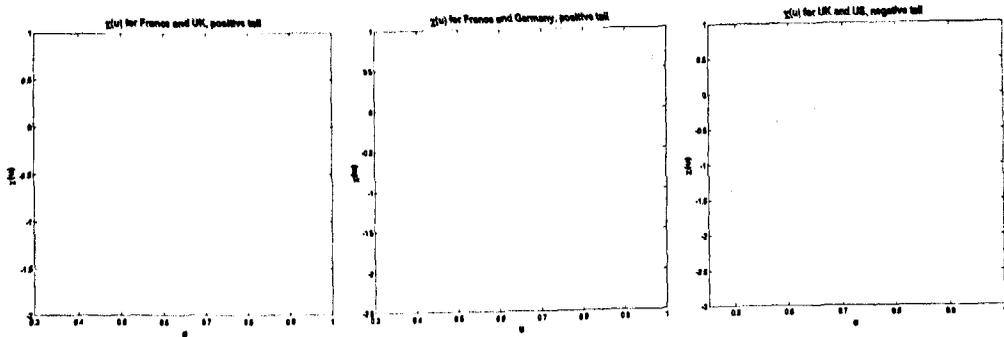


Figure D.19:  $\chi(u)$  of filtered pairs: a) France-UK, b) France-Germany, and c) UK-US.

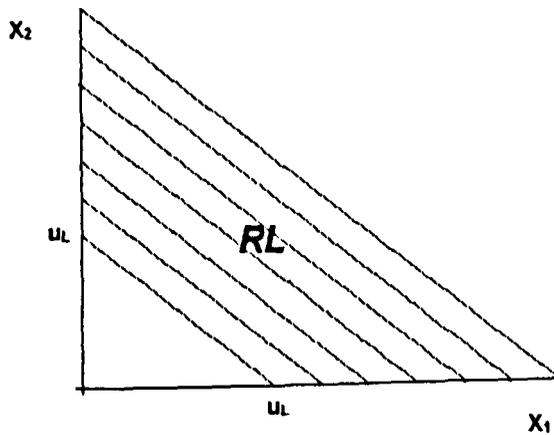


Figure D.20: Portfolio analysis tail region *RL*.

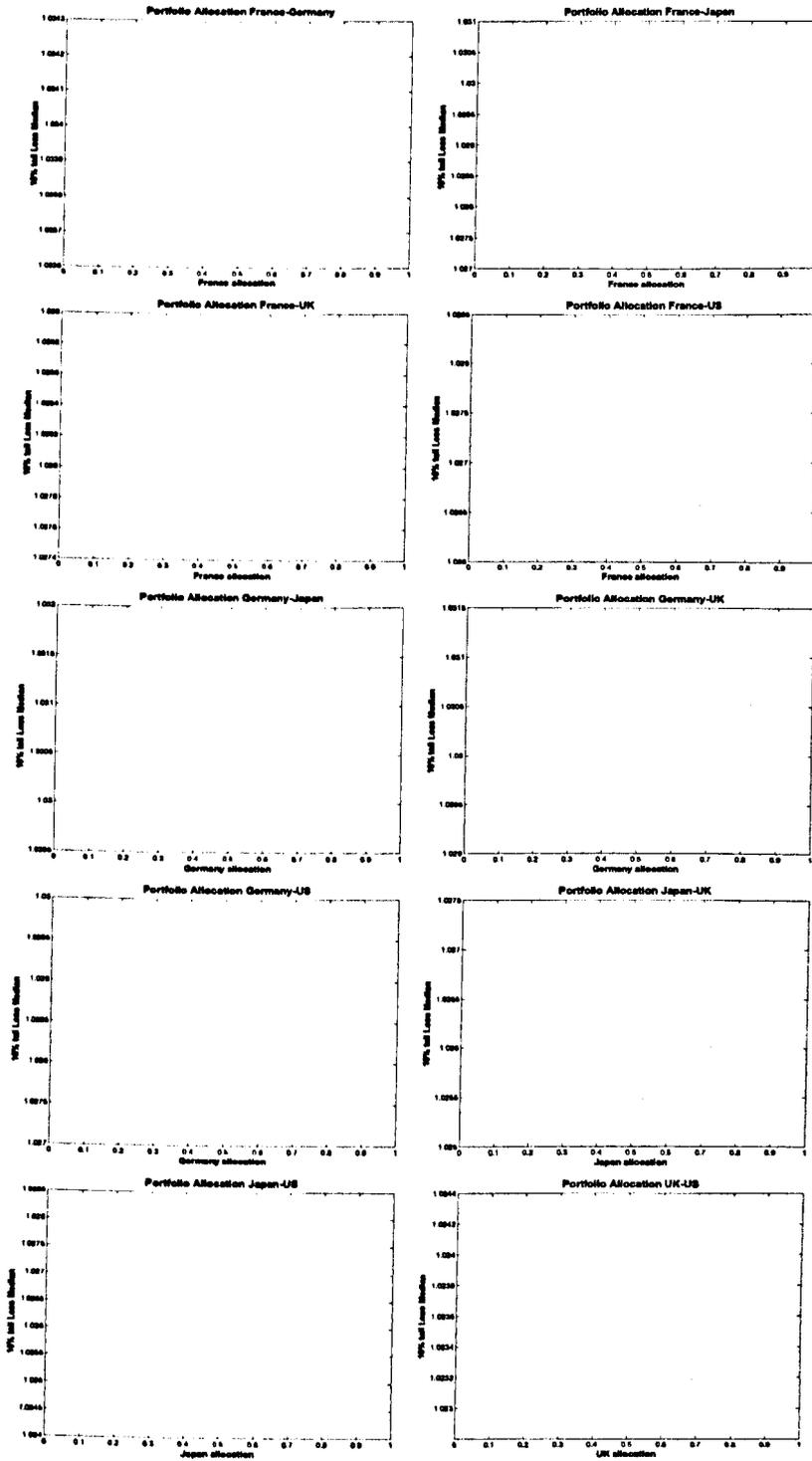


Figure D.21: Portfolio optimum allocations plots for different Global Indices.

Computing time: 184.5 minutes per pair.

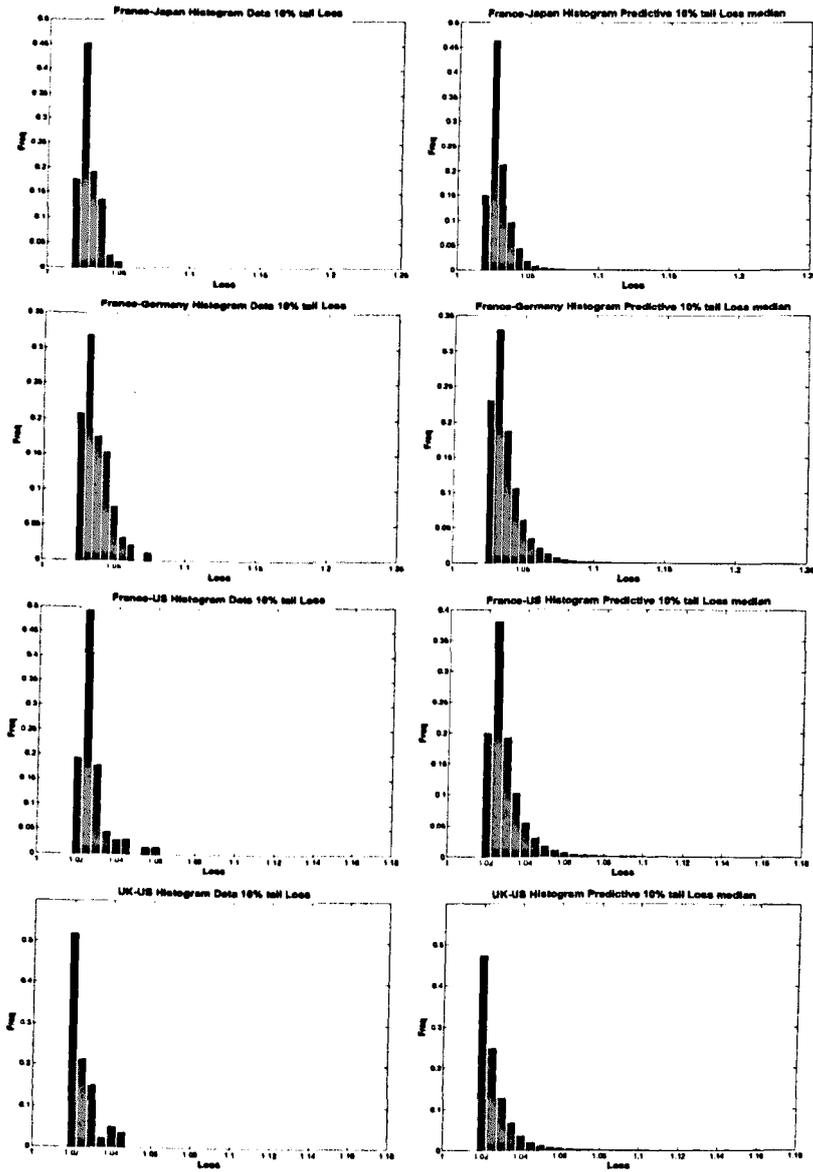


Figure D.22: Portfolio Loss histograms. Left hand: historical, right hand: predictive.

# Bibliography

- Ahmed, M.S.(1961). "On a Locally Most Powerful Boundary Randomized Similar Test for the Independence of two Poisson Variables". *Ann. Math. Statist.*, **32**, pp. 809-827.
- Apputhurai, P. and Stephenson, A.G.(2011). "Accounting for Uncertainty in Extremal Dependence Modeling using Bayesian Model Averaging Techniques". *Journal of Statistical Planning and Inference.*, **141**, pp. 1800-1807.
- Balkema, A.A. and de Haan L. (1974). "Residual Life Time at Great Age". *The Annals of Probability*, **5**, pp. 792-804.
- Barbie P. and McCormick W.P. (2005). "Asymptotic Expansions of Convolutions of Regularly Varying Distributions". *J. Aust. Math. Soc.*, **78**, pp. 1-33.
- Bardsley, W. E. (1977). "A Test for Distinguishing Between Extreme Value Distributions". *Journal of Hydrology*, **34**, pp. 377-381.

- Beirlant, J., de Wet, T. and Goegebeur, Y. (2006). "A Goodness-of-fit Statistic for Pareto-type Behaviour". *J. Comput. Appl. Math.*, **186**, pp. 99-116.
- Bernardo, J. and Adrian F.M. Smith (1995). *Bayesian Theory*. Chichester : John Wiley and Sons.
- Bingham, N.H., Goldie, C.M., and Teugels, J.L., (1987). *Regular Variation*. Cambridge: Cambridge University Press.
- Birkhoff, G.D.(1942). "What is the Ergodic Theorem?" *The American Mathematical Monthly*, **49**, No. 4, pp. 222-226.
- Bottolo, L., Consonni, G., Dellaportas, P. and Lijoi, A. (2003). "Bayesian Analysis of Extreme Values by Mixture Modeling". *Extremes*, **6**, pp. 25-47.
- Burr, I.W.(1942). "Cumulative Frequency Functions". *The Annals of Mathematical Statistics*, **13**, No. 2, pp. 215-232.
- Box, G.E.P. and Tiao G.C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley Pub. Co.
- Casella, G. and Robert, C.P. (1997). *Monte Carlo statistical methods*. New York: Springer.
- Choulakian, V. and Stephens, M. A. (2001). "Goodness-of-Fit Tests for the Generalized Pareto Distribution". *Technometrics*, **43**, No. 4, pp. 478-484.
- Clayton, D.G.(1978). "A Model for Association in Bivariate Life Tables and

- Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence". *Biometrika*, **65**, No. 1, pp. 141-151.
- Coles, S.G. (2001). "An introduction to statistical modeling of extreme values". London: Springer.
- Coles, S. G., Heffernan, J. and Tawn, J.(1999). "Dependence Measures for Extreme Value Analyses". *Extremes*, **2**, pp. 339-365.
- Coles, S.G. and Pauli, (2002). "Models and Inference for Uncertainty in Extremal Dependence". *Biometrika*, **89**, No. 1, pp. 183-196.
- Coles, S.G. and Powell, E. (1996). "Bayesian Methods in Extreme Value Modelling: A Review and New Developments". *International Statistical Review*, **64**, No. 1, pp. 119-136.
- Coles, S.G. and Tawn, J. (1990). "Statistics of Coastal Flood Prevention". *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, **332**, No. 1627. pp. 457-476.
- Coles, S.G. and Tawn, J. (1991). "Modelling Extreme Multivariate Events". *J. R. Statist. Soc. B*, **53**, pp. 377-392.
- Coles, S.G., and Tawn, J. (1996). "A Bayesian Analysis of Extreme Rainfall Data". *Applied Statistics*, **45**, No. 4, p.p. 463-478.
- Coles, S.G. and Tawn, J (2005). "Bayesian Modelling of Extreme Surges on

- the UK East Coast". *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, **363**, pp. 1387-1406.
- Cooley, D., Doug, N., and Naveau, P. (2007). "Bayesian Spatial Modeling of Extreme Precipitation Return Levels". *Journal of the American Statistical Association*. **102**, pp. 824-840.
- Dagsvik, J. (1988) "Markov chains generated by maximizing components of multidimensional extremal processes". *Stoch. Processes Applic.*, **28**, 31-45.
- Davison, A.C. and Smith, R.L. (1990) "Models for Exceedances over High Threshold". *J.R. Statist. Soc. B*, **52**, No.3, 393-442.
- de Haan, L. (1977) "On Regular Variation and Its Application to the Weak Convergence of Sample Extremes". *Mathematical Centre Tract*, **32**, Mathematics Centre, Amsterdam, Holland.
- de Haan, L., and Resnick, S.I. (1977) "Limit Theory for Sample Extremes". *Z. Wahrscheinlichkeitstheorie*, **40**, pp. 317-337.
- de Haan, L. (1984) "A Spectral Representation for Max-stable Processes". *The Annals of Probability*, **12**, No. 4, pp. 1194-1204.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events: for Insurance and Finance*. Springer, Berlin Heidelberg New York.
- Falk, M. and Michel, R. (2006). "Testing for tail independence in extreme

- value models". *Annals of the Institute of Statistical Mathematics*, **58**, pp. 261-290.
- Ferraz F., Gamerman D. and Freitas H. (2011). "A semiparametric Bayesian approach to extreme value estimation". *Pakistan Journal of Statistics, Series A*, **9**, pp. 1-24.
- Fisher, R. A. and Tippett, L.H.C. (1928). "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample". *Proc. Cambridge Philos. Soc.*, **24**, pp. 180-190.
- Fry, T.R.L.(1993). "Univariate and Multivariate Burr Distributions: A Survey". *Pakistan Journal of Statistics. Series A*, **9**, pp. 1-24.
- Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2003). *Bayesian Data Analysis*. 2nd ed. Chapman & Hall.
- Glosten, L. R., Jagannathan, R., and Runkle D.E. (1993). "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks". *Journal of Finance. American Finance Association*, **48**, No. 5, December, pp. 1779-1801.
- Green, P.J. (1995). "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination". *Biometrika*, **92**, No. 4, pp. 711-732.
- Guillote, S., and Perron F. (2008). "A Bayesian Estimator for the Dependence

Function of a Bivariate Extreme-Value Distribution". *The Canadian Journal of Statistics*, **36**, No. 3, pp. 383-396.

Gumbel, E.J. (1958) *Statistics of Extremes*. New York: Columbia University Press.

Heffernan, J.E. (2000). "A Directory of Coefficients of Tail Dependence". *Extremes*, **3**, pp. 279-290.

Hill, B.M. (1975). "A Simple General Approach to Inference about the Tail of a Distribution". *Ann. Statist.*, **3**, pp. 1163-1174.

Hosking, J.R.M. (1984). "Testing whether the shape parameter is zero in the Generalized Extreme Value Distribution". *Biometrika*, **71**, pp. 367-374.

Jenkinson, A.F. (1955). "The frequency distribution of the annual maximum (or minimum) of meteorological elements". *Quart. J. R. Met. Soc.*, **81**, pp. 158-171.

Joe, H. (1990). "Families of min-stable multivariate exponential and multivariate extreme value distributions.". *Statist. Probab. Letters*, **9**, pp. 75-81.

Joe, H. (1994). "Multivariate Extreme Value Distributions with Applications to Environmental Data". *Canad J. Statistics*, **22**, pp. 47-64.

Joe, H. (1997). "Multivariate Models and Dependence Concepts". Chapman and Hall, London.

- Johnson, N.L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, 1, 2nd ed. John Wiley, New York.
- Kotz, S. and Nadarajah S. (2002). *Extreme Value Distributions: Theory and Applications*. London: Imperial College River Edge. N.J.: World Scientific.
- Leamer, E.E. (1978). *Specification searches : ad hoc inference with nonexperimental data*. New York: Wiley.
- Ledford, A.W. and Tawn, J. (1996). "Statistics for Near Independence in Multivariate Extreme Values". *Biometrika*, **83**, pp. 169-187.
- Ledford, A.W. and Tawn, J. (1997). "Modelling Dependence within Joint Tail Regions". *Journal of Royal Statistical Society Series B*, **59**, No. 2, pp. 475-499.
- Lindley, D.V. (1985). *Making Decisions*. 2nd ed. J. Wiley, New York.
- McFadden, D. (1978). "Modelling the choice of residential location". In *Spatial Interaction Theory and Planning Models*, Ed. A. Karlqvist, L. Lundquist, F. Snickers and J. Weibull, pp. 75-96. Amsterdam: North-Holland.
- Mikosch, T. (2005). "Copulas: Tales and Facts". *Extremes*, **9**, pp. 3-20.
- Newey, W.K. and McFadden, D. (1994). *Large Sample Estimation and Hypothesis Testing*. In *Handbook of Econometrics*, Ch. 36. Elsevier Science.
- Otten, A. and Van Montfort, M.A.J. (1978). "The Power of Two Tests on the Type of Distributions of Extremes". *Journal of Hydrology*, **37**, pp. 195-199.

- Pickands, J. (1975). "Statistical Inference Using Extreme Order Statistics". *Annals of Statistics*, **3**, pp. 119-131.
- Pickands, J. (1981). "Multivariate Extreme Value Distributions". *Proc. 43rd Session of the ISI, Buenos Aires*. **49**, pp. 859-878.
- Pickands, J. (1997). "Bayes Quantile Estimation and Threshold Selection for the Generalized Pareto Family". *Extreme Value Theory and Applications*. Ed. J. Galambos, J. Lechner, y E. Simiu. Dordrecht: Kluwer. p.p. 123-138.
- Poon, S. H., Rockinger M., and Tawn, J. (2004). "Extreme Value Dependence in Financial Markets: Diagnostics, Models and Financial Implications". *Review of Financial Studies*, **17**, No. 2, pp. 581-610.
- Raftery A., Madigan D., and Volinsky C.T. (1995). "Accounting for Model Uncertainty in Survival Analysis Improves Predictive Performance". *Bayesian Statistics 5*, Oxford University Press, pp. 323-349.
- Ramos, A. and Ledford A. (2009). "A New Class of Models for Bivariate Joint Tails". *Journal of The Royal Statistical Society Series B-statistical Methodology*, **71**, No. 1, pp. 219-241.
- Resnick, S.I. (1987). *Extreme Values, Point Processes and Regular Variation*. Springer-Verlag, New York.
- Robert, C.P. (1994). *The Bayesian Choice: A Decision-Theoretic Motivation*. Springer-Verlag, New York.

- Rodriguez, R.N. (1960). "A Guide to the Burr Type XII Distributions". *Biometrika*, **64**, pp. 129-134.
- Samuelson, P.A. (1948). *Foundations of Economic Analysis*. Cambridge, Harvard University Press.
- Sarabia, J.M. and Gómez-Déniz, E. (2008). "Construction of Multivariate Distributions: a Review of some Recent Results,". *Statistics and Operations Research Transactions*, **32**, pp. 3-36.
- Schervish, Mark J. (1995). *Theory of Statistics*. Springer-Verlag, New York.
- Sibuya, M. (1960). "Bivariate Extreme Statistics I". *Annals of the Institute of Statistical Mathematics*, **11**, No. 2, pp. 195-210.
- Smith, A.F.M. and Roberts, G.O. (1993). "Bayesian computation via the Gibbs sampler and related Markov Chain Monte-Carlo methods". *Journal of The Royal Statistical Society Series B*, **55**, pp. 3-24.
- Smith, R.L. (1989). "Extreme Value analysis of environmental time series: an application to trend detection in ground level ozone". *Statistical Science*, **4**, p.p. 367-393.
- Smith, R.L. (1985). "Maximum Likelihood Estimation in a Class of Non-Regular Cases". *Biometrika*, **72**, pp. 67-90.
- Smith, R.L., Tawn, J., and Yuen, H.K. (1990). "Statistics of Multivariate Extremes". *International Statistical Review*, **58**, No. 1, pp. 47-58.

- Smith, R.L. (1994). "Multivariate threshold methods". In *Extreme Value Theory and Applications*, Ed. J. Galambos, J. Lechner and E. Simiu, pp. 225-48.
- Stephens, M.A. (1977). "Goodness of Fit for the Extreme Value Distribution". *Biometrika*, **64**, No. 3, pp. 583-588.
- Tadikamalla, P.R. (1980). "A Look at the Burr and Related Distributions". *International Statistical Review*, **48**, No. 3, pp. 337-344.
- Tawn, J. (1988). "Bivariate Extreme Value Theory: Models and Estimations". *Biometrika*, **75**, No. 3, pp. 397-415.
- Tawn, J. (1990). "Modelling Multivariate Extreme Value Distributions". *Biometrika*, **77**, No. 2, pp. 245-253.
- Tawn, J. (1994). "Application of Multivariate Extremes". *Extreme Value Theory and Its Applications*, eds. J.Galambos, J.Lechner, and E Simiu, Kluwer.
- Tiago de Oliveira, J. (1958). "Extremal Distributions". *Faculdade de Ciencias de Lisboa*, Ser A, **7**, pp. 215-227.
- Tiago de Oliveira, J. (1980). "Bivariate extremes; foundations and statistics". In *Multivariate Analysis V*. Ed. P. R. Krishnaiah, Amsterdam: North-Holland, pp. 349-66.
- Tierney, L. (1994). "Markov Chains for exploring Posterior Distributions". *The Annals of Statistics*, **22**, No. 4, pp. 1701-1728.

- 
- Walhin, J.F. and Paris, J. (1980). "Recursive Formulae for some Bivariate Counting Distributions Obtained by the Trivariate Reduction Method". *ASTIN Bulletin*, **30**, pp. 141-155.
- Wiener, N. (1939). "The Ergodic Theorem". *Duke Math. J.*, **5**, No. 1, pp. 1-18.