THE UNIVERSITY
of EDINBURGH

Bayesian Analysis of Jointly Heavy-Tailed Data

Karla Vianey Palacios Ramírez

Doctor of Philosophy
University of Edinburgh
2022

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(Karla Vianey Palacios Ramírez)

Acknowledgements

After almost four years of studying at the University of Edinburgh, I can say that it is a privilege being able to finish this PhD journey. I am grateful for all the support and encouragement I received during this hard process. This experience has taught me more than I ever expected.

First, I want to express my sincere and deep gratitude to my supervisor Miguel de Carvalho, who made this thesis possible. Words cannot express my gratitude, but I will try; thank you for all the encouragement, motivation and patience that helped me during all the time of research and writing, thank you for all the support and guidance provided when I had new ideas and for always believing in me. I will always cherish these four years, and I could not have imagined having a better supervisor and mentor.

Second, my deepest thank you to my family for being there through good and bad times. A special thank you to my mom, who has always supported me and encouraged me to follow my dreams, to my grandma who always listened and provided the best advice, to my sisters Abril and Sharie for being great friends during this time and motivated me, and to my dad for all the good advice. Thank you to all my friends, for the laughs during the hard times and the help provided. To Nannette, who has been there for as long as I can remember, for all your support and friendship. To Fraser, for being there with me during this journey and for all your love and caring.

I would also like to thank my PhD colleagues, from the Statistical Intelligence Laboratory for all the discussions and encouragements. I am grateful to Conacyt (the Mexican National Council for Science and Technology) for the financial support through the program "Becas al Extranjero Demanda Libre 2018-1".

My deepest thank you to Luis Gutierrez, who provided guidance and expertise for great parts of the contributions.

Finally, thank you to all the people in Edinburgh who made this possible, and received me with open arms in this beautiful country.

Abstract

This thesis develops novel Bayesian methodologies for statistical modelling of heavy-tailed data. Heavy tails are often found in practice, and yet they are an Achilles heel of a variety of mainstream random probability measures such as the Dirichlet process. The first contribution of this thesis is the study of random probability measures that can be used to model heavy-tailed data, focusing on the characterization of the tails of the so-called Pitman–Yor process, which includes the Dirichlet process as a particular case. We show that the right tail of a Pitman–Yor process, known as the stable law process, is heavy-tailed, provided that the centering distribution is itself heavy-tailed. Empowered by this finding we developed two classes of heavy-tailed mixture models and assessed their relative merits. Multivariate extensions of the proposed heavy-tailed mixtures are also devised along with a predictor-dependent version so to learn about the effect of covariates on a multivariate heavy-tailed response.

Another contribution of the thesis is a framework for learning about the frequency and magnitude of extreme values in the joint tail as well as in other risk sets induced by a suitable aggregation of variables on a common scale. The proposed framework reveals explicit links between nonstationary multivariate extremes and heteroscedastic extremes, and hence can be used to track the dynamics governing the degree of association between the extremes of a random vector over time. Bayesian inference methods are conducted for the two targets of interest: the structure scedasis function and the coefficient of tail dependence for structure processes. To learn about the structure scedasis function from data, we resort to a Bayesian nonparametric approach that defines a prior in the space of structure scedasis functions. A battery of numerical experiments is used to study the finite sample performance of all methodological developments made in this thesis. Data from the fields of neuroscience as well as from finance are used to showcase the proposed methodologies in practice.

Lay Summary

Heavy-tailed data analysis is a challenging area of study and with a lot of open questions, however considerable effort is being directed into this field. This motivates the question: Why is heavy-tailed analysis important? Heavy-tailed data are observed every day in different fields, including engineering, finance, hydrology, and medicine and there is a great need for studying heavy-tailed phenomenon.

To study and model heavy-tailed phenomenon we need a good understanding on regularly varying theory, probability theory and statistical inference. Consequently, heavy-tailed data analysis is a challenging field of study, but it is an extremely important one for real-world applications. Heavy-tailed analysis is fundamental to the study of extreme value theory and is the main topic of this thesis.

Bayesian statistics will provide flexible models and practical methods for performing the inference, and by combining theory developed in Bayesian statistics and Extreme value statistics we are able to solve interesting problems related to heavy-tailed analysis. The main idea behind this field is focused on fitting a probability model to the data of interest and summarizing the result by a probability distribution or a family of probability distributions. The space of the parameter of interest can be finite dimensional (parametric Bayesian statistics) or infinite dimensional (non-parametric Bayesian statistics), this defines the two main areas of interest in Bayesian statistics. The following references, Gelman (2014); Quintana and Müller (2004); Ghosal and Van der Vaart (2015) provide a good understanding of the standard practice and main results developed in both areas. Bayesian statistics will provide flexible models and practical methods for performing the inference, and by combining theory developed in Bayesian statistics and Extreme value statistics we are able to solve interesting problems related to heavy-tailed analysis.

Contents

Part I—Preparations

Chapter 1

Motivation and Background

This chapter offers preparations and motivations as well as technical details required to identify and solve the problems presented in this thesis.

1.1 Getting started

Heavy-tailed analysis is an area of increasing interest and of active development—both from an applied and theoretical outlook. Loosely speaking, it focuses on modelling data that with large probability can attain large values. Heavy-tailed data are observed every day in different fields, including engineering, finance, hydrology, and medicine. Although heavy-tailed data analysis is widely needed in practice, it is a challenging field of study as several standard concepts (e.g. moments) and tools (e.g. Dirichlet process) are not tailored for dealing with it. To put this into context, we demonstrate the importance of studying heavy tails by commenting on some real life applications. In risk analysis we see heavy-tailed signals all the time and they are usually related with extreme events. For instance, there is evidence showing that floods caused more than half million deaths between 1980–2009 (Doocy et al., 2013); and floods are an example of a extreme event showing a heavy tail behavior. There are a variety of other meteorological phenomena with a heavy-tailed behavior, including heatwaves, tropical storms, hurricanes, large wildfires, and other extreme weather events that affect society as well as the ecosystem (e.g. Mendes et al., 2010; Lin et al., 2010; Naveau et al., 2016; Davis and Uryasev, 2016). Also, there are numerous applications on reinsurance and finance exhibiting heavy-tailed behavior; such as, the extreme losses on a portfolio of stocks, large insurance claims, insurance payouts, and so on (e.g. Embrechts et al., 1997; Longin, 2017; Albrecher et al., 2017).

Heavy-tailed distributions go beyond risk analysis, they also play an important role in modelling data in neurology and medicine. For instance, in neuroscience are known to exhibit a heavy-tailed behavior and is related to important medical conditions, such as Alzheimer's disease, dementia, depression, anxiety, and many others (e.g. Frank, 2009; Roberts et al., 2015).

These real world applications showcase the need for developing methodologies capable of modelling heavy-tailed behavior and quantifying its risk. Motivated by the applications discussed above, a main goal of this thesis will be to address some of the open problems on heavy-tailed modelling. To do this, we will rely on one of the most active fields on statistical inference, namely that of Bayesian nonparametrics. Roughly speaking, the Bayesian nonparametric approaches are endowed with a high level of flexibility for modelling, in the sense that they are centered a priori around a parametric model but allow for the data to take over whenever there is evidence against that model. As a consequence, the heavy-tailed models developed on this thesis will lie at the interface of two fast-evolving fields of statistics—extreme value theory (Coles, 2001; Beirlant et al., 2004) and Bayesian nonparametrics (Müller et al., 2015a; Ghosal and Van der Vaart, 2015)—and hence this thesis will aim to combine these fields and solve interesting problems related to heavy-tailed analysis.

In the following sections we provide an introduction to the fields of extreme value theory and Bayesian nonparametrics, including some preparations for the following chapters.

1.2 Main Contributions of this Thesis

The main contributions of this thesis entail the development of novel concepts, methods, and inference approaches of Bayesian heavy-tailed modelling. Specifically, the main contributions are as follows.

Contribution 1: Heavy-Tailed Pitman–Yor Mixture Models

The big challenge of modelling heavy-tailed data using random probability measures is one of the motivations of this contribution. From a theoretical perspective, the characterization of the tails of a random probability measure provides a framework of appropriate random measures that can be used to model heavy-tailed data. For instance, it is well known that the tails of the Dirichlet process are exponentially thinner than the tails of the baseline, which leads to inappropriate models for heavy-tails. From an applied viewpoint, mixture models over a random probability measure and generalizations of them lead to extremely flexible models that can be adjusted to a variety of practical problems including heavy-tailed signals.

The contributions of this chapter are twofold. First, we characterize the tails of a particular instance of the Pitman–Yor process, known as the stable law process. In particular, we show that the tails of the stable law process are heavy-tailed, provided that the baseline is itself heavy-tailed, leading to appropriate models for heavy-tails and derive envelopes for the trajectories of the tail. Second, we develop two classes of heavy-tailed mixture models and rely on the characterization of the tails of the random measures to assessed their relative merits. We extend the methodology to the multivariate setup and provide a covariate-dependent version.

Contribution 2: Interfacing Nonstationary Multivariate Extremes and Heteroscedastic Extremes

Plenty of applications on multivariate extremes have a time-evolving nature and a lot of effort has been devoted to track how the dependence between the extreme values evolves over time—or according to another covariate—which is the main focus of nonstationary multivariate extremes. Yet, it is still a very challenging goal. This chapter propose a statistical framework that aims to link two mainstream fields of extreme value theory—*nonstationary multivariate extremes* and *heteroscedastic extremes*—and to deal with the nontrivial challenge of tracking the dependence over time using the *heteroscedastic extremes* regime.

We show how the proposed framework captures key aspects on the dynamics of extremal dependence over time. In particular, that the scedasis function of the joint tail can be regarded as a standardized time-varying version of the coefficient of extremal dependence. More general, using a class of aggregation rules for extreme values, we show that a constant scedasis corresponds to stationary multivariate extremes. Another contribution of this chapter is to propose Bayesian inference for the magnitude and frequency of extremes. In particular, for the structure scedasis functions, we propose inference via finite mixtures of Polya trees supported on the unit interval, which avoid boundary bias.

1.3 Outline of this Thesis

The remainder of the thesis unfolds as follows. The thesis is partitioned into four parts. Part I, consists of Chapters 1, 2, and 3 and contains background and lays the groundwork. Part II, consists of Chapters 4 and 5 and contains the proposed methodologies and main contributions of the thesis. Part III, which consists of Chapter 6 discusses the main results. Finally, Part IV contains two supporting chapters that complement the main contributions.

The details are as follows. In Chapter 1, we provide some motivation for modelling heavy-tailed data and introduce basic concepts and preparations on extreme value theory and Bayesian nonparametric theory. In Chapter 2, we revise literature related to Bayesian nonparametric priors with a particular emphasis on the prior processes that will play a key role in Part II, namely: the Pitman–Yor and the Polya-tree. Details on the construction of these process priors and posterior distributions are provided in this chapter. In Chapter 3, we revise literature on nonstationary extremes focusing on the study of dependence of extremes changing over time, providing information on measures of dependence for extremes, a brief introduction to copulas to study and model different dependence structures, and then an extension to time-varying version when we assume a nonstationary setup. In Chapter 4, we introduce a model that induces a prior on the space of multivariate distributions with heavy-tailed margins. We provide results on the tail behavior of some well known random process, such as the Stable process and the Dirichlet

process. The model is extended to the predictor-dependent version to include covariates on a heavy-tailed response. This is a flexible multivariate model able to estimate the body and the tail of a heavy-tailed distribution simultaneously. In Chapter 5, we introduce a Bayesian model for learning about the frequency and magnitude of extreme values in the joint tail as well as in other regions, whose shape will be defined via a class of natural rules for aggregating extreme values on a common scale. We demonstrate how the resulting structure scedasis function carries key information on the dynamics governing the degree of association of extremes of the basis process over time, and developed Bayesian inference methods for the two targets of interestthe structure scedasis function and the extreme value index. Finally, in Chapter 6 we present some closing remarks, limitations, and comment on open problems to be addressed in research.

Chapter 2

Nonparametric Bayesian Analysis

This chapter will briefly review theory and methods of nonparametric Bayesian inference, with a special emphasis on the priors that will play a key role in the main contributions of this dissertation.

2.1 Introduction

Bayesian nonparametric priors has been studied by several authors, the following books Ghosal and Van der Vaart (2015) and Müller et al. (2015b), provide full details about Bayesian nonparametric priors, theoretical properties and an extended view on applications. There are a lot of ways to construct a prior in the space of probability measures. For instance, construction through a stochastic process, construction through normalization of a random measure, construction through stick-breaking, and construction though a tree. The two priors of interest in this thesis will be constructed via stick-breaking (Pitman–Yor) and via a tree (Polya tree). Note that there are other options to construct the Pitman–Yor prior, but for now we will focus on the stick-breaking construction, Chapter 4 will provide a connection with the construction of a Pitman–Yor process via normalization of a random measure. The first prior introduced in this chapter is the Dirichlet process prior, and then we introduce extensions as they provide more flexibility to perform inference and the methodologies developed in this thesis will take advantage of their flexibility. Chapter 4 will study the tails of the Pitman–Yor process prior (PYP) and variations of it (covariance dependent version), and will use this result to developed mixture models. Chapter 5, will developed methodology using the mixture of Polya tree priors to induce a prior in the space of structure scedasis functions. The following sections provide a brief introduction to some priors on spaces of probability measures.

2.2 Bayesian Inference via Prior Processes

The main contributions of this thesis will involve methodologies from nonparametric Bayesian inference, and hence below I will introduce some preparations on the subject. Recall that, by definition a statistical model consists of a family of probability measures indexed by a parameter, $\mathcal{G} = \{G_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$. Statistical models that are described through a finite vector are referred as finite-dimensional models or parametric models. A Bayesian nonparametric (BNP) model is described through an infinite-dimensional parameter, and hence the parameter space is for example a function space. Using this approach allows more flexibility by relaxing the parametric assumptions. To perform Bayesian inference in a nonparametric model we need to define a prior on the infinite-dimensional parameter space, this parameter is usually a function or a measure (random probability measure, regression function, etc.). This section will be focus on the most commonly used prior process—the Dirichlet process—and review interesting properties as well as different representations.

2.2.1 Dirichlet Process

The Dirichlet process (DP) is considered a canonical example of a Bayesian nonparametric prior, mainly due to the mathematical tractability and its various representations. There are several methods to construct the Dirichlet process (e.g. Ferguson, 1973; Sethuraman, 1994), in this subsection we will review some of this constructions.

Definition 1. *(Dirichlet process) A random measure G on a Borel space $(S, \mathscr{B}_S)$ is a Dirichlet process with total mass parameter $M = G(S)$ and with base measure G_0, if*

$$(G(B_1), \ldots, G(B_k)) \sim Dir(MG_0(B_1), \ldots, MG_0(B_k)) \tag{2.1}$$

for every finite measurable partition $B_1, \ldots, B_k$ of S.

If the random measure G follows a Dirichlet process, we will write $G \sim \mathrm{DP}(M, G_0)$, where M is called the *precision* parameter and G_0 the *centering distribution*. Properties of the DP will be derived directly from the Dirichlet distribution (2.1); in the following proposition we state the mean and variance of the Dirichlet process.

Proposition 1. *Let $G \sim DP(M, G_0)$. Then, for any measurable sets A and B,*

1. $E\{G(A)\} = G_0(A)$.
2. $Var\{G(A)\} = \{G_0(A)(1 - G_0(A))\}/M + 1$.
3. $Cov\{G(A), G(B)\} = \{G_0(A \cap B) - G_0(A)G_0(B)\}/M + 1$.

Proposition 1 shows the effect of the *precision* parameter M in the variance of the DP. The proof of Proposition 1 can be found in Ghosal and Van der Vaart (p.60 2015). For instance,

for large M, the variance of the process G goes to zero, and G is concentrated in G_0. Whereas, for small M, the variance is very large and G is concentrated in the empirical distribution. Figure 2.1 illustrate the effect of the *precision* parameter M on realizations of a DP.

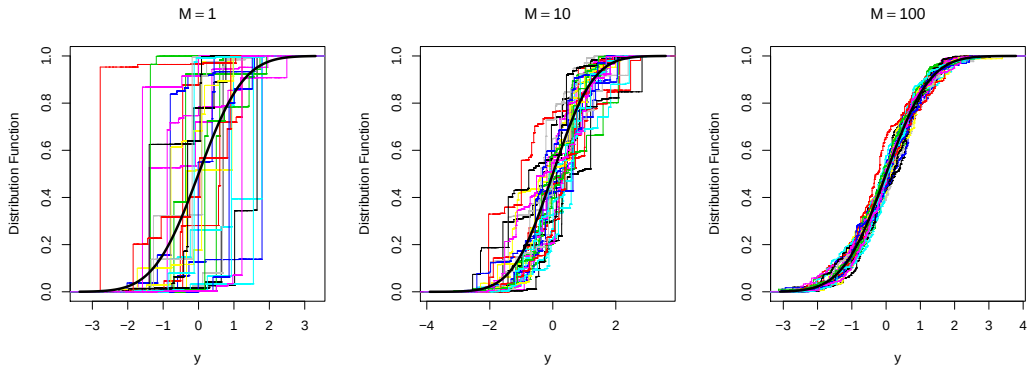


Figure 2.1: Plot of 20 sample trajectories from the $\text{DP}(M, G_0)$, with $G_0 = N(0, 1)$ a standard Normal distribution. For different values of M.

Stick-Breaking Representation

The Dirichlet process have another useful representation, motivated by its almost surely discrete nature, the so-called stick-breaking representation, given by Sethuraman (1994). A random probability measure G follows a Dirichlet process—here denoted as $G \sim \text{DP}(M, G_0)$—if G admits a stick-breaking representation of the type,

$$G = \sum_{h=1}^{\infty} \pi_h \delta_{Y_h}, \tag{2.2}$$

where

$$Y_h \overset{\text{iid}}{\sim} G_0, \qquad \pi_h = V_h \prod_{k<h} (1 - V_k), \qquad V_h \overset{\text{iid}}{\sim} \text{Beta}(1, M), \qquad h \in \mathbf{N}.$$

and δ_x the Dirac delta over x.

Posterior conjugacy

The Dirichlet process is conjugated with respect to a random sample of the probability measure G. This property means the posterior given a sample will also be a DP but with parameters updated.

Proposition 2. *Let* $Y_1, \ldots, Y_n \mid G \overset{iid}{\sim} G$ *with* $G \sim DP(M, G_0)$. *Then*

$$G \mid Y_1, \ldots, Y_n \sim DP(M + n, \frac{M}{M+n} G_0 + \frac{n}{M+n} G_n)$$

where $G_n = n^{-1} \sum_{i=1}^{n} \delta_{Y_i}$ *is the empirical distribution of* $Y_1, \ldots, Y_n$.

The proof of Proposition 2 involves Dirichlet-Multinomial conjugacy and further details on it can be found in Ghosal and Van der Vaart (p.62 2015). Combining Proposition 2 and Proposition 1, we can obtain the posterior mean and the posterior variance of the DP,

$$E\{G(A) \mid Y_1, \ldots, Y_n\} = \frac{M}{M+n} G_0(A) + \frac{n}{M+n} G_n(A), \tag{2.3}$$

$$\text{Var}\{G(A) \mid Y_1, \ldots, Y_n\} = \frac{G_n(A)\{1 - G_n(A)\}}{1 + M + n}. \tag{2.4}$$

As in can be seen from (2.3), the posterior mean is a convex combination of the prior mean G_0 and the empirical distribution, where for fixed M and large n the posterior mean has the same asymptotic properties as the empirical distribution G_n.

2.2.2 Dirichlet Process Mixtures

The Dirichlet process, as mentioned before is discrete with probability one. This section will focus on Dirichlet process mixtures, an approach introduced to overcome the lack of smoothness of the DP when estimating a density. The idea is to provide more smoothness by convolving the trajectories of the DP with some continuous kernel $\mathbb{K}$.
Consider $\mathbb{K}(y;\theta)$, the density of a kernel with $\theta \in \Theta \subseteq \mathbb{R}^q$, then the densities

$$f(y) = \int_\Theta \mathbb{K}(y;\theta) \mathrm{d}G(\theta) \tag{2.5}$$

are known as *Dirichlet process mixtures* (DPM), if $G \sim \text{DP}(M, G_0)$. The kernel choice will depend on the sample space. For example, in the unit interval we consider kernels such as Beta; on the positive line we consider kernels such as Gamma, Weibull or Log-Normal and for the real line we consider kernels such as Normal. Note that using the stick-breaking representation of the DP, we can write the DPM model (2.5) as,

$$f(y) = \int_\Theta \mathbb{K}(y;\theta) \mathrm{d}G(\theta) = \sum_{h=1}^{\infty} \pi_h \mathbb{K}(y;\theta_h). \tag{2.6}$$

The representation of the DPM given by (2.6) is very useful for posterior sampling. Walker (2007) and Kalli et al. (2011) developed an slice sampler for the posterior of the DPM. Below we describe briefly the algorithm; extend (2.6) by considering a latent variable u such that

$$f(y,u) = \sum_{h=1}^{\infty} 1(u < \pi_j) \mathbb{K}(y;\theta_h).$$

Then, given u we have a finite mixture,

$$f(y \mid u) = N_u^{-1} \sum_{j \in A_u} \mathbb{K}(y;\theta_h),$$

where $A_u = \{k : \pi_k > u\}$ and $N_u = \sum_{j=1}^{\infty} 1(\pi_h > u)$. Next, add another latent variable d, a membership variable with $\mathsf{p}(d_i = h \mid \text{else}) \propto 1(u_i < \pi_h)$. Then, the variables of the slice sampler can be updated as follows,

- $\mathsf{p}(\theta \mid \text{else}) \propto G_0(\theta_h) \prod_{d_i=h} \mathbb{K}(y_i; \theta_h)$.
- $\mathsf{p}(V_h \mid \text{else}) \propto \text{Beta}(V_h; 1 + \sum_{i=1}^{n} 1(d_i = h), M + \sum_{i=1}^{n} 1(d_i > h))$.
- $\mathsf{p}(u_i \mid \text{else}) \propto 1(0 < u_i < \pi_{d_i})$.
- $\mathsf{p}(d_i = k \mid \text{else}) \propto 1(k : \pi_k > u_i)\mathbb{K}(y_i; \theta_k)$.

There are other algorithms to sample from the Dirichlet process mixtures, using different approaches such as the blocked-Gibbs sampling algorithm MacEachern (1994), Neal's algorithm and Reversible Jump Markov Chain Monte Carlo Jain and Neal (2004). We focus this section on the algorithm proposed by Kalli et al. (2011), as the Gibbs sampler presented on Chapter 4 will be an extension of this sampling method.

Beyond the Dirichlet process, there are other options of Bayesian nonparametric priors; Next, we introduce two extensions of the Dirichlet process prior that can be used for density estimation problems and they will be key for the methodology developed in this thesis.

2.3 Construction of a Prior Process

This section, introduced different constructions of a prior distribution on the space of probability measures. The Dirichlet process introduced in Subsection 2.2.1 allows different representations, depending on the construction. Similarly, other prior processes will allow different constructions. In this section, we provide more information about the different constructions that will be used to define the Pitman–Yor process and the Polya tree process. Computational methods for both prior processes are available in the Supporting Material A and B, respectively.

Construction through Stick-Breaking

The stick-breaking construction consists in breaking a stick of length one, at random to each element of $\mathbb{N}$. Consider a stick on length one and random variables on the unit interval $(V_i)_{i \leqslant 1}$, the stick-breaking construction consists on cutting the stick according to $(V_i)_{i \geqslant 1}$. First, cut the stick into two parts, one of length $\pi_1 = V_1$ and the reminder of length $(1 - \pi_1) = (1 - V_1)$. Second, cut the leftover stick of length $(1 - \pi_1)$ into two parts, one proportional to V_2, obtaining a stick of length $\pi_2 = V_2(1 - V_1)$ and the reminder of length $(1 - V_2)(1 - V_1)$. Continue this process, so at step j, we have j sticks of length $\pi_1, \ldots, \pi_j$, where $\pi_j = V_j \prod_{i=1}^{j}(1 - V_j)$ and a reminder stick of length $1 - \sum_{j \geqslant 1} \pi_j$. We can see that by construction, $\sum_{j \geqslant 1} \pi_j \leqslant 1$, with $0 \leqslant \pi_j \leqslant 1$. Figure 2.2 illustrates the stick-breaking construction.

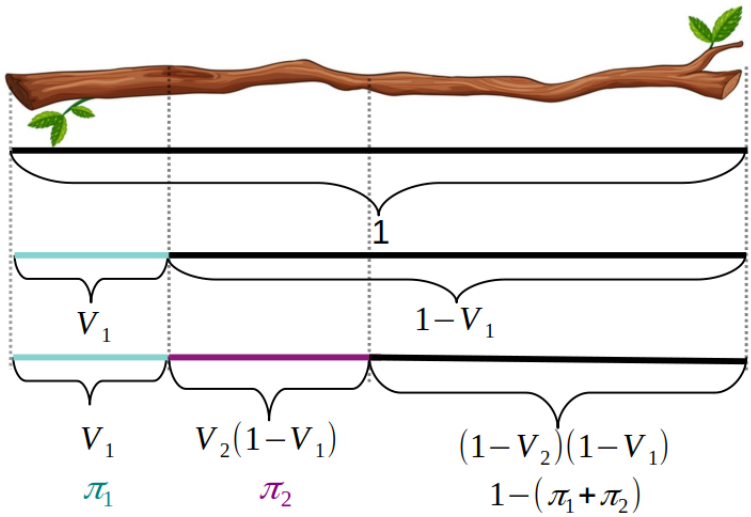


Figure 2.2: Stick-breaking construction, first steps.

Definition 2 (Stick-breaking process)**.** *A random probability measure μ is called a stick-breaking process if it has atomic decomposition of the form*

$$\mu = \sum_{h=1}^{\infty} \pi_h \delta_{Y_h} \quad \pi_h = V_h \prod_{k<h} (1 - V_k),$$

where $(Y_h)_{h\geqslant 1}$ are a collection of atoms independent of the weights $(\pi_h)_{h\geqslant 1}$, for some random variables $(V_h)_{h\geqslant 1}$ on the unit interval.

The stick-breaking process μ will be proper if and only if $\lim_{j\to\infty} E\{\prod_{i=1}^{j}(1 - V_i)\} = 0$, if μ is proper, this implies that $\sum_{j\geqslant 1} \pi_j = 1$. Note that the Dirichlet process is a particular case of stick-breaking process with $V_h \sim \text{Beta}(1, M)$, and satisfying the condition $\lim_{j\to\infty} E\{\prod_{i=1}^{j}(1 - V_i)\} = 0$.

Construction through a Tree

For this construction, we will consider a sequence $\Pi^0 = \{\mathfrak{X}\}$, $\Pi^1 = \{B_{1,1}, B_{1,2}\}$, $\Pi^2 = \{B_{2,1}, B_{2,2}, B_{2,3}, B_{2,4}\}$, and so on, of partitions of the sample space $\mathfrak{X}$. Let G be a probability measure, satisfying the tree additivity property, i.e, $G(B_{j-1,k}) = G(B_{j,k+1}) + G(B_{j,k})$ for $k = 1, \ldots, 2^j$ and consider the conditional probabilities for $l = 1, \ldots, 2^j$ and all j,

$$p_{j,2l} = G(B_{j,2l} \mid B_{j-1,l}) \quad p_{j,2l-1} = G(B_{j,2l-1} \mid B_{j-1,l}). \tag{2.7}$$

Figure 2.3 illustrates the tree construction.

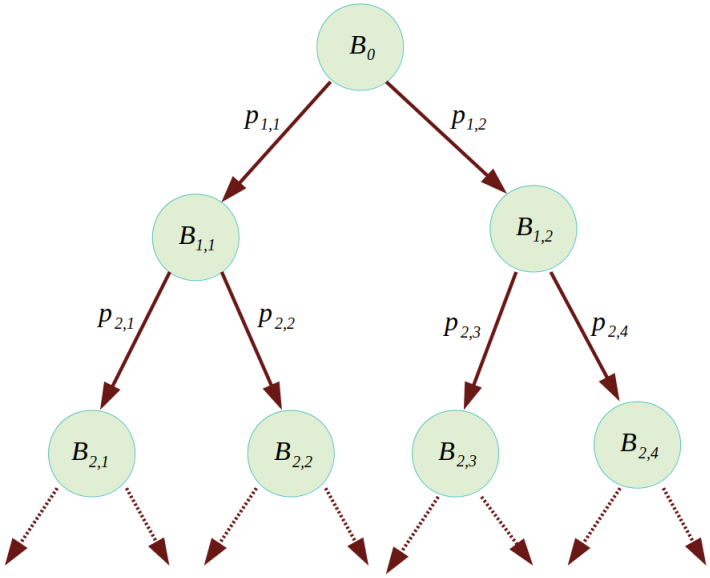


Figure 2.3: Tree diagram construction, first steps.

Then, if we define the probability of a set $B_{J,l}$ as,

$$G(B_{J,l}) = \prod_{j=1}^{J} p_{j,\lceil l2^{j-J} \rceil},$$

and $p_{j,k+1} + p_{j,k} = 1$, G will define a finitely additive measure (see Ghosal and Van der Vaart, 2015, Section 3).

Definition 3 (Tail-free process)**.** *The random measure G, satisfying the tree additivity property and with conditional probabilities defined as (2.7) is a tail-free process with respect to the partition $\Pi = \{\Pi^j : j = 1, \cdots\}$ if $\{p_0\}, \{p_{1,1}, p_{1,2}\}, \ldots,$ are independent.*

Note that the Dirichlet processes has a representation as a particular case of a tail-free process, if in Definition 3 we define $\{p_0\}, \{p_{1,1}, p_{1,2}\} \ldots$ to be Beta distributed with parameters adding up to the upper branch of the tree; that is, if we add the parameters of $p_{1,1}$ we recover the first parameter of p_0, and so on.

Construction through Normalization of a Random Measure with Independent Increments

For this construction, we first characterized the class of random measures with independent increments and then define the normalized prior process.

Definition 4. *[Complete random measure CRM] A measure μ on a Borel space $(S, \mathscr{B}_S)$ is a complete random measure if for every disjoint set $B_1, \ldots, B_n \in \mathscr{B}_S$, the random variables $(\mu(B_1), \ldots, \mu(B_n))$ are independent.*

In simple words, it is a random measure, which produce independent random variables when evaluated over disjoint sets. Inhere, whenever we work with a completely random measure we will assume that it has no fixed jumps and that the Laplace transform of μ admits the following representation:

$$E\left[e^{-\lambda\mu(B)}\right] = \exp\left[-\int_{\mathbb{R}^+\times B}(1-e^{-\lambda v})\nu(dv,dy)\right]$$

where $\lambda > 0$ and ν is a measure referred to as the Levy measure of μ. The Lévy measure is a measure satisfying $\int_{\mathbb{R}^+\times B}\min\{v,1\}\nu(dv,dy) < \infty$, and can be decomposed as $\nu(dv,dy) = \rho(dv \mid y)\alpha(dy)$, where ρ will control the jump intensities and α the location of the jumps. Now, our main topic of interest is to define random probability measures, and we will be using completely random measures to build them through normalization.

Definition 5. *[Normalized random measure with independent increments NRMI] Let μ be a CRM, with Lévy measure $\nu(dv,dy) = \rho(dv \mid y)\alpha(dy)$, such that $0 < \mu(S) < \infty$. Then G is a NRMI defined as,*

$$G = \frac{\mu}{\mu(S)}$$

The condition $0 < \mu(S) < \infty$, provides a sufficient condition for the normalization to be well defined, which is satisfied when the Lévy measure is such that α is finite and $\int_{\mathbb{R}^+}\rho(dv \mid y) = \infty$. We called G a NRMI prior process, and it is fully characterized by their Lévy measure $\nu(dv,dy)$.

The Dirichlet process can be defined as a NRMI, when we take the CRM μ to be the gamma process, when $\nu(dv,dy) = e^{-v}v^{-1}G(dy)$, with G_0 the centering distribution, and M analogous to Definition 1 of the Dirichlet process.

2.4 Pitman–Yor Process

The Pitman–Yor process (PYP) stems from the two parameter-parameter Poisson–Dirichlet (PD) distribution of Pitman and Yor (1997), and it is a key instance of a stick-breaking prior. The two parameter PD distribution is a family of distributions designed to generate stick-breaking-type weights; note, however, that Pitman and Yor do not focus on atoms, nor comment on the random probability measures that can be devised by PD distributions. Ishwaran and James (2001) later took advantage of the rich class of weights stemming from PD distributions in order to construct random probability measures that extend the Dirichlet process are now known as Pitman–Yor Processes.

Stick-Breaking Representation

A random probability measure G follows a Pitman–Yor process, here denoted as $G \sim \text{PYP}(D, M, G_0)$, if G admits a stick-breaking representation of the type,

$$G = \sum_{h=1}^{\infty} \pi_h \delta_{Y_h}, \tag{2.8}$$

where

$$Y_h \overset{\text{iid}}{\sim} G_0, \qquad \pi_h = V_h \prod_{k<h} (1 - V_k), \qquad V_h \overset{\text{ind}}{\sim} \text{Beta}(1 - D, M + hD), \qquad h \in \mathbf{N}.$$

The parameters of the Pitman–Yor and are known as *discount* (D), *precision* (M), and *centering distribution* (G_0) and are subject to the constraints $0 \leqslant D < 1$ and $M > -D$. The Dirichlet process is a particular case, with $D = 0$ and the case where $M = 0$ is known as the class of stable law processes. In terms of the two parameter-parameter Poisson–Dirichlet (PD) terminology of Pitman and Yor, there is the following correspondence:

- Dirichlet process: $\text{PD}(0, M)$.
- Stable law process: $\text{PD}(D, 0)$.

The *stick-breaking representation* is handy for a variety of purposes, however it can also be derived via normalization of a D-stable subordinator or a generalized gamma process. The next proposition, states the mean and variance of the Pitman–Yor process.

Proposition 3. *Let $G \sim PYP(D, M, G_0)$. Then, for any measurable sets A and B,*

1. $E\{G(A)\} = G_0(A)$.
2. $E\{G(A)G(B)\} = \frac{1-D}{M+1} G_0(A \cap B) + \frac{M+D}{M+1} G_0(A)G_0(B)$.
3. $Cov\{G(A), G(B)\} = \frac{1-D}{M+1}\{G_0(A \cap B) - G_0(A)G_0(B)\}$.

Proposition 3 provides insight on the effects of the parameters M and D on the PYP. For instance, for fixed M, when D approaches 1, we get that the variance goes to zero, and G is concentrated in G_0. Alternatively, if M approaches zero, for fixed D, the variance is large and the G_0 approaches the empirical distribution. Figure 2.4 illustrates the effect of the parameters M and D on realizations of a PYP.

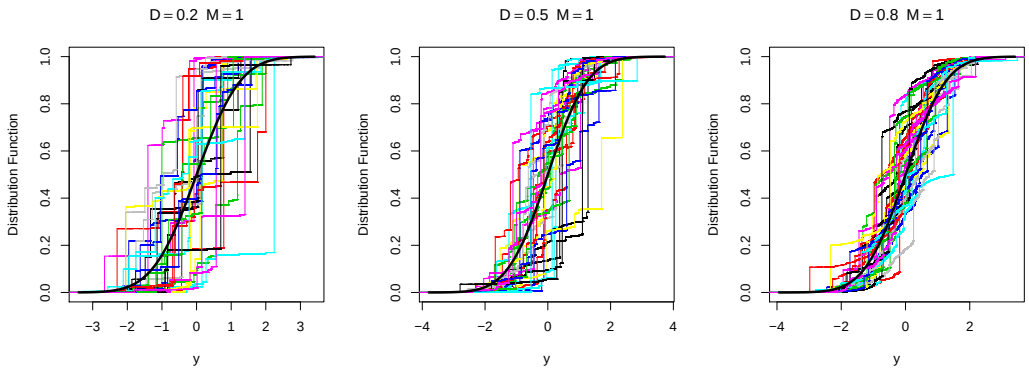


Figure 2.4: Plot of 20 sample trajectories from the PYP(D, M, G_0), with $G_0 = N(0, 1)$ a standard Normal distribution. For different values of D and fixed $M = 1$.

Posterior conjugacy

The Pitman–Yor process is not conjugate in general, only on the case when we recover the Dirichlet process, but we still can characterize the posterior in the following proposition.

Proposition 4. *Let* $Y_1, \ldots, Y_n \mid G \overset{iid}{\sim} G$ *with* $G \sim PYP(D, M, G_0)$ *for* $D \geqslant 0$. *Then the posterior* $G \mid Y_1, \ldots, Y_n$ *is the distribution of the random measure,*

$$R_n \sum_{j=1}^{K_n} \hat{V}_j \delta_{\hat{Y}_j} + (1 - R_n) Q_n$$

where $R_n \sim Beta(n - K_n D, M + K_n D)$, $(\hat{V}_1, \ldots, \hat{V}_{K_n}) \sim Dir(K_n; N_{1,n} - D, \ldots, N_{K_n,n} - D)$ *and* $Q_n \sim PYP(D, M + DK_n, G_0)$ *independently. And* $N_{1,n}, \ldots, N_{K_n,n}$ *are the multiplicities of the* K_n *distinct values* $(\hat{Y}_1, \ldots, \hat{Y}_{K_n})$ *of the sample.*

The proof of Proposition 3 and Proposition 4 can be found on Ghosal and Van der Vaart (Chapter 19 2015). The posterior distribution of the Pitman–Yor process will be consistent in weak topology for two specific cases: when the true distribution is discrete and when the centering distribution G_0 is proportional to the continuous part of the true distribution. In the special case of recovering the Dirichlet process, it means a Pitman–Yor process with parameters PYP$(D = 0, M, G_0)$ then the posterior distribution will always be consistent; see Ghosal and Van der Vaart (Theorem 14.38 2015). In summary, the Pitman–Yor process is only consistent under fairly unrealistic situations—except for the case of the Dirichlet process which turns out to be consistent under fairly mild assumptions. Another type of important results concerning the posterior distribution of a nonparametric prior are the class of Bernstein–von Mises theorems that provides asymptotic approximations for the posterior distributions to a Gaussian, and are important for the construction and interpretation of Bayesian credible regions. Ray and van der

Vaart (2021) prove a kind of Bernstein–von Mises theorem for the Dirichlet process posterior, specifically show that the Laplace transform of the posterior Dirichlet process centered at the empirical measure converges to the Laplace transform of a centered Normal distribution. Ghosal and Van der Vaart (Lemma 14.39 2015), show a Bernstein–von Mises theorem for the Pitman–Yor process posterior, in the particular case of an atomless distribution. Then, Franssen and van der Vaart (2021) provide a Bernstein–von Mises theorem when the true distribution of the data is discrete and with atoms that decrease not too slowly.

2.5 Polya Tree Process

The Polya tree prior, is a random measure construct via a tree, which is a special case of a tail-free random measure. This class of priors are flexible and contain the Dirichlet process as a special case. Polya tree constructions were first considered by Ferguson (1974); Lavine et al. (1992, 1994). The idea is similar to defining an histogram, but adding a random component, consider the bins of the histogram as $\{B_{j,l} : l = 1, \ldots, 2^j\}$ for $j = 1, 2, \ldots$, but now the probability on each bin will be random $G(B_{j,l})$ according to a random probability measure G, and defining the recursive random probabilities $p_{j,2l} = G(B_{j,2l} \mid B_{j-1,l})$ and $p_{j,2l-1} = G(B_{j,2l-1} \mid B_{j-1,l})$. This is made precise in the following definition from Müller et al. (2015a).

Definition 6. *(Polya tree prior) Let $\Pi = \{\Pi^j\}$ be a sequence of nested partitions of the form $\Pi^j = \{B_{j,l} : l = 1, \ldots, 2^j\}$, and $\mathcal{A} = \{\alpha_{j,l} : l = 1, \ldots, 2^j\}$ a collection of nonnegative numbers. Then, a random probability measure G in $\mathfrak{X}$ is a Polya tree prior with parameters $(\Pi, \mathcal{A})$ if for every $l = 1, \ldots, 2^j$ and*

$$G(B_{J,l}) = \prod_{j=1}^{J} p_{j,\lceil l2^{j-J} \rceil}, \quad j = 1, \ldots,$$

where the conditional probabilities $p_{j,l}$ are mutually independent random variables

$$p_{j,l} \sim Beta(\alpha_{j,2l-1}, \alpha_{j,2l})$$

and $p_{j,2l} = 1 - p_{j,2l-1}$. We write $G \sim PT(\Pi, \mathcal{A})$.

Note that from the definition, we can see that Polya trees are a special case of tail-free processes, when we consider the conditional distributions to be Beta distributed. The Dirichlet Process (DP) can be obtained as a specific case of a Polya tree prior when we add the condition $\alpha_{j,2l} + \alpha_{j,2l-1} = \alpha_{j,l}$, this special case has the advantage that the choice of $\alpha_{j,l}$ does not affect the inference. However, the Polya tree prior will allow for choices of $\alpha_{j,l}$ such that G is absolutely continuous with probability one, which we could never achieve with the DP (Ferguson, 1974).

Posterior conjugacy

The Polya tree prior is conjugated under a random sample of the probability measure G, it means, given a sample, the posterior will also be a Polya tree process but with parameters updated by the number of observations y_i in each interval $B_{i,j}$ of the partition.

Proposition 5. *Let* $Y_1, \ldots, Y_n \mid G \overset{iid}{\sim} G$ *with* $G \sim PT(\Pi, \mathcal{A})$. *Then*

$$G \mid Y_1, \ldots, Y_n \sim PT(\Pi, \mathcal{A}^*),$$

where $\alpha^*_{j,l} \in \mathcal{A}^*$ *are* $\alpha^*_{j,l} = \alpha_{j,l} + k_{j,l}$, *and* $k_{j,l} = \sum_{i=1}^{n} 1_{Y_i \in B_{i,j}}$.

The proof of Proposition 5 can be found in Lavine et al. (1992, 1994), with more information regarding the marginal distribution of the random sample. Specific choices of partitions Π, will allow the random measure G to be center on a parametric distribution G_0, that the $E(G) = G_0$ (Lavine et al., 1992).

Prior centering on a parametric distribution

Let G_0 be a continuous distribution function, and G_0^{-1} the inverse distribution function, define the partition Π in terms of the quantiles of G_0,

$$B_{j,l} = [G_0^{-1}\{(l-1)/2^j\}, G_0^{-1}\{l/2^j\}),$$

where $j = 0, \ldots, 2^j$, and take $\alpha_{j,2l} = \alpha_{j,2l-1}$. Then,

$$E\{G(B_{j,l})\} = \prod_{j=1}^{J} E\{p_{j,l}\} = \frac{1}{2^J} = G_0(B_{j,l}).$$

We write $G \sim PT(G_0, \mathcal{A})$. We can additionally consider a prior on the parameters of G_0 so that the partition Π is updated at each step; this is known as a Mixture of Polya tree priors and has the advantage of changing the jumps, leading to smooth densities (Hanson and Johnson, 2002). Conditions for posterior consistency in density estimation of Polya trees in weak topology can be found in Lavine et al. (1994) and Ghosal et al. (1999), and results on posterior consistency under Hellinger topology can be found in Barron et al. (1999). A nonparametric Bernstein–von Mises theorem for the Polya tree process was given by Castillo (2017) along with the study of contraction rates.

2.6 Miscellanea

Until now we have focus on the priors of most interest for the contribution chapters, the Dirichlet, Pitman–Yor and Polya tree, however there are many other prior processes that have

interesting properties. In this section, we will review other prior processes that are based on the constructions from Section 2.3.

Extensions using Stick-Breaking Representation

There are many stick-breaking approaches for constructing prior processes, we provide below some other examples beyond the DP and the Pitman–Yor.

Example 1 (Geometric process)**.** *A random probability measure G follows a geometric process, denoted as $G \sim G(\nu_0, G_0)$ if G admits a stick-breaking representation of the type,*

$$G = \sum_{h=1}^{\infty} \pi_h \delta_{Y_h}, \tag{2.9}$$

where

$$Y_h \overset{\text{iid}}{\sim} G_0, \qquad \pi_h = V_h \prod_{k<h}(1 - V_k), \qquad V_h = \lambda(1-\lambda)^{i-1}, \quad \lambda \sim \nu_0, \qquad h \in \mathbf{N}.$$

Here, G_0 is the center distribution and ν_0 is a random variable, usually defined as a Beta distribution. The popularity of this model arises from the decreasing structure of the weights and their simple stick-breaking representation. The geometric process was introduced by Fuentes-Garcia et al. (2010), and Mena et al. (2011) proposed posterior algorithms for this prior process.

Example 2 (Probit stick-breaking process)**.** *A random probability measure G follows a probit stick-breaking process, denoted as $G \sim PSBP(\mu, \sigma, G_0)$ if G admits a stick-breaking representation of the type,*

$$G = \sum_{h=1}^{\infty} \pi_h \delta_{Y_h}, \tag{2.10}$$

where

$$Y_h \overset{\text{iid}}{\sim} G_0, \qquad \pi_h = V_h \prod_{k<h}(1 - V_k), \qquad V_h = \Phi(\alpha_h), \quad \alpha_h \sim N(\mu, \sigma), \qquad h \in \mathbf{N}.$$

Here, Φ is the standard normal cumulative distribution, G_0 is the center distribution and μ and σ will control the variability of the trajectories around G_0. For further details on this prior process, see Rodríguez and Dunson (2011).

Extensions using NRMI

As mentioned in Section 2.3, different CRM will lead to different NRMI priors, below we present one of the most interesting priors the so-called normalized generalized Gamma processes that are of great interest due to their generality.

Example 3 (Normalized generalized Gamma)**.** *A random probability measure G in $(S, \mathscr{B}_S)$ follows a normalized generalized Gamma process, denoted as $G \sim NGG(M, k, \gamma, G_0)$ if G is defined as,*

$$G = \frac{\mu}{\mu(S)},$$

where μ is a CRM, with Lévy measure

$$\nu(dv, dy) = \frac{e^{-kv}}{\Gamma(1-\gamma)v^{1+\gamma}} dv G(dx).$$

Here, G_0 is the centering distribution and M is the total mass, with $k \leqslant 0$ and $\gamma \in [0, 1)$ the parameters controlling the variability of the trajectories around G_0.

Note that the class of normalized generalized gamma priors contains a lot of interesting priors as special cases:

- If $G \sim \text{NGG}(M, 1, 0, G_0)$ we recover the Dirichlet process.
- If $G \sim \text{NGG}(1, k, 1/2, G_0)$ we recover the normalized inverse Gaussian (N-IG) process.
- If $G \sim \text{NGG}(1, 0, \gamma, G_0)$ we recover the stable process process.

See Lijoi et al. (2007) for more details on NGG processes and James et al. (2009) for posterior inference algorithms. An interesting application of this general class of processes, is the construction of mixture models. Barrios et al. (2013), considered infinite mixture models, based on NGG measures. Similar to Section 2.5, consider $\mathbb{K}(y; \theta)$, the density of a kernel with $\theta \in \Theta \subseteq \mathbb{R}^q$, and let

$$f(y) = \int_{\Theta} \mathbb{K}(y; \theta) \mathrm{d}G(\theta) \tag{2.11}$$

where now the mixing distribution is $G \sim \text{NGG}(M, k, \gamma, G_0)$, these type of mixtures are extremely flexible and computational methods are available, Barrios et al. (2013) provided details on posterior inference for these mixture models.

Chapter 3

Nonstationary Heavy-Tailed Models

In this chapter, we overview some key methods and results for modelling nonstationary extremes that have some links with the methodology developed in the main contributions of the thesis.

3.1 Introduction

The statistical models under the assumption of stationary data can be strong and restrictive for some applications. For instance, in modelling financial data where it is common to study the time-varying process over time and the assumption of stationary observations can be restrictive, in fact it has been documented in the literature (Poon et al., 2003, 2004; Castro, 2015; Einmahl et al., 2016) that the price returns of stock markets are heavy-tailed and nonstationary. For this reason, there is need for modelling nonstationary extremes. Standard practice for modelling univariate nonstationary data is to express the parameters of an extreme value model as a function of covariates. For instance, Davison and Smith (1990) and Coles (2001) considered a linear trend in the parameters of generalized Pareto (GP) distributions. Different approaches to model univariate nonstationary extremes following the ideas on Davison and Smith (1990) have been studied by Chavez-Demoulin and Davison (2005), Eastoe and Tawn (2009), Chavez-Demoulin et al. (2016), introducing conditional models for the GEV (Generalized Extreme Value) distribution. Heteroscedasic extremes is a recent methodology proposed by Einmahl et al. (2016) that offers another paradigm for modelling nonstationary extremes where the nonstationarity arise from changes in the frequency of extremes. For modelling multivariate nonstationary extremes less is known, but several models have been proposed (Eastoe and Tawn, 2009; Heffernan and Tawn, 2004; de Carvalho, 2016b; Marcon et al., 2016; Castro and de

Carvalho, 2017; Castro et al., 2018) where the main object of study is the conditional angular density, object we will introduce in the following sections.

One of the main focuses of nonstationary extremes is to model the dependence structure of extreme values when it is changing over time. Copulas are powerful tools for modeling dependence, and there is a strong connection between copulas and extreme value theory, Gudendorf and Segers (2010) contains a survey on modelling dependence via extreme-value copulas. Modelling the dependence changing over time can be achieved via time-varying copulas, these type of copulas were developed by Patton (2006), who extended the class defined by Sklar (1959) for conditional distributions. More contributions for modelling time-varying dependence via copulas are Van den Goorbergh et al. (2005); Jondeau and Rockinger (2006); Aloui et al. (2013).

The main contributions of Chapter 4 and Chapter 5 will have some links with the concepts and methodologies introduced in this chapter. Particularly Chapter 4 will use copulas to construct a multivariate heavy-tailed distribution, with heavy tailed marginals, and Chapter 5 will interface *nonstationary multivariate extremes* and *heteroscedastic extremes* to model frequency and magnitude of extremes on different regions. The remainder of this chapter is organized as follows. Section 3.2 review methodology on extreme value theory. Section 3.3 examine univariate methodologies for modelling nonstationary extremes. Section 3.4 and Section 3.5, provide some preparations for the extension to the multivariate case; introducing the time-varying copula and the main objects for studying dependence of multivariate extremes. Finally, Section 3.6 connects these concepts and review methodology for modelling multivariate nonstationary extremes.

3.2 Extreme Value Theory

3.2.1 Univariate Theory of Extreme Values

In this section, we present a concise introduction to extreme value theory. Extreme value theory focuses on the study of extreme events on the tail of a distribution, which can be categorized into three families depending on the rate of decay of the tail: light-tailed distributions, short-tailed distributions, and heavy-tailed distributions. In particular, the study of heavy-tailed distributions is of great interest and regular variation is an important tool to describe them. There are different ways to define a heavy-tailed distribution, the following definition is based on regular variation, later in this chapter we will provide a connection with equivalent definitions of a heavy-tailed distribution. However, the following definition is what we refer to as a heavy-tailed distribution throughout the thesis.

Definition 7 (Regular variation)**.** *A distribution F is said to have a regularly varying tail with*

index α, if for some $\alpha > 0$,

$$1 - F(y) \sim y^{-\alpha}\mathcal{L}(y), \quad as \quad y \to \infty,$$

where $\mathcal{L}$ is a slowly varying function at infinity; i.e. $\mathcal{L}(zy)/\mathcal{L}(y) \to 1$, as $y \to \infty$ for all $z > 0$.

Definition 7 is equivalent to say that F is a heavy-tailed distribution, where $\xi = 1/\alpha$ is called the extreme value index. Note that with increasing values of ξ, the tail becomes heavier.

Extreme value theory is not only about the study of heavy-tailed distributions in the sense of Definition 7, but they are an important framework when studying extreme value theory, the methodology provided below will be made general for any rate of decay of the tail. We present a concise introduction of the main methods for modelling extreme values: block maxima methods and threshold methods. The first one, considers asymptotics of the block maxima (minima) and the second one considers observations exceeding certain threshold. The results presented in this section are well known in the literature and can be found, for instance, in Coles (2001) and Beirlant et al. (2004).

Block Maxima Methods

For the first method, let us focus on the behavior of $M_n = \max(Y_1, \ldots, Y_n)$, where $Y_1, \ldots, Y_n$ are independent and identically distributed random variables, with distribution function F. It is of probabilistic interest to study the possible limit distributions of the maxima M_n. The distribution of M_n can be easily obtained for all values of n, simply using the fact of independence between the variables:

$$P(M_n \leqslant z) = P(Y_1 \leqslant z) \times \cdots \times P(Y_n \leqslant z) = \{F(z)\}^n$$

Under this expression, the problem simplifies to studying the behavior of $\{F(z)\}^n$ in the limit, by definition, for any $z < z^*$, where z^* is the end-point of F, the limit of $\{F(z)\}^n$ will be zero. Hence, to get a non-degenerate distribution, considering a standardization seems natural:

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

The following theorem, attributed to Fisher and Tippett (1928) and Gnedenko (1943), is a key result on extreme value theory as it derives the general form of all possible limit laws for M_n^*.

Theorem 1. *(Extremal limit theorem) If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that,*

$$P\{(M_n - b_n)/a_n \leqslant z\} \to G(z), \quad as \quad n \to \infty,$$

for G a non-degenerate distribution function, then G is of the form:

$$G(z) = \exp\left[-\left\{1+\xi\left(\frac{z-\mu}{\sigma}\right)\right\}^{-1/\xi}\right], \tag{3.1}$$

for $\{z : 1+\xi(z-\mu)/\sigma > 0\}$, $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$.

Proof. See de Haan and Ferreira (2006, Section 1.1.3). □

Theorem 1, states that the limit distribution of M_n^* belong to the generalized extreme value (GEV) family, distribution obtained in (3.1), with three parameters of interest, μ the location parameter; σ the scale parameter and ξ the shape parameter, denoted by GEV(μ, σ, ξ). The shape parameter ξ in the GEV distribution (3.1) is of great importance in extreme value theory, it controls the behavior of the tail and divides the space into three classes of distributions. For $\xi > 0$ the class of heavy-tailed distributions, refer to as Fréchet; for $\xi = 0$ the class of light-tailed distributions, refer to as Gumbel and $\xi < 0$ the class of short-tailed distributions, refer to as Weibull. This class of distributions is called the domain of attraction of G and will be denoted by $D(G)$ and ξ is called the extreme value index. Note that for $\xi > 0$, the class of heavy-tailed distributions will satisfy Definition 7. Figure 3.1 shows the density from a GEV under each domain of attraction. Example of distributions on each domain attraction are:

- Fréchet domain of attraction ($\xi > 0$): Pareto, Generalized Pareto, Inverse Gamma, Log-Gamma, Fréchet.
- Gumbel domain of attraction ($\xi = 0$): Normal, Gamma, Exponential, Logistic, Log-Normal.
- Weibull domain of attraction ($\xi < 0$): Uniform, Beta, Reverse Burr.

Application of Theorem 1 requires splitting the data into blocks of equal length, and then fitting the GEV to the set of block maxima. We have to consider two main things, the estimation of the parameters of the GEV and the choice of the block size. Estimation of the parameters in the GEV is usually made by maximum likelihood techniques, in here we focus on Bayesian inference along the lines of Beirlant et al. (2004, Chapter 11). The choice of the size of the blocks is a trade-off between bias and variance (not too large or too small). For very small blocks we have an increase on the bias of the estimation, while for very large blocks we increase the variance of the estimation. There can be cases, where it is appropriate to model block minima, then we can easily adapt the GEV in Theorem 1 to the block $m_n = \min(Y_1, \ldots, Y_n)$ (Theorem 3.3 Coles and Tawn, 1991), this implies fitting the GEV distribution for the block maxima $(-Y_1, \ldots, -Y_n)$.

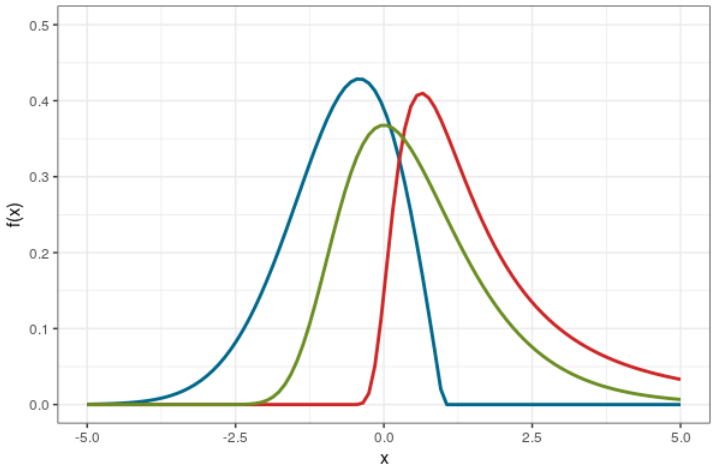


Figure 3.1: Density of three instances of generalized extreme value distributions with common scale parameter $\sigma = 1$, location parameter $\mu = -1, 0, 1$, and shape parameter $\xi = -0.5, 0, 0.5$ for the domain of attraction respectively. Weibull (green), Gumbel (blue) and Fréchet (red).

Bayesian Inference for the GEV

Bayesian inference can be conducted as follows. Let $Y_1, \ldots, Y_m$ be a random sample from a GEV(μ, σ, ξ) distribution. The likelihood is based on:

$$Y_i \mid \sigma, \mu, \xi \sim \text{GEV}(\mu, \sigma, \xi), \quad i = 1, \ldots, m.$$

let $\boldsymbol{\theta} = (\mu, \sigma, \xi)$ be the parameter of interest, then

$$f(y \mid \boldsymbol{\theta}) = \frac{1}{\sigma}\left\{1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right\}^{-1/\xi-1} \exp\left[-\left\{1 + \xi\left(\frac{y-\mu}{\sigma}\right)^{-1/\xi}\right\}\right].$$

For the prior distribution, consider the maximal average data information (MDI) prior (Zellner, 1996):

$$\pi(\boldsymbol{\theta}) = \exp\left[E\{\log f(Y \mid \theta)\}\right] \propto \frac{1}{\sigma} e^{-\psi(1)(1+\xi)},$$

where $\psi(1)$ is the Euler's constant. The Jeffreys prior on the GEV only exists when $\xi > -0.5$ and it is complicated to obtain in other cases. Under the above prior a Metropolis–Hastings algorithm is performed with proposal density equal to a multivariate Normal on $(\mu, \log\sigma, \xi)$ to obtain the posterior of the parameters.

Threshold Method

Until now, we discussed approaches considering the block maxima method, where the intention is to model the asymptotic behavior of the maximum or minimum value in the data. Now, we focus on the second method to model extremes, based on the behavior of observations above certain threshold.

Let $X_1, \ldots, X_n$ be a random sample, with marginal distribution F. For arbitrary X in the sequence we can focus on the conditional probability, exceeding certain threshold u:

$$P(X > u + y \mid X > u) = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0.$$

Approximations to this quantity are required, the main result is given in the following theorem.

Theorem 2. *Define $M_n = \max(X_1, \ldots, X_n)$, where $X_1, \ldots, X_n$ are independent and identical distributed random variables, with distribution equal to F. For arbitrary X in the sequence, suppose that Theorem 1 holds, it means,*

$$P(M_n^* \leqslant z) \to G(z) \quad as \quad n \to \infty$$

where G is a non-degenerate distribution function of the form:

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

for some $-\infty < \mu < \infty$, $-\infty < \xi < \infty$ and $\sigma > 0$. Then, for large u the conditional probability $P(X - u \leqslant y \mid X > u) \approx H(y)$ with,

$$H(y) = 1 - \left(1 + \frac{\xi y}{\hat{\sigma}}\right)^{-1/\xi} \tag{3.2}$$

for $\hat{\sigma} = \sigma + \xi(u - \mu)$ defined on $\{y : y > 0$ and $(1 + \xi y/\hat{\sigma}) > 0\}$. The exponential distribution $H(y) = 1 - \exp\{-y/\hat{\sigma}\}$ is used for the $\xi = 0$ case.

Proof. Result from Pickands (1975). □

Theorem 2, states that if the limit distribution of M_n^* belong to the GEV family, then the threshold excesses belong to the generalized Pareto family (GP), distribution obtained in (3.2) and denoted by $\text{GP}(\hat{\sigma}, \xi)$, with parameters of interest uniquely determined by the associated GEV distribution of the block maxima. The shape parameter ξ in the GP distribution (3.1) controls the behavior of the tail. When $\xi < 0$ the excesses have a distribution with upper bound equal to $u - \hat{\sigma}/\xi$; when $\xi > 0$ the distribution has no upper limit and when $\xi = 0$ the distribution is unbounded and equal to an exponential distribution with parameter $1/\hat{\sigma}$. We have to consider two main things, the estimation of the parameters of the GP and the choice of the threshold u, as in the block maxima method, we have to consider a trade-off between the bias and the variance. For very large threshold we have an increase on the bias of the estimation, while for very small threshold blocks we increase the variance of the estimation. Note that with the GP distribution we can obtain the non-conditional distribution, if (3.2) is

a suitable model for $P(X - u \leqslant y \mid X > u)$, then it follows that

$$P(X > x) = \zeta_u \left\{1 + \xi \left(\frac{x-u}{\sigma}\right)\right\}^{-1/\xi}, \tag{3.3}$$

where $\zeta_u = P(X > u)$. The distribution obtained in (3.3) is of great interest, because sometimes it is more convenient to interpret extreme value models in terms of quantiles or the return levels. The return level x_m is the level exceeded on average once every m observations, and it can be obtain as the solution of $P(X > x_m) = 1/m$.

Bayesian Inference for the GP Distribution

Bayesian inference can be conducted as follows. Let $Y_j = X_i - u$ for $X_i > u$ and $i = 1, \ldots, n$, be a sample of threshold excesses from a $\text{GP}(\sigma, \xi)$, then for sufficiently large threshold u, the ordered exceedances $Y_j = X_{n-u+1,n} - X_{n-u,n}$ with $X_{k,n}$ the k-order statistic of a sample of size n, will have distribution function,

$$F(y \mid \sigma, \xi) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}$$

for $j = 1, \ldots, k$. For the prior distribution, consider Jeffreys' prior:

$$\pi(\sigma, \xi) = \frac{1}{\sigma(1+\xi)\sqrt{1+2\xi}},$$

which is finite for $\xi > -1/2$. Another option is the maximal average data information (MDI) prior given by,

$$\pi(\sigma, \xi) \propto E\{\log f(Y \mid \sigma, \xi)\} = \frac{1}{\sigma} e^{-\xi}.$$

Under the above prior one can then use a Metropolis–Hastings algorithm with proposal density equal to a multivariate Normal on $(\log \sigma, \xi)$ to obtain the posterior of the parameters.

3.2.2 Multivariate Theory of Extreme Values

Many applications of extreme values are inherently multivariate, and statistical methods to analyze multivariate extremes are also of great interest. For instance, in financial data, multivariate modelling of extremes is essential given the nature of dependence between the financial markets and of the stocks in a portfolio. For example, multivariate extreme value modelling has been used for studying the event of joint extreme return losses of pairs of exchange rates (Longin and Solnik, 2001), as well as to investigate the dependence between international markets in periods of high volatility, (Poon et al., 2003, 2004; Castro et al., 2018).

Whenever studying multivariate extremes there are two main targets of interest: The marginal distributions and the dependence structure. To understand the dependence struc-

ture it is convenient to standardize the margins, and a common choice is to standardize to Fréchet margins; other choices for standardization can be found on Beirlant et al. (Section 8.2.6. 2004). It is important to notice that the class of possible limiting dependence structures for the multivariate case cannot be captured in a finite-dimensional parametric family as in the univariate case, which makes the study and theory for the multivariate case more challenging.

Bivariate Case

For ease of exposition we first start with the bivariate case and then extend the main ideas and concepts for the multivariate setup. Following the same structure as in the univariate case we focus on block maxima methods and threshold methods.

Block Maxima Methods

Block maxima approaches for multivariate extremes can be devised in a similar way as in the univariate case but now using componentwise maxima as $M_n = (M_{x,n}, M_{y,n})$, where $M_{x,n} = \max(X_1, \ldots, X_n)$ and $M_{y,n} = \max(Y_1, \ldots, Y_n)$. It is of probabilistic interest to study the possible limiting distributions of the maxima M_n. For this purpose, assume without a loss of generality that for all i both X_i and Y_i have standard Fréchet distributions of the form

$$F(z) = \exp(-1/z), \quad z > 0,$$

this can be easily obtained after applying a suitable transformation. To obtain standard univariate results, consider the re-scaled vector

$$M_n^* = (M_{x,n}^*, M_{y,n}^*) = \{\max(X_1, \ldots, X_n)/n, \max(Y_1, \ldots, Y_n)/n\}. \tag{3.4}$$

The following theorem, Coles (2001, Theorem 8.1), derived the general form of all possible limit laws for M_n^*. This can be regarded as the bivariate version of Theorem 1.

Theorem 3. *Let M_n^* be defined as in (3.4), with (X_i, Y_i) independent vectors with standard Fréchet margins. Then if,*

$$P\left(M_{x,n}^* \leqslant x, M_{y,n}^* \leqslant y\right) \to G(x, y), \quad as \quad n \to \infty$$

with G a non-degenerate distribution function of the form

$$G(x, y) = \exp\{-V(x, y)\}, \quad x > 0, y > 0, \tag{3.5}$$

where

$$V(x, y) = 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) \mathrm{d}H(w), \tag{3.6}$$

and H is a distribution function on $[0,1]$ with the mean constraint:

$$\int_0^1 w \, \mathrm{d}H(w) = 1/2.$$

Theorem 3 determined that the possible limits for (3.4) are in the class of bivariate extreme value distributions, defined in (3.5). This class is in completely identified by the set of distributions H satisfying the mean constraint, H is called spectral measure. The spectral measure H in (3.6) plays a main role characterizing the limit of standardized componentwise maxima. Notice that H is not necessarily differentiable and hence it may not admit a density distribution. One option is to use non-parametric methods (Einmahl et al., 2001; Kiriliouk et al., 2015) to estimate the spectral measure H. In practice, it is usual to resort to parametric models, and some popular examples are listed below

- Logistic family (Coles, 2001, p.146):

$$h(w) = \frac{1}{2}\{a^{-1} - 1\}\{w(1-w)\}^{-1-1/a}\{w^{-1/a} + (1-w)^{-1/a}\}^{a-2} \tag{3.7}$$

  for $w \in (0,1)$. And we can obtain the explicit form of the bivariate extreme value distribution:

$$G(x,y) = \exp\{-(x^{-1/a} + y^{-1/a})\}$$

  with $a \in (0,1)$. Note that because of the symmetry of the model, the mean constraint is automatically satisfied.

- Dirichlet model (Coles and Tawn, 1991):

$$h(w) = \frac{ab\Gamma(a+b+1)(aw)^{a-1}\{b(1-w)\}^{b-1}}{2\Gamma(a)\Gamma(b)\{aw + b(1-w)\}^{a+b+1}} \tag{3.8}$$

  for $w \in (0,1)$ and $a, b > 0$. Note that the model is symmetric when $a = b$.

Figure 3.2 shows how the spectral density h for the logistic model in (3.7) changes with respect to different values of a. The parameter a controls extremal dependence so that as $a \to 1$ the bivariate GEV corresponds to independence, and as $a \to 0$ the bivariate GEV corresponds to perfect dependence.

Threshold Method

The second method to model bivariate extremes is based on the behavior of observations above certain thresholds. Following the ideas of the univariate case, the aim is to obtain a bivariate extension of (3.3), this means, a family of joint distributions $F(x,y)$ defined on $x > u_x$ and $y > u_y$ above large enough thresholds u_x and u_y.

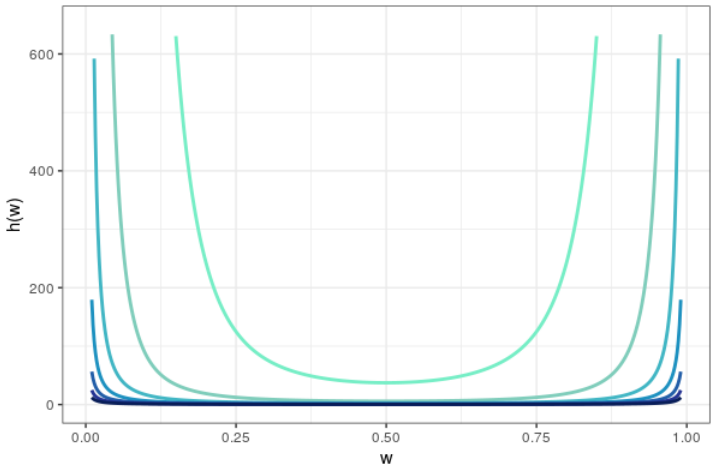


Figure 3.2: Spectral density for the logistic model, for different dependent structures, $a = (0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8)$ colors going form light blue (small $a \to 0$) to dark blue ($a \to 1$) respectively.

Assume that the vectors $\{(x_i, y_i)\}_{i=1}^n$ are independent and with distribution $F(x, y)$. Then, for large enough thresholds u_x and u_y, the marginal distributions of F have approximately the form of (3.3), with parameters $(\zeta_x, \sigma_x, \xi_x)$ and $(\zeta_y, \sigma_y, \xi_y)$ for X and Y respectively. Now, applying a suitable transformation of the form:

$$\tilde{X} = -\left(\log\left[1 - \zeta_x\left\{1 + \frac{\xi_x(X - u_x)}{\sigma_x}\right\}^{-1/\xi}\right]\right)^{-1}$$

for both X and Y, in order to get $(\tilde{X}, \tilde{Y})$ with margins $\tilde{F}$ approximately standard Fréchet for $X > u_x$ and $Y >_y$. Using Equation (3.5) it follows that,

$$\tilde{F}(\tilde{X}, \tilde{Y}) \approx [\exp\{-V(\tilde{x}/n, \tilde{y}/n)\}]^{1/n} = \exp\{-V(\tilde{x}, \tilde{y})\}$$

and since $F(x, y) = \tilde{F}(\tilde{x}, \tilde{y})$ we get for $x > u_x$ and $y > u_y$,

$$F(x, y) \approx G(x, y) = \exp\{-V(\tilde{x}, \tilde{y})\}$$

with thresholds large enough to justify the limit. This can be regarded as the bivariate extension of (3.5). The choice of the threshold is again a trade-off between bias and variance.

Multivariate Case

We will extend some of the bivariate results above to the multivariate setup. Theorem 3 can be extended to the multivariate case. Let $\{\boldsymbol{X}_i\}_{i=1}^n$ be a vector on $\mathbb{R}^d$ of independent and identically distributed random vectors, with distribution F. Define the componentwise standardized maximum as $M_n^* = \{\max_{i=1\ldots n}(X_{i,1})/n, \ldots, \max_{i=1\ldots n}(X_{i,d})/n\}$ and a natural

extension of (3.5) is the d-variate extreme value distribution

$$G(\boldsymbol{x}) = \exp\{-V(\boldsymbol{x})\}, \quad \boldsymbol{x} = (x_1, \ldots, x_d) \in (0, \infty)^d, \tag{3.9}$$

where V is the exponent measure of the form,

$$V(\boldsymbol{x}) = d \int_0^1 \max\left(\frac{w_1}{x_1}, \ldots, \frac{w_d}{x_d}\right) \mathrm{d}H(\boldsymbol{w}), \tag{3.10}$$

and H is a spectral measure defined on the unit simplex,

$$S_d = \{\boldsymbol{w} \in [0, \infty)^d : w_1 + \cdots + w_d = 1\}$$

with the moment constraints:

$$\int_{S_d} w_j \, \mathrm{d}H(\boldsymbol{w}) = 1/d, \quad j = 1, \ldots, d.$$

The spectral measure H defined in (3.10), will play a role in the independence or dependence of the extremes. Independence is achieved if and only if H puts equal mass $1/d$ at vertices $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d$ of the simplex S_d, while extremal dependence occurs when H puts mass only at the barycenter $(1/d, \ldots, 1/d)$.

3.3 Nonstationary Univariate Extremes

Nonstationary GEV

A standard practice to model univariate nonstationary extremes consists on expressing the parameters of the extreme value distribution defined in (3.1) as a function of covariates. Consider a nonstationary sequence of random variables, then the limiting behaviour of the standarized maximum Z_t is described by the conditional GEV,

$$Z_t \sim \mathrm{GEV}(\mu(t), \sigma(t), \xi(t)), \quad t = 1, \ldots, m. \tag{3.11}$$

where the parameters of interest are $\mu(t)$ the time-varying location parameter, $\sigma(t)$, the time-varying scale parameter and $\xi(t)$ the time-varying shape parameter. Define each parameter as a function of covariates, let η denote the parameter μ, σ or ξ, then

$$\eta(t) = h(\boldsymbol{X}^T \boldsymbol{\beta}),$$

for h a link function, $\boldsymbol{X} \in \mathbb{R}^k$ a vector of covariates and $\boldsymbol{\beta} \in \mathbb{R}^k$ a vector of regression parameters.

Bayesian inference for the conditional GEV

Bayesian inference on the time-varying parameters can be conducted as follows, let $Y_1, \ldots, Y_n$ be a random sample from a conditional GEV distribution $\text{GEV}(\mu(t), \sigma(t), \xi(t))$. The likelihood model conditional on t is:

$$Y_i \mid \sigma(t), \mu(t), \xi(t) \sim \text{GEV}(\mu(t), \sigma(t), \xi(t)), \quad i = 1, \ldots, n,$$

let $\boldsymbol{\theta}(t) = (\mu(t), \sigma(t), \xi(t))$ be the parameter of interest, then

$$f(y \mid \boldsymbol{\theta}(t)) = \frac{1}{\sigma(t)} \left\{ 1 + \xi(t) \left(\frac{y - \mu(t)}{\sigma(t)} \right) \right\}^{-1/\xi(t)-1} \exp \left[- \left\{ 1 + \xi(t) \left(\frac{y - \mu(t)}{\sigma(t)} \right)^{-1/\xi(t)} \right\} \right].$$

For the prior distribution, a natural option is to consider a multivariate Normal distribution on the regression coefficients $\boldsymbol{\beta}$ of each parameter $\mu(t), \sigma(t), \xi(t)$ and adapt MCMC models to sample from the posterior distributions. Laplace priors, along the same lines of de Carvalho et al. (2022), would be another natural option and it would induced shrinkage and Bayesian lasso versions of the later models.

Heteroscedastic extremes

Recently, Einmahl et al. (2016) proposed another framework for modelling univariate nonstationary extremes, where the main object of interest is the *scedasis function*. Heteroscedastic extremes assumes independent but non-identically distributed observations, where the non-stationarity arise from changes in the frequency of extremes. We made this precise in the following definition,

Definition 8 (Scedasis function)**.** *Let $Y_1, \ldots, Y_T$ be independent observations from distribution functions $F_1, \ldots, F_T$ respectively and with common infinite upper endpoint, the scedasis function is defined as*

$$c\left(\frac{t}{T}\right) = \lim_{z \to \infty} \frac{1 - F_t(y)}{1 - F(y)}, \tag{3.12}$$

for some unknown baseline distribution $F(y)$ with an infinite upper endpoint.

If the following condition is imposed,

$$\int_0^1 c(s) \mathrm{d}s = 1,$$

the scedasis function c is uniquely defined, and can now be interpreted as the frequency of extremes. If the scedasis is uniform, $c = 1$, we refer to *homoscedastic extremes*, and changes in the scedasis we refer to as *heteroscedastic extremes*. An important consequence of (3.12) is that if F is in the Fréchet domain of attraction with extreme value index ξ, then all $F_1, \ldots, F_T$ are in the Fréchet domain of attraction with the same extreme value index ξ. In other words,

the extreme value index is constant over time. Note that the extreme value index contains information regarding the magnitude of extremes.

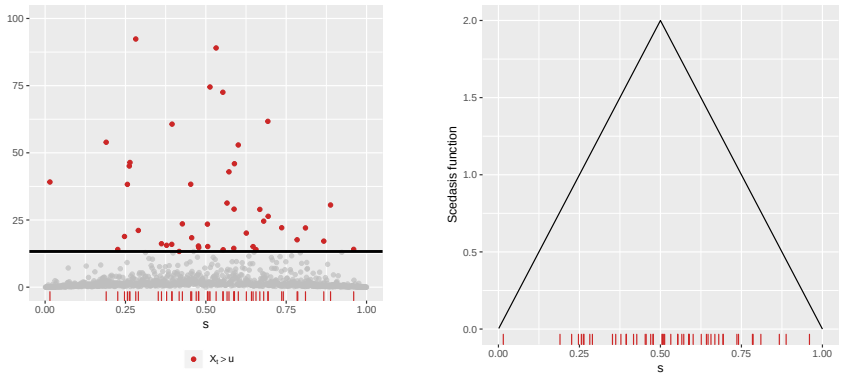


Figure 3.3: Simulated data from $F_t(x) = \exp\{-c(t/T)/x\}$, with $c(s) = 2s/.5$ for $s \in [0, .5)$ and $c(s) = (2 - 2s)/.5$ for $s \in [.5, 1]$, along with the true scedasis function $c(s)$.

Inference for the scedasis function and the extreme value index ξ, was introduced by Einmahl et al. (2016) using a kernel estimation method for the scedasis function and the usual Hill estimator for the extreme value index. For a suitable sequence $k = k(T)$, consider the following estimators,

$$\hat{c}(s) = \frac{1}{kh} \sum_{t=1}^{T} 1(Y_t > Y_{(T-k)}) \mathbb{K}\left(\frac{s - t/T}{h}\right),$$

$$\hat{\xi} = \frac{1}{k} \sum_{j=1}^{k} \log(Y_{(T-j+1)}) - \log(Y_{(T-k)}),$$

where $Y_{(i)}$ is the i-th order statistic of $Y_1, \ldots, Y_T$ and $\mathbb{K}$ is a continuous kernel. Asymptotic properties of the estimators were also studied for the case of heteroscedastic extremes. Chapter 5 will propose Bayesian inference for these parameters, along with extensions to the multivariate case.

3.4 Nonstationary Copulas

In this section, we introduce some theory on nonstationary copulas, but before we need to introduce basic concepts of copulas and the class of extreme value copulas. The notion of copulas became increasingly popular in recent years, this class of functions was introduced by Sklar (1959) and have been widely studied, standard references for theory of copulas are Embrechts et al. (2001), Nelsen (2006).

Basic concepts

First, we will give a basic introduction to theory of copulas. Second, we will introduce the class of extreme value copulas and the class of time-varying copulas.

Definition 9. *A d-dimensional copula (d-copula), is a d-variate distribution function on* $\mathbb{I}^d$ *with marginals univariate uniform distributions on* $\mathbb{I}$*, for every* $d \geqslant 2$*, with* $\mathbb{I}$ *the unit closed interval and the Cartesian product* $\mathbb{I}^d = \mathbb{I} \times \cdots \times \mathbb{I}$ *of d unit closed intervals.*

Theorem 4. *A function* $C : \mathbb{I}^d \to \mathbb{I}$ *is a copula if and only if, it satisfies the following properties:*

1. $C(\boldsymbol{u}) = u_j$ *when* $\boldsymbol{u}$ *is equal to one in all the components, with exception of the* $j-th$ *component which is equal to* u_j*, for all* $j \in 1, \ldots, d$*.*
2. $C(\boldsymbol{u}) \leqslant C(\boldsymbol{v})$ *for all* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{I}^d$*, such that* $\boldsymbol{u} \leqslant \boldsymbol{v}$*.*
3. C *is d-increasing.*

As a direct consequence of Theorem 4, it can be prove that $C(\boldsymbol{u}) = 0$ if $u_j = 0$ for any j.

Some canonical examples of copulas are:

- Independent copula: $\Pi_d(\boldsymbol{u}) = u_1 \times \cdots \times u_d$, with $\boldsymbol{u} = (u_1, \ldots, u_d)$.
- Commonotonicity copula: $M_d(\boldsymbol{u}) = \min\{u_1, \ldots, u_d\}$, with $\boldsymbol{u} = (u_1, \ldots, u_d)$.
- Countermonotonicity copula: $W_2(\boldsymbol{u}) = \max\{u_1 + u_2 - 1, 0\}$, with $\boldsymbol{u} = (u_1, u_2)$.

The following theorem provides the foundation of copulas, by linking the multivariate joint distribution with their marginal distributions via a copula.

Theorem 5 ((Sklar 1959))**.** *Let* F *be a d-dimensional distribution function with marginals* $F_1, \ldots, F_d$*. Denote* R_j *and the range of* F_j*. Then, there exists a copula* C *such that for all* $(x_1, \ldots, x_d) \in \mathbb{R}^d$*,*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{3.13}$$

where C *is uniquely determined on* $R_1 \times \ R_2 \times \cdots \times R_d$ *and if* $F_1, \ldots, F_d$ *are continuous then* C *it is unique.*

The converse implication of Theorem 5 is also satisfied, if C is a d-copula and $F_1, \ldots, F_d$ are univariate distribution functions, then $F : \mathbb{R}^d \to \mathbb{I}$ defined in (3.13) is a d-dimensional

Class of Archimedean Copulas

Here, we revise an important family of copulas refer to as Archimedean copulas, this class include a great variety of dependence structures and have convenient properties. An Archimedean copula is defined through a generator φ, as

$$C(u_1, \ldots, u_d) = \varphi^{-1}\{\varphi(u_1) + \cdots + \varphi(u_d)\} \tag{3.14}$$

where $(u_1, \ldots, u_d) \in [0,1]^d$, φ should be strictly decreasing and convex and φ^{-1} is the inverse of φ with φ^{-1} a d-monotonic function. We introduce the most commonly Archimedean copulas used in the literature:

- Clayton copula: $C(u_1, \ldots, u_d) = \max\{(u_1^{-\alpha} + \cdots + u_d^{-\alpha} - 1)^{-1/\alpha}, 0\}$, the generator is $\varphi(t) = (t^{-\alpha} - 1)/\alpha$, with $\alpha \in [-1, \infty) \backslash \{0\}$.
- Frank copula: $C(u_1, \ldots, u_d) = \frac{-1}{\alpha} \ln[1 + \frac{\{\exp(-\alpha u_1)-1\} \times \cdots \times \{\exp(-\alpha u_d)-1\}}{\exp(-\alpha)-1}]$, the generator is $\varphi(t) = -\ln[\{\exp(-\alpha t) - 1\}/\{\exp(-\alpha) - 1\}]$ for $\alpha \in (-\infty, \infty) \backslash \{0\}$.
- Gumbel copula: $C(u_1, \ldots, u_d) = \exp[-\{(-\log(u_1))^\alpha + \cdots + (-\log(u_d))^\alpha\}]^{1/\alpha}$, the generator is $\varphi = (-\ln t)^\alpha$ with $\alpha \in [1, \infty]$.

Figure 3.4 shows an example of the Archimedean copulas introduced above, for specific parameters.

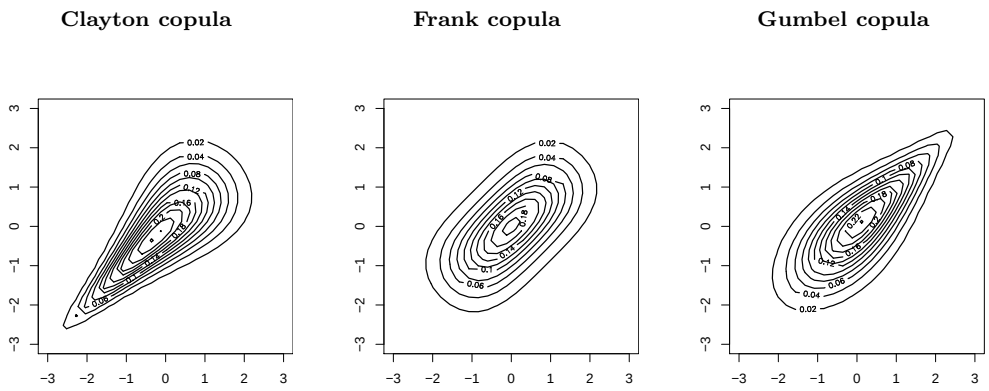


Figure 3.4: Contour plots of bivariate Archimedean copulas, $\alpha = \{2, 4, 2\}$ from left to right respectively.

Extreme Value Copulas

Extreme value theory and copulas have some points of intersection, for example, the class of extreme value copulas is appropriate for modelling the dependence structure of extreme values. Let $Y_i = (Y_{i1}, \ldots, Y_{id})$, $i \in \{1, \ldots, n\}$ be a sample of random vectors with distribution functions F and marginals $F_1, \ldots, F_d$ and copula C_F. Denote the component wise maxima by

$$M_n = \left(\bigvee_{i=1}^{n} \{Y_{i1}\}, \ldots, \bigvee_{i=1}^{n} \{Y_{id}\} \right)$$

where $\vee$ is the maximum. Then, it follows that the copula C_n of the vector M_n is equal to

$$C_n(u_1, \ldots, u_d) = C_F(u_1^{1/n}, \ldots, u_d^{1/n}).$$

for all $(u_1, \ldots, u_d) \in [0,1]^d$. We will make this idea formal in the following definition,

Definition 10. *A d-dimensional copula is called an extreme-value copula is there exists a copula C_F such that*

$$C_F(u_1^{1/n}, \ldots, u_d^{1/n}) \to C(u_1, \ldots, u_d) \quad (n \to \infty) \tag{3.15}$$

for all $(u_1, \ldots, u_d) \in [0,1]^d$.

And we say that the copula C_F is in the domain of attraction of C.

Theorem 6. *A copula is an extreme value copula if and only if it is max-stable, it means, that it satisfies the following relationship,*

$$C(u_1, \ldots, u_d) = C(u_1^{1/m}, \ldots, u_d^{1/m})^m,$$

for all $(u_1, \ldots, u_d) \in [0,1]^d$ and $m \geqslant 1$ an integer.

Time-varying Copulas

Here, we introduce the class of conditional copulas, an extension of classic copulas and we add focus on the conditional time version.

Definition 11 (Conditional copula)**.** *The conditional copula of $(X, Y) \mid Z$, with conditional margins $F_{X|Z}(\cdot \mid z)$ and $F_{Y|Z}(\cdot \mid z)$, is the conditional joint distribution function of $U \equiv F_{X|Z}(X \mid z)$ and $V \equiv F_{X|Z}(X \mid z)$ given $Z = z$.*

The following theorem extends Sklar theorem to the conditional version.

Theorem 7. *Let F be the conditional joint distribution function of $(X, Y) \mid Z$, with marginals $F_{X|Z}(\cdot \mid z)$ and $F_{Y|Z}(\cdot \mid z)$, continuous in x and y, for all z. Then, there exists a unique copula $C(\cdot \mid z)$ such that*

$$F(x, y \mid z) = C(F_{X|Z}(x \mid z), F_{Y|Z}(y \mid z) \mid z). \tag{3.16}$$

The converse implication of Theorem 7 is also satisfied, if $C(\cdot \mid z)$ is a conditional copula and let the conditional margins be $F_{X|Z}(\cdot \mid z)$ and $F_{Y|Z}(\cdot \mid z)$, then $F(x, y \mid z)$ defined in (3.16) is a conditional joint distribution function with marginals $F_{X|Z}(\cdot \mid z)$ and $F_{Y|Z}(\cdot \mid z)$. To simplify notation, denote the conditional copula as $C_z(\cdot) = C(\cdot \mid z)$. Theorem 7 can be easily extended to the d-dimensional setting, as long as we use the same conditional variable(s) for all margins as well as for the copula. Note that using this conditional setup we can construct copulas with the dependence structure conditional on time, a version of time-varying copulas. Fermanian

and Wegkamp (2012) introduced the concept of *pseudo-copulas*, as another method of studying time-varying dependence structures. A *pseudo-copulas* is a d-variate distribution function on $\mathbb{I}^d$ with arbitrary margins on $\mathbb{I}$, for every $d \geqslant 2$. This new setup will allow the marginal distributions to be conditional only on the past observations of each particular margin, and not all the past information, providing flexibility on real world applications.

3.5 Dependence Structures for Extreme Values

Asymptotic Dependence and Asymptotic Independence

In this section, we focus on a class of models describing the tails of asymptotically independent and dependent distributions, starting with the bivariate case and then extending them to the multivariate case.

Asymptotic Dependence

Let (Y_1, Y_2) a random vector with distribution function F and continuous marginals F_1 and F_2 respectively. Define the coefficient of extremal dependence as:

$$\chi = \lim_{y \to y^*} P(Y_2 > y \mid Y_1 > y), \tag{3.17}$$

if the limit exists, where y^* is the right end-point of the common marginal distribution. In the case where the marginal distributions are non-identical distributed we can extend the result in (3.17). Define the transform variables $U_j = F_j(X_j)$ for $j = 1, 2$, with U_j now to be uniform distributed on $(0, 1)$. Now, define the coefficient of extremal dependence as,

$$\chi = \lim_{u \to 1} P(U_2 > u \mid U_1 > u) \tag{3.18}$$

The coefficient of extremal dependence χ is interpreted as the tendency of one variable being extreme given that the other is extreme. If $\chi = 0$, the variables are asymptotically independent, if $0 < \chi \leqslant 1$ the variables are asymptotically dependent. We can define also, for $0 < u < 1$,

$$\chi(u) = 2 - \frac{\log P(U_1 < u, U_2 < u)}{\log u},$$

and it follows that,

$$\chi = \lim_{u \to 1} \chi(u),$$

and for the extreme value distribution $G(x, y) = \exp\{-V(x, y)\}$,

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

uniformly over u. Observe that the condition $\chi = 0$ also implies that if F_1 and F_2 are in the domain of attraction of a GEV G_1 and G_2 respectively, then $\chi = 0$ if and only if the bivariate distribution F is in the domain of attraction of a bivariate GEV of the form $G(x, y) = G_1(x)G_2(x)$. Estimation of χ can help to understand the limit behavior as $u \to 1$.

Asymptotic independence

When we have asymptotically dependent variables, it means, $0 < \chi \leqslant 1$ and the value of χ increases according to the degree of dependence at extreme values. However, when $\chi = 1$ it does not provide any measure of discrimination for asymptotically independent distributions. For this reason we will define an alternative measure of extremal dependence, for $0 < u < 1$,

$$\overline{\chi}(u) = \frac{2(1-u)}{\log P\{U_1 > u, U_2 > u\}} - 1 \tag{3.19}$$

it follow that,

$$\overline{\chi} = \lim_{u \to 1} \overline{\chi}(u) \tag{3.20}$$

if the limit exists. The coefficient in (3.20) can take values between $[-1, 1]$. For asymptotically independent variables, it follows that $-1 \leqslant \overline{\chi} < 1$. Then when we look together at $(\chi, \overline{\chi})$ we can have a summary of extremal dependence:

- If $\overline{\chi} = 1$ and $0 < \chi \leqslant 1$ the variables are asymptotically dependent and χ help understand the strength of dependence, within this class.
- If $\chi = 0$ and $-1 \leqslant \overline{\chi} < 1$ the variables are asymptotically independent and $\overline{\chi}$ help understand the strength of dependence, within this class.

Bivariate tail dependence

Coefficient of tail dependence

The coefficient of tail dependence was introduced by Ledford and Tawn (1996). It is convenient to transform the marginals to standard Fréchet distributions using $Z_j = -1/\log F_j(Y_j)$ for $j = 1, 2$. Then, to define the coefficient of tail dependence, assume that the joint distribution of (Z_1,Z_2) is regularly varying,

$$P(Z_1 > z, Z_2 > z) = \mathcal{L}(z) z^{-1/\eta}, \quad z > 0, \tag{3.21}$$

where η is a positive constant known as the coefficient of tail dependence, and $\mathcal{L}$ is a slowly varying function, that is, $\mathcal{L}$ such that $\mathcal{L}(zl)/\mathcal{L}(z)$ as $z \to \infty$, for all $l > 0$. Note that if we define the variable $Z = \min\{Z_1, Z_2\}$, it follows that $P\{Z_1 > z, Z_2 > z\} = P\{Z > z\}$, according to (3.21), η will be the extreme value index of the variable Z. If $\eta = 1$ and $\lim_{z \to \infty} \mathcal{L}(z) = c$ for

some $0 < c \leqslant 1$, then the variables are asymptotically dependent of degree c. If $0 < \eta < 1$ or if $\eta = 1$ and $\lim_{z\to\infty} \mathcal{L}(z) = 0$, then the variables are asymptotically independent.

Multivariate tail dependence

Stable tail dependence function

For the multivariate case, a quantity of interest is the stable tail dependence function, linked to the exponent measure V, and defined by

$$l(\boldsymbol{v}) = V(1/v_1, \ldots, 1/v_d), \quad \boldsymbol{v} = (v_1, \ldots, v_d) \in [0, \infty)^d \tag{3.22}$$

The stable tail dependence function l has the following properties:

- $l(s\cdot) = sl(\cdot)$ for $0 \leqslant s \leqslant \infty$.
- $l(\boldsymbol{e}_j) = 1$ for $j = 1, \ldots, d$, and $\boldsymbol{e}_j$ the jth unit vector in $\mathbb{R}^d$.
- $v_1 \vee \cdots \vee v_d \leqslant l(\boldsymbol{v}) \leqslant v_1 + \cdots + v_d$ for $\boldsymbol{v} = (v_1, \ldots, v_d) \in [0, \infty)^d$.

Note that, the lower and upper bounds for the stable tail dependence function corresponds to complete dependence and independence respectively. More details about multivariate extreme values can be found in Pickands (1981), Coles and Tawn (1991), Coles (2001), Beirlant et al. (2004).

3.6 Nonstationary Multivariate Extremes

In this section, we introduce the *conditional angular density*, which is the main object of study of multivariate nonstationary extremes. The two previous sections provided preparations to understand the definitions revised in this section.

Conditional extreme value distributions

The conditional bivariate extreme value distribution (BEV) introduced by de Carvalho (2016b), Castro et al. (2018) is defined as a conditional varying version of (3.5),

$$G_x(y_1, y_2) = \exp\{-V_x(y_1, y_2)\}, \quad y_1 > 0, y_2 > 0 \tag{3.23}$$

where

$$V_x(y_1, y_2) = 2\int_0^1 \max\left(\frac{w}{y_1}, \frac{1-w}{y_2}\right) \mathrm{d}H(w \mid X = x), \tag{3.24}$$

and $H_x(\cdot) = H(\cdot \mid X = x)$ is a conditional angular measure on $[0, 1]$ for $x \in \mathcal{X} \subseteq \mathbb{R}$ with the moment constrain:

$$\int_0^1 w \mathrm{d}H_x(w) = 1/2, \quad x \in \mathcal{X}.$$

If $H_x(\cdot)$ is absolutely continuous, let $h_x = \mathrm{d}H_x/\mathrm{d}w$ be the *conditional angular density* and define the angular surface as the set conditional angular densities $\{h_x(w) : w \in [0,1] x \in \mathcal{X}\}$. The bivariate extreme value distribution can be defined in terms of the *conditional tail dependence function* as follows

$$G_x(y_1, y_2) = \exp\left\{-l_x\left(\frac{1}{y_1}, \frac{1}{y_2}\right)\right\}, \quad y_1 > 0, y_2 > 0 \tag{3.25}$$

Note that the conditional tail dependence function, is an extension of (3.22) to the conditional setting.

Example 4 (Conditional logistic family)**.** *Considered as an extension of the logistic family defined in* (3.7)*; if we index by a covariate, say time, the conditional logistic angular density is,*

$$h(w) = \frac{1}{2}\{a_t^{-1} - 1\}\{w(1-w)\}^{-1-1/a_t}\{w^{-1/a_t} + (1-w)^{-1/a_t}\}^{a_t - 2},$$

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

Figure 3.5 illustrates the conditional logistic angular surface for different values of a_t, this figure is adapted from Castro et al. (Example 1 2018).

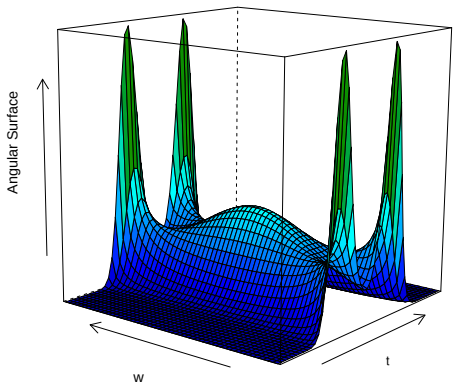


Figure 3.5: Angular surface from the logistic conditional family with $a_t = \Phi(t^2)$.

For more details on conditional angular measures see de Carvalho (2016b), Castro and de Carvalho (2017), Castro et al. (2018). Note that the *conditional angular density* is equivalent to a conditional extreme value copula, providing a general and flexible tool to model *conditional angular densities* via a copula. In order to make this connection clear, assume that the margins are unit Fréchet distributed, then we can represent (3.25) via a conditional extreme value copula, as follows

$$C_x(u, v) = G_x\left(\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-u)}\right), \quad (u, v) \in [0,1]^2 \tag{3.26}$$

with C_x a conditional extreme value copula, following Definition 11. Finally, notice that the results in this section will allow us to write the extreme value distributions via copulas, providing a flexible tool to construct multivariate heavy-tailed distributions. Different approaches have been proposed to model the *conditional angular density* via Bayesian nonparametric models. For instance, for the bivariate case, Guillotte et al. (2011) proposed a Bayesian framework using a censored likelihood. Boldi and Davison (2007) proposed a model based on finite Dirichlet mixtures, this model was also studied by Sabourin and Naveau (2014) who proposed a novel parametrization where the moment constraint is automatically satisfied. Another approach was proposed by Marcon et al. (2016), modelling the extremal dependence via Bernstein polynomials. Hanson et al. (2017) proposed a similar approach by considering mean-constrained Bernstein polynomials.

Part II—Main Contributions

Chapter 4

Heavy-Tailed Pitman–Yor Mixture Models

The first contribution of this chapter focuses on the characterization of the tails of the so-called Pitman–Yor process, which includes the Dirichlet process as a particular case. We show that the right tail of a Pitman–Yor process, known as the stable law process, is heavy-tailed, provided that the centering distribution is itself heavy-tailed. A second contribution of the chapter rests on the development of two classes of heavy-tailed mixture models and the assessment of their relative merits. Multivariate extensions of the proposed heavy-tailed mixtures are here devised along with a predictor-dependent version so to learn about the effect of covariates on a multivariate heavy-tailed response. The simulation study suggests that the proposed method performs well in a variety of scenarios, and we showcase the application of the proposed methods in a neuroscience dataset. This chapter is based on a paper of the same name, submitted to JRSS B, with co-authors Miguel de Carvalho and Luis Gutierrez. The methodology is available in the R package `pityoR`*, soon to be published in CRAN.*

4.1 Introduction

Thousands of heavy-tailed signals are produced on a day-to-day basis across the globe in fields as diverse as engineering, finance, and medicine. And yet, despite the widespread need for modelling these, heavy tails remain a weak spot of several established random probability measures such as the Dirichlet process (e.g. Ghosal and Van der Vaart, 2015, Section 4.3).

Prior to introducing the main problems to be addressed, and the main contributions of this chapter, we first lay the groundwork. The Pitman–Yor process is a random probability

measure that has received considerable attention in recent years (e.g. Pitman and Yor, 1997; Ishwaran and James, 2001; Teh, 2006; Bassetti et al., 2014; Miller and Harrison, 2014; Canale et al., 2017; Arbel et al., 2019; Lijoi et al., 2020; Corradin et al., 2021) and that includes the Dirichlet process as a particular case. Some Bayesian nonparametric approaches, such as the Pitman–Yor process, can be understood as an extension of standard parametric methods in the sense that they are centered a priori around a parametric model, $\{G_{0,\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q\}$, but assign positive mass to a variety of alternatives. Thus, a recurring theme in much of the Bayesian nonparametric literature is to regard a parametric approach—known as the baseline or centering distribution—as a reference, while allowing for other alternative models to take over when data suggests that the parametric model is inappropriate. See the monographs of Müller et al. (2015a) and Ghosal and Van der Vaart (2015) or the review paper of Müller and Mitra (2013) for an introduction to Bayesian nonparametric inference.

It is well-known that the tails of the Dirichlet are exponentially much thinner than those of the baseline (Ghosal and Van der Vaart, 2015, Section 4.2.3). Motivated by this, this chapter opens with the question on whether this is simply a property of the Dirichlet process or whether it is more generally an attribute of the Pitman–Yor process. Hence, the first contribution of this chapter will focus on the characterization of the tails of the Pitman–Yor process. In particular, we will derive envelopes for the trajectories of the tail of a particular instance of the Pitman–Yor process, known as the stable law process. As will be discussed below (Section 4.2.1), the latter envelopes combined with those of Doss and Sellke (1982) offer an almost complete portrait of the tails of the Pitman–Yor process. In addition, we then show that the tail of the stable law process is only moderately thinner than that of the baseline; in particular, our results imply that the tail of the stable law process is heavy-tailed, provided that the baseline is itself heavy-tailed. This result is in sharp contrast with the Dirichlet process, given that, as mentioned earlier, its tail is exponentially much lighter than that of the centering.

A second contribution of the chapter rests on the development of two classes of heavy-tailed mixture models and the assessment of their relative merits. The heavy-tailed mixture models devised here have links with the phase-type scale mixtures of Bladt and Rojas-Nandayapa (2018) and the infinite mixtures of Pareto distributions of Tressou (2008). Our focus differs however from these papers in several important ways. Some key differences are that we take a general view of heavy-tailed Pitman–Yor mixtures and take advantage of our novel characterization of its tail. In addition, by keeping a general focus in mind, our theoretical and numerical analyses will reveal that there are some good reasons for preferring stable process scale mixtures over Pitman–Yor mixtures built from heavy-tailed kernels. Finally, motivated by the fact that heavy-tailed data are frequently multivariate—and since covariates are often available—we further extend the proposed heavy-tailed scale mixture models to model these as well. In other words, multivariate extensions of the proposed heavy-tailed mixtures are also devised below along with a predictor-dependent version to learn about the effect of covariates on a multivariate heavy-

tailed response. A final comment on the jargon of heavy tails is in order. Following the standard convention in the literature on heavy tails (e.g. Resnick, 2007), here we will characterize these via regular variation (Bingham et al., 1989). A distribution function $F(y) = P(Y \leqslant y)$, or its density $f = \mathrm{d}F/\mathrm{d}y$ in case it exists, is said to have a regularly varying tail, with tail index $\alpha \equiv \alpha(F) > 0$, if

$$\lim_{y\to\infty} \frac{P(Y > yt)}{P(Y > y)} = t^{-\alpha}. \tag{4.1}$$

The smaller the tail index, the slower the decay of the tail, $1 - F(y)$, to 0 as $y \to \infty$, and thus the more heavy-tailed is the distribution. Throughout, the notation $1 - F \in \mathrm{RV}_{-\alpha}$ is used to denote that F verifies (4.1).

The remainder of this chapter unfolds as follows. In Section 4.2, we study the tails of the Pitman–Yor process and construct two classes of heavy-tailed Pitman–Yor mixture models. In Section 4.3 we expand the proposed toolbox to the multivariate setting as well as to a conditional framework. Section 4.4 illustrates the performance of the proposed methods and reports the main findings of our numerical studies. An application of the proposed methods to a neuroscience case study is given in Section 4.5. Finally, in Section 4.6 we present closing remarks. Proofs and further technical details are available in Section 4.7; the R package `pityoR`, available from the Supporting materials A, implements instances of the methods proposed herein.

4.2 Heavy-tailed Pitman–Yor Mixture Models

4.2.1 On the Tails of the Pitman–Yor Process

Subordinator-Type Representations: We start by studying the tails of the Pitman–Yor process since its properties will be vital for constructing our class of mixture models. The main result of this section is Theorem 9, but before we can discuss its implications, we first lay the groundwork. Recall that a random probability measure G follows a Pitman–Yor process, here denoted as $G \sim \mathrm{PYP}(D, M, G_0)$, if it admits a stick-breaking representation of the type,

$$G = \sum_{h=1}^{\infty} \pi_h \delta_{Y_h}, \qquad Y_h \overset{\text{iid}}{\sim} G_0. \tag{4.2}$$

Here, δ_Y denotes a point mass at Y, and the weights are generated by a stick-breaking scheme, that is, $\pi_h = V_h \prod_{k<h}(1 - V_k)$ with $V_h \overset{\text{ind}}{\sim} \mathrm{Beta}(1 - D, M + hD)$, for $h \in \mathbb{N} = \{1, 2, \dots\}$. The parameters of the Pitman–Yor process are known as discount (D), precision (M), and centering distribution function (G_0) and are subject to the constraints $0 \leqslant D < 1$ and $M > -D$. The Dirichlet process, $\mathrm{DP}(M, G_0)$, is a particular case with $D = 0$, and the case $M = 0$ is known as the class of stable law processes, $\mathrm{SP}(D, G_0)$.

In this subsection, we will examine the behavior of $1 - G(y)$, as y approaches the right endpoint, $y_+ = \sup\{y : G(y) < 1\}$, where $G(y)$ is the random distribution function of a

$\mathrm{PYP}(D, M, G_0)$. Recall that both $G \sim \mathrm{PYP}(D, M, G_0)$ and G_0 are supported over the same set, and thus the right endpoints of G and G_0 coincide. Following the standard convention in extreme value analysis, we focus on the right tail, but all claims below apply to the left tail with minor adjustments.

While the stick-breaking representation of the Pitman–Yor process in (4.2) is handy for a variety of contexts, the process also admits subordinator-type representations, which turn out to be more suitable for studying its tails. A subordinator, $\{S(t) : t \geqslant 0\}$, is an increasing stochastic process over the positive real line that has independent and homogeneous increments (e.g. Applebaum, 2009, Ch. 1). By the so-called Lévy–Khintchine representation (e.g. Bertoin, 1999, Section 1.2), a subordinator is fully characterized by its Laplace exponent,

$$\Phi(\lambda) = \mathtt{k} + \mathtt{d}\lambda + \int_0^\infty (1 - e^{-\lambda\, u})\nu(\mathrm{d}u), \qquad \lambda \geqslant 0,$$

that obeys $\mathrm{E}[\exp\{-\lambda S(t)\}] = \exp\{-t\Phi(\lambda)\}$, for $t \geqslant 0$; here, $\mathtt{k} > 0$ is the killing rate, $\mathtt{d} > 0$ is the drift coefficient, and ν is a measure on $(0, \infty)$—known as Lévy measure—that governs the law of the increments and which obeys the constraint $\int_0^\infty \min(1, u)\, \nu(\mathrm{d}u) < \infty$. It is well-known that both the Dirichlet process and stable law processes admit subordinator representations (e.g. Pitman and Yor, 1997, Propositions 5–6), a fact that we briefly review in the following examples.

Example 5 (Subordinator representation of $\mathrm{PYP}(0, M, G_0)$)**.** If $G \sim \mathrm{DP}(M, G_0)$, then

$$G(y) = \frac{\gamma\{MG_0(y)\}}{\gamma(M)}, \quad y \in \mathbb{R}. \tag{4.3}$$

See, for instance, Ferguson (1973, 1974) and Ghosal and Van der Vaart (2015, Section 4.2.3). Here, γ is a Gamma process, that is, $\mathtt{k} = \mathtt{d} = 0$ and $\nu(\mathrm{d}u) = u \exp(-u)\, \mathrm{d}u$, for $u > 0$.

Example 6 (Subordinator representation of $\mathrm{PYP}(D, 0, G_0)$)**.** If $G \sim \mathrm{SP}(D, G_0)$, then

$$G(y) = \frac{S\{G_0(y)\}}{S(1)}, \quad y \in \mathbb{R}. \tag{4.4}$$

Here, S is a D-stable subordinator, that is, $\mathtt{k} = \mathtt{d} = 0$ and $\nu(\mathrm{d}u) = D/\Gamma(1-D)u^{-1-D}$, for $u > 0$, where $\Gamma(z) = \int_0^\infty u^{z-1} \exp(-u)\, \mathrm{d}u$ is the gamma function. The stable process representation in (4.4) is an immediate consequence of Pitman and Yor (1997, Proposition 6).

On the Tails of the Pitman–Yor Process: The tails of the Dirichlet process are much lighter than those of the centering distribution, with probability one, a fact that can be shown using the representation from Example 5.

Theorem 8 (Tails of $\mathrm{PYP}(0, M, G_0)$)**.** *Let $G(y)$ be the distribution of a $DP(M, G_0)$ process,*

with non-atomic G_0. Then,

$$\liminf_{y\to y_+} \frac{1-G(y)}{g_s\{1-G_0(y)\}} = \begin{cases} 0, & \text{if } s<1, \\ \infty, & \text{if } s>1, \end{cases} \quad a.s.,$$

$$\limsup_{y\to y_+} \frac{1-G(y)}{h_r\{1-G_0(y)\}} = \begin{cases} 0, & \text{if } r>1, \\ \infty, & \text{if } r\leqslant 1, \end{cases} \quad a.s.,$$

with $g_s(t) = \exp\{-s\log|\log t|/t\}$ and $h_r(t) = \exp\{-1/(t|\log t|^r)\}$ for $0<t<1$.

Proof. See Doss and Sellke (1982) or Ghosal and Van der Vaart (2015, Theorem 4.22). □

It follows from Theorem 8 that for large values of y, with probability one, we have the following lower and upper envelopes for the tail,

$$\exp\left[-\frac{s\log|\log M\{1-G_0(y)\}|}{M\{1-G_0(y)\}}\right] \leqslant 1-G(y) \leqslant \exp\left[-\frac{1}{M\{1-G_0(y)\}|\log M\{1-G_0(y)\}|^r}\right], \tag{4.5}$$

for $s<1$ and $r>1$, where $G\sim \mathrm{PYP}(0,M,G_0)$. Hence, the tails of the Dirichlet process are almost exponential, even if G_0 is heavy-tailed. Despite being one of the most popular Bayesian nonparametric priors, such deficiency of the DP rules out its use when the goal is to model heavy tails, extreme values, and risk. As shown next, what happens with the stable law process is substantially different as its tails are close to those of the baseline.

Theorem 9 (Tails of $\mathrm{PYP}(D,0,G_0)$)**.** *Let $G(y)$ be the distribution of an $SP(D,G_0)$ process, with non-atomic G_0. Then,*

$$\liminf_{y\to y_+} \frac{1-G(y)}{l\{1-G_0(y)\}} = D(1-D)^{(1-D)/D}/S(1) \quad a.s.,$$

$$\limsup_{y\to y_+} \frac{1-G(y)}{u_r\{1-G_0(y)\}} = \begin{cases} 0, & r>1, \\ \infty, & r\leqslant 1, \end{cases} \quad a.s.,$$

with $l(t) = t^{1/D}\{\log|\log t|\}^{1-1/D}$ for $0<t<e^{-1}$, and $u_r(t) = t^{1/D}|\log t|^{r/D}$ for $0<t<e^{-r}$.

Theorem 9 warrants some remarks. The upshot of Theorem 9 is that the tails of a random distribution following a $\mathrm{PYP}(D,0,G_0)$ are almost as heavy as those of the centering, G_0. Indeed, a consequence of Theorem 9 is that for large values of y, with probability one,

$$\{1-G_0(y)\}^{1/D}[\log|\log\{1-G_0(y)\}|]^{1-1/D} \leqslant 1-G(y) \leqslant \{1-G_0(y)\}^{1/D}|\log\{1-G_0(y)\}|^{r/D}, \tag{4.6}$$

for $r>1$ and $1-G_0(y)<e^{-r}$. Numerical illustrations of the asymptotic envelopes in (4.6) are presented in Figure 4.1. Another implication of Theorem 9 is that if the tail of the centering distribution of $\mathrm{PYP}(D,0,G_0)$ is heavy-tailed, then so will be that of the corresponding process,

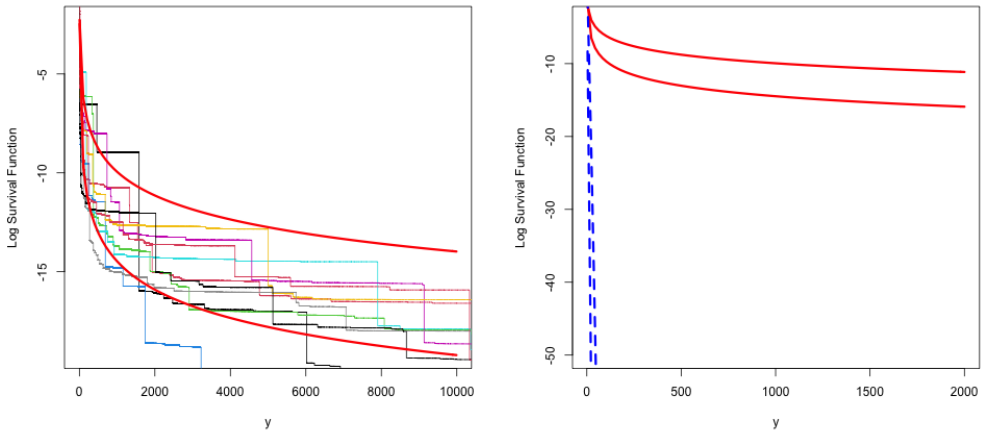


Figure 4.1: Asymptotic Envelopes for Example 7. Left: Asymptotic envelopes that follow from Theorem 9 along with random trajectories of the log survival functions from a PYP$(0.5, 0, G_0)$. Right: The same envelopes for log survival function of PYP$(0.5, 0, G_0)$ (solid) against those of PYP$(0, 1, G_0)$ (dashed); G_0 is the standard unit Pareto distribution.

though with a lighter tail. Figure 4.2 presents the asymptotic envelopes that stem from Theorem 2 in the chapter along with 10 random trajectories of a Pitman–Yor process with discount parameter $D = 0.4$, centered at a variety of parametric models. The random trajectories were simulated using Algorithm 1 from Arbel et al. (2019), for exact simulation from an ε-PYP. As it can be seen from Figure 4.2, the random trajectories follow closely the asymptotic envelopes that stem from Theorem 9.

Theorem 10 (Stability of the Heavy-tail Property)**.** *If $G \sim PYP(D, 0, G_0)$, with $D \in (0, 1)$, and G_0 is non-atomic and has a regularly varying tail, with tail index $\alpha_0 \equiv \alpha(G_0) > 0$, then G has a regularly varying tail with tail index $\alpha(G) = \alpha(G_0)/D$ almost surely.*

Note, that we can control the tail index of G through the choices of $\alpha(G_0)$ and D, for instance, for larger parameter D the tighter the envelopes are around the tail index of G_0 and for smaller parameter D, the lighter the tail of G_0.
Next, we compare Theorems 8 and 9 in a parametric example.

Example 7 (Pareto centering distribution—DP *versus* SP)**.** Suppose first that $G \sim \mathrm{DP}(M, G_0)$, where G_0 is a standard Pareto distribution, that is, $1 - G_0(y) = 1/y$, for $y > 1$. Then (4.5) yields

$$\exp[-sy/M \log|\log(M/y)|] \leqslant 1 - G(y) \leqslant \exp[-y/\{M|\log(M/y)|^r\}], \tag{4.7}$$

for $s < 1$ and $r > 1$, and hence the tails of $G \sim \mathrm{DP}(M, G_0)$ are almost exponential, despite the fact that G_0 is heavy-tailed. Indeed, as shown in the supporting materials A, the two bounds in (4.7) are in the Gumbel maximum domain of attraction, although G_0 is in the Fréchet maximum

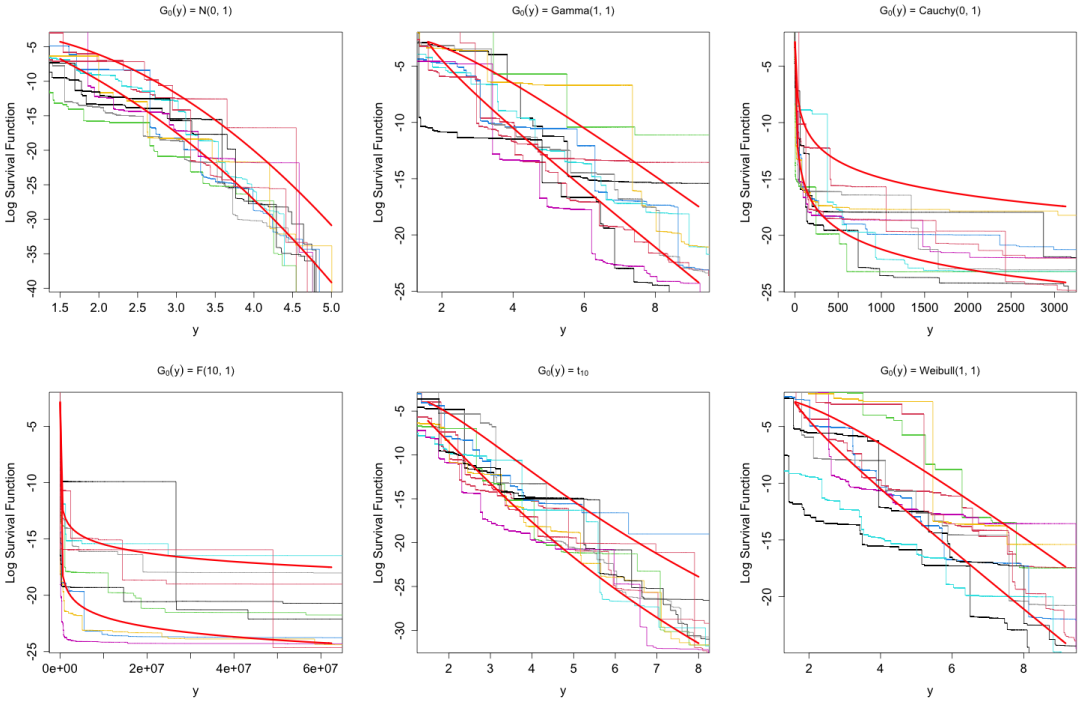


Figure 4.2: Asymptotic envelopes that follow from Theorem 9 in the main chapter along with the log survival function of 10 random trajectories of a $\mathrm{PYP}(0.4, 0, G_0)$.

domain of attraction. Suppose now that $G \sim \mathrm{SP}(D, G_0)$. Then,

$$y^{-1/D}\{\log|\log y^{-1}|\}^{1-1/D} \leqslant 1 - G(y) \leqslant y^{-1/D}|\log y^{-1}|^{r/D}.$$

It follows from Theorem 10 that $1 - G$ is regularly varying at infinity with tail index $1/D$. Figure 4.1 illustrates that, as predicted by Theorem 9, the trajectories of the stable process follow the asymptotic envelopes in (7); in addition, the same figure illustrates that the envelopes of the DP in (4.5) fall abruptly in comparison with those of the SP.

4.2.2 Heavy-Tailed Pitman–Yor Mixture Models

Empowered by the main findings from Section 4.2.1, this section shows that two classes of heavy-tailed Pitman–Yor mixture models can be devised, and it discusses the relative merits of each option. Below, we focus on univariate mixtures; comments on multivariate as well as conditional extensions are given in Section 4.3.

Heavy-tailed PYP Scale Mixtures: The first class of mixture models to be considered will be of the type,

$$f(y) = \int_0^\infty K_\sigma(y; \eta_\sigma)\,\mathrm{d}G(\sigma), \qquad G \sim \mathrm{PYP}(D, 0, G_0(\sigma)), \qquad 1 - G_0(\sigma) = \frac{\mathscr{L}(\sigma)}{\sigma^{\alpha_0}}, \tag{4.8}$$

with $y \in \mathbb{R}$ and $\alpha_0 \equiv \alpha(G_0) > 0$. Here, $K_\sigma(\cdot) = K(\cdot/\sigma; \eta_\sigma)/\sigma$ with K being a kernel, $\sigma > 0$ is a

scale parameter, η_σ denotes additional parameters (possibly related with σ), and $\mathscr{L}$ is a slowly varying function, that is, $\mathscr{L}(yt)/\mathscr{L}(y) \to 1$ as $y \to \infty$ for any $t > 0$.

Heavy-tailed PYP Shape Mixtures: The second class of mixture models to be considered is

$$f(y) = \int_0^\infty \Bbbk(y; \alpha, \eta_\alpha)\, \mathrm{d}G(\alpha), \qquad G \sim \mathrm{PYP}(D, M, G_0(\alpha)), \qquad 1 - \mathbb{K}(y; \alpha, \eta_\alpha) = \frac{\mathscr{L}(y)}{y^\alpha}, \quad (4.9)$$

with $y \in \mathbb{R}$ and $\alpha \geqslant 0$. Here, $\Bbbk$ is an Pareto-type kernel with distribution function $\mathbb{K}$, η_α denotes additional parameters (possibly related with α), and $\mathscr{L}$ is a slowly varying function.

The following theorem implies that PYP mixture models in (4.8) and (4.9) are indeed heavy-tailed, and in addition it shows how their tail indices relate with that of the centering. Here and below $(a)_+ = \max(a, 0)$ denotes the positive part function.

Theorem 11 (Heavy-tailed PYP Mixtures)**.** *The following results hold, for F the cdf of f:*

a) If (4.8) holds, with $U \sim K_\sigma(\cdot; \eta_\sigma)$, $\mathrm{E}(U_+^{\alpha_0}) < \infty$, $P(U_+ > \sigma) = o\{1 - G_0(\sigma)\}^{1/D}$ and $\liminf_{\sigma \to \infty} \mathscr{L}(\sigma) > 0$, then the tail of f is regularly varying with tail index $\alpha(F) = \alpha_0/D$ almost surely.

b) If (4.9) holds, then the tail of f is regularly varying with tail index $\alpha(F) = \inf\{\alpha : G_0(\alpha) > 0\}$, almost surely, for any $G \sim PYP(D, M, G_0(\alpha))$.

Some comments on Theorem 11 are in order:

- Theorem 11 a) shows that if the centering of the stable process is heavy-tailed, and if the tail of the kernel is not heavier than that of the mixing, then f in (4.8) is heavy-tailed—with the same tail index as that of the mixing; that is, $\alpha(F) = \alpha(G) = \alpha_0/D$ with probability one, where $\alpha_0 = \alpha(G_0)$. Interestingly, this result offers a partial answer to an insightful open problem raised by Li et al. (2019, Theorem 3.5) on the range of $\alpha(F)$ under a polynomial decay of G_0. Indeed, Theorem 11 a) illustrates that under a polynomial decay of G_0, it holds that $\alpha(F) = \alpha_0/D$ differs from 0 and ∞, almost surely, and thus $\alpha(F)$ for (4.8) does not concentrate on a singleton when assigned hyperpriors on D and α_0. The range of $\alpha(F)$ can be the positive real line, provided that $\mathsf{p}(D) > 0$ or $\mathsf{p}(\alpha_0) > 0$, for all $0 < D < 1$ and $\alpha_0 > 0$ with $\mathsf{p}(\alpha_0)$ and $\mathsf{p}(D)$ denoting the prior densities of α_0 and D, respectively.

- Theorem 11 b) is a folklore result that formalizes the idea that infinite shape mixtures of heavy-tailed kernels are themselves heavy-tailed, with tail index equal to that of the heaviest component (i.e., equal to the left endpoint of the centering in the case of (4.9)). As it is evident from the proof, the argument holds more generally for any random distribution G and not just for Pitman–Yor processes.

Table 4.1: Instances of Pareto-type kernels $\Bbbk$ following (4.9)

Distribution	**Kernel** ($\Bbbk$)	**Slowly varying function** ($\mathscr{L}$)	**Tail index** (α)
Burr	$\propto y^{c-1}/(1+y^c)^{a+1}$	$\propto (y^{-c}+1)^{-(a+1)}$	ca
F	$\propto y^{a/2-1}(a+by)^{-(a+b)/2}$	$\propto (a/y+b)^{-(a+b)/2}$	$b/2$
Generalized Pareto	$\propto (1-ay/\sigma)^{1/a-1}$	$\propto (1/y-k/\sigma)^{1/a-1}$	-1/a
Pareto	$\propto y^{-(a+1)}$	$\propto 1$	a
Student-t	$\propto (1+y^2/a)^{-(a+1)/2}$	$\propto (1/y^2+1/a)^{-(a+1)/2}$	a

For simplification, throughout we refer to (4.8) as scale mixtures but as it will be shown below (4.8) also includes scale–shape mixtures of light-tailed kernels. Next, we present instances of the specifications in (4.8) and (4.9), that showcase the generality of the latter and how they relate with some mainstream approaches.

Example 8 (PYP Scale Mixtures of an Erlang Kernel)**.** As an example of (4.8) consider

$$f(y) = \sum_{h=1}^{\infty} \pi_h \, \mathrm{Er}(y; a, \sigma_h), \tag{4.10}$$

where $\mathrm{Er}(y; a, \sigma_h)$ is the density of the Erlang distribution with shape $a \in \mathbb{N}$ and scale $\sigma_h > 0$, and $\pi_h = V_h \prod_{k<h}(1-V_k)$ with $V_h \overset{\text{ind}}{\sim} \mathrm{Beta}(1-D, M+hD)$, for $h \in \mathbb{N}$. In the notation of (4.8), with $K(y;\eta_\sigma) = \mathrm{Er}(y; a, 1)$ and $\eta_\sigma = a$. For this kernel, the infinite mixture in (4.8) shows a connection with the phase-type scale mixtures of Bladt and Rojas-Nandayapa (2018). Two key differences are that: *i*) since the mixing in (4.8) is over a Pitman–Yor process inference can be conducted by standard Bayesian nonparametric samplers, whereas fitting phase-type scale mixtures is far from straightforward; *ii*) Theorem 11 a) allows for a general kernel with a lighter tail than that of the mixing, whereas Bladt and Rojas-Nandayapa (2018) consider phase-type kernels. Another version of (4.8), along the same lines as (4.10), which we have found to perform well in practice is the following scale–shape mixture of PYP,

$$f(y) = \sum_{h=1}^{\infty} \pi_h \, \mathrm{Er}(y; \lceil \sigma_h \rceil, \sigma_h/\lambda), \tag{4.11}$$

where $\lceil \cdot \rceil$ is the ceiling function, and $K(y;\eta_\sigma) = \mathrm{Er}(y; \lceil \sigma \rceil, 1/\lambda)$ with $\eta_\sigma = (\lceil \sigma \rceil, 1/\lambda)^{\intercal}$ in the notation of (4.8). We will revisit (4.11) later in the chapter.

Example 9 (PYP Shape Mixtures of a Pareto-Type Kernel)**.** Since the Burr, F, and generalized Pareto distributions are particular cases of Pareto-type kernels (Table 4.1), the mixture model in (4.9) includes as particular cases infinite mixtures of such distributions with a PYP mixing. In addition, (4.9) also includes the Pareto kernel Dirichlet process mixtures of Tressou (2008).

There are some reasons for preferring the stable process scale mixtures in (4.8) over (4.9). In particular, scale mixtures of stable processes offer a more natural link between the tail of the

centering and the tail of F than (4.9). For example, if in (4.8) the centering is a Pareto Type II distribution over $(0,\infty)$ (i.e., $1-G_0(\sigma)=(1+\sigma)^{-\alpha_0}$) then $\alpha(F)=\alpha_0/D$, and hence both the centering and F are heavy-tailed. Yet, if the centering in (4.9) is a Pareto Type II distribution over $(0,\infty)$ (i.e., $1-G_0(\alpha)=(1+\alpha)^{-\beta}$), then $\alpha(F)=0$ and hence while the centering is heavy-tailed, F is super-heavy tailed. To avoid this behaviour, a truncated prior would be required so that G_0 is defined over $(\hat{\alpha},\infty)$ with an hyper-prior on $\hat{\alpha}$, however this can be computationally difficult. Motivated by this, in the next sections, we will emphasize scale mixtures of stable processes.

4.3 Extensions

4.3.1 Multivariate Variants

We now discuss how Section 4.2.2 can be extended to define priors on the space of multivariate heavy-tailed distributions, that is, the class of joint distributions with heavy-tailed marginals. Our construction follows the principles in Sarabia Alegría et al. (2008, Section 3.1), and it entails assuming conditional independence among components (margins) along with a common parameter shared by all components. This yields the following multivariate versions of (4.8) and (4.9) for $\mathbf{y}\in\mathbb{R}^d$. First, extending (4.8) consider the following multivariate heavy-tailed stable process scale mixture model:

$$f(\mathbf{y})=\int_{\mathbb{R}_+^d}\prod_{k=1}^{d}K_{\sigma_k}(y_k;\eta_{\sigma_k})\,\mathrm{d}G(\boldsymbol{\sigma}),\quad G\sim\mathrm{PYP}(D,0,G_0(\boldsymbol{\sigma})),\quad 1-G_{0,k}(\sigma)=\frac{\mathscr{L}_k(\sigma)}{\sigma^{\alpha_{0,k}}},\tag{4.12}$$

where $\boldsymbol{\sigma}=(\sigma_1,\ldots,\sigma_d)\in\mathbb{R}_+^d$, and $G_{0,k}(\sigma)$ is the kth marginal distribution of $G_0(\boldsymbol{\sigma})$, for $k=1,\ldots,d$. Second, extending (4.9), consider the following multivariate heavy-tailed PYP shape mixtures:

$$f(\mathbf{y})=\int_{\mathbb{R}_+^d}\prod_{k=1}^{d}\Bbbk(y_k;\alpha_k,\eta_{\alpha_k})\,\mathrm{d}G(\boldsymbol{\alpha}),\quad G\sim\mathrm{PYP}(D,M,G_0(\boldsymbol{\alpha})),\quad 1-\Bbbk(y;\alpha_k,\eta_{\alpha_k})=\frac{\mathscr{L}_k(y)}{y^{\alpha_k}},\tag{4.13}$$

where $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_d)\in\mathbb{R}_+^d$, $\mathscr{L}_k$ is a sequence of slowly varying functions, and $G_{0,k}(\alpha)$ will denote the kth marginal distribution of $G_0(\boldsymbol{\alpha})$, for $k=1,\ldots,d$.

Theorem 12 (Multivariate Heavy-tailed PYP Mixtures)**.** *The following results hold, for F_k the cdf of the kth-marginal distribution of f:*

a) If (4.12) holds, with marginal k such that, $U_k\sim K_{\sigma_k}(\cdot;\eta_{\sigma_k})$ satisfying $\mathrm{E}(U_{k+}^{\alpha_{0,k}})<\infty$, $P(U_{k+}>\sigma)=o\{1-G_0(\sigma)\}^{1/D}$ and $\liminf_{\sigma\to\infty}\mathscr{L}(\sigma)>0$, then the k-th marginal of f has a regularly varying tail with tail index $\alpha(F_k)=\alpha_{0,k}/D$, almost surely for $k=1,\ldots,d$.

b) If (4.13) holds, then the kth marginal of f has a regularly varying tail with tail index $\alpha(F_k)=\inf\{\alpha:G_{0,k}(\alpha)>0\}$, almost surely for any $G\sim PYP(D,M,G_0(\boldsymbol{\alpha}))$.

The previous result naturally extends Theorem 11 to the multivariate setting.

Remark 1. Some comments on the multivariate heavy-tailed stable process scale mixtures in (4.12) are in order:

- A concrete instance of (4.12) that we have found to work well in practice is the following extension of Example 8,

$$f(\mathbf{y}) = \sum_{h=1}^{\infty} \pi_h \prod_{k=1}^{d} \mathrm{Er}(y_k; \lceil \sigma_{k,h} \rceil, \sigma_{k,h}/\lambda), \tag{4.14}$$

  where once more $\mathrm{Er}(y; a, b)$ is the density of the univariate Erlang distribution and $\pi_h = V_h \prod_{k<h}(1 - V_k)$ with $V_h \overset{\text{ind}}{\sim} \mathrm{Beta}(1 - D, hD)$, for $h \in \mathbb{N}$.

- The specification in (4.12) requires that the centering is itself a multivariate heavy-tailed distribution. These can be easily constructed from copulas (Nelsen, 2006), and we recommend opting for Pareto Type II margins; that is, for $k = 1, \ldots, d$,

$$1 - G_{0,k}(\sigma) = \left(1 + \frac{\sigma}{\beta}\right)^{-\alpha_{0,k}}, \tag{4.15}$$

  with $\sigma \geqslant 0$, $\alpha_k > 0$, and $\beta > 0$. Such margins for the centering are convenient since they lead to a closed-form posterior of the extreme value index, as can be seen from the supporting materials A.

- To complete (4.12) we recommend a Jeffrey's prior on the tail index $\mathsf{p}(\alpha_{0,k}) \propto 1/\alpha_{0,k}$ and a Beta prior on the discount parameter, $D \sim \mathrm{Beta}(a_D, b_D)$. Finally, one may set a prior on the remainder parameters of the kernel, and for instance in the Erlang kernel example in (4.14), we will opt for $\lambda \sim \mathrm{Gamma}(a_\lambda, b_\lambda)$. This implies a prior for the tail index α_0/D, with infinite expectation. Alternatively, with a proper prior on α_0, e.g. $\alpha_0 \sim \Gamma(a_{\alpha_0}, b_{\alpha_0})$, the induced prior on the tail index would have prior expectation $E(\alpha_0)E(1/D) = \{a_{\alpha_0}(a_D + b_D - 1)\}/\{b_{\alpha_0}(a_D - 1)\}$.

4.3.2 Modelling Conditional Joint Densities

We now show how to extend the proposed models to include the effect of covariates by using a single atoms dependent Pitman–Yor process extending the principles from Barrientos et al. (2012, Definition 3) and Quintana et al. (2022, Section 2.3). For conciseness, we focus on multivariate heavy-tailed PYP scale mixtures in (4.12), but the principles discussed below can be easily adapted for the multivariate shape mixtures from Section 4.3.1 as well as to the univariate methods from Section 4.2. Consider the following predictor-dependent model,

$$f(\mathbf{y} \mid \mathbf{x}) = \int_{\mathbb{R}_+^d} \prod_{k=1}^{d} K_{\sigma_k}(y_k; \eta_{\sigma_k}) \, \mathrm{d}G_{\mathbf{x}}(\boldsymbol{\sigma}), \tag{4.16}$$

where $\mathbf{y}, \boldsymbol{\sigma} \in \mathbb{R}^d$ and $\{G_{\mathbf{x}}\}$ is a family of random probability measures indexed by a covariate $\mathbf{x} \in \mathbb{R}^p$. Specifically, we consider the following dependent stable process

$$G_{\mathbf{x}} = \sum_{h=1}^{\infty} \pi_h(\mathbf{x}) \delta_{\boldsymbol{\sigma}_h}, \qquad \boldsymbol{\sigma}_h \overset{\text{iid}}{\sim} G_0(\boldsymbol{\sigma}). \tag{4.17}$$

Here, the weights of the stick-breaking representation and the discount parameter D of the Pitman–Yor process are indexed over the covariate as follows, $\pi_h(\mathbf{x}) = V_h(\mathbf{x}) \prod_{k<h}\{1 - V_k(\mathbf{x})\}$, and

$$V_h(\mathbf{x}) \sim \text{Beta}(1 - D_h(\mathbf{x}), hD_h(\mathbf{x})), \qquad D_h(\mathbf{x}) = \frac{e^{\mathbf{x}^{\top}\boldsymbol{\beta}_h}}{1 + e^{\mathbf{x}^{\top}\boldsymbol{\beta}_h}}, \tag{4.18}$$

where $\boldsymbol{\beta}_h$ is a parameter in $\mathbb{R}^p$. Clearly, since (4.17) is a PYP for every $\mathbf{x}$, Theorem 12 a) implies that the joint density mixture model in (4.16) yields a multivariate heavy-tailed distribution, for every $\mathbf{x}$. The model is completed with a prior distribution for $\boldsymbol{\beta}_h$, given by $\boldsymbol{\beta}_h \overset{\text{iid}}{\sim} N_p(\mathbf{0}, s^2\boldsymbol{I})$, for $h \in \mathbb{N}$. A specific embodiment of the approach discussed in this section, that will be revisited later in the chapter, is the following extension of Example 8,

$$f(\mathbf{y} \mid \mathbf{x}) = \sum_{h=1}^{\infty} \pi_h(\mathbf{x}) \prod_{k=1}^{d} \text{Er}(y_k; \lceil \sigma_{k,h} \rceil, \sigma_{k,h}/\lambda), \tag{4.19}$$

where $\pi_h(\mathbf{x}) = V_h(\mathbf{x}) \prod_{k<h}\{1 - V_k(\mathbf{x})\}$, with $V_h(\mathbf{x}) \sim \text{Beta}(1 - D_h(\mathbf{x}), hD_h(\mathbf{x}))$, and $D_h(\mathbf{x}) = e^{\mathbf{x}^{\top}\boldsymbol{\beta}_h}/(1 + e^{\mathbf{x}^{\top}\boldsymbol{\beta}_h})$, for $h \in \mathbb{N}$.

4.4 Simulation Study

4.4.1 Simulation Scenarios and Preparations

This section describes the true data generating processes and the settings used over the Monte Carlo simulation study from Section 4.4.2.

Data Generating Processes: We consider one scenario for the univariate version of the model from Section 4.2.2, three scenarios for the multivariate version from Section 4.3.1, and three scenarios for the multivariate conditional version from Section 4.3.2. The univariate scenario is a standard unit Pareto distribution with tail function $1 - F(y) = 1/y$, for $y > 1$, and its main aim will be to highlight that for heavy-tailed data, stable process mixing leads to much better fits at the tails than Dirichlet process mixing. Beyond the univariate scenario, we also considered bivariate and conditional scenarios that contemplate different dependence levels and complexities of the marginals. Table 4.2 summarizes the bivariate and conditional scenarios, which are marginally characterized by

$$\begin{cases} f_k(y) = w\, f_{\text{LG}}\{y \mid a_1, b_1\} + (1 - w)\, f_{\text{LG}}\{y \mid a_2, b_2\}, \\ f_k(y \mid x) = w\, f_{\text{LG}}\{y \mid a_1(x), b_1(x)\} + (1 - w)\, f_{\text{LG}}\{y \mid a_2(x), b_2(x)\}, \end{cases} \tag{4.20}$$

Table 4.2: Bivariate and Conditional Simulation Scenarios. The marginals are mixtures of log-Gamma distributions as in (4.20), and dependence is set by a Gumbel copula with parameter θ

Scenario	**Marginal** (f_1)	**Marginal** (f_2)	**Copula** (θ)
Bivariate 1,2	$a_1 = b_1 = 5; w = 1$	$a_1 = b_1 = 5; w = 1$	3,1
3	$a_1 = 13; b_1 = 7;$ $a_2 = 10; b_2 = 8; w = .4$	$a_1 = 8; b_1 = 7;$ $a_2 = 15; b_2 = 8; w = .4$	1
Conditional 1,2	$a_1(x) = 1 + 4x; b_1(x) = 3; w = 1$	$a_1(x) = 1 + 4x; b_1(x) = 3; w = 1$	1,3
3	$a_1(x) = 11 + 5x; b_1(x) = 8 + 5x;$ $a_2(x) = b_2(x) = 7; w = .4$	$a_1(x) = 6 + 5x; b_1(x) = 12 + 5x;$ $a_2(x) = b_2(x) = 8; w = .4$	1

for $k = 1, 2$, where $f_{\text{LG}}(y; a, b) = b^a \log(y)^{a-1} y^{-(b+1)}/\Gamma(a)$ is the density of a log-Gamma distribution with shape $a > 0$ and rate $b > 0$; parenthetically, we note that the log-Gamma distribution is in the Fréchet domain of attraction with tail index b (Beirlant et al., 2004, Table 2.1).

The dependence is modeled via a Gumbel copula, so that data for the bivariate and conditional scenarios are respectively simulated from

$$\begin{cases} F(y_1, y_2) = C_\theta\{F_1(y_1), F_2(y_2)\}, \\ F(y_1, y_2 \mid x) = C_\theta\{F_1(y_1 \mid x), F_2(y_2 \mid x)\}. \end{cases} \tag{4.21}$$

Here, $C_\theta(u, v) = \exp[-\{(-\log u)^\theta + (-\log v)^\theta\}^{1/\theta}]$, for $(u, v) \in (0, 1)^2$, whereas $\theta \geqslant 1$ is the parameter controlling dependence, and F_1 and F_2 are the distribution functions of f_1 and f_2. For all scenarios, we have simulated $n = 1\,000$ observations, and for the conditional scenarios covariates were drawn from a standard uniform distribution.

MCMC and Model Specification: All models are develop using the slice sampler (Walker, 2007) available from the supporting materials A. We considered a burn-in period of 5 000 iterations, and after that scanned 5 000 samples from the posterior targets of interest. For the univariate scenario we fit a stable process scale mixture with an uninformative Gamma prior and an Erlang kernel (i.e., (4.11) with prior $\lambda \sim \text{Gamma}(0.1, 0.1)$); for the bivariate and conditional scenarios we fitted stable process scale mixtures with an Erlang kernel based on Remark 1 and Equation (4.19), respectively. For the latter, a Gumbel copula was used for the centering, and empirical Bayes was used to set the hyperparameter for θ via maximum likelihood. For the conditional version of the model in the regression parameters in (4.18), we consider the prior $\boldsymbol{\beta}_h \overset{\text{iid}}{\sim} N_p(\mathbf{0}, s^2\boldsymbol{I})$, and set the hyperparameter to be $s^2 = 100$. For the hyperparameters of the marginals Pareto for the base measure, we set $\beta_k = 1$ and tail index $\alpha_{0,k} = 2$, for $k = 1, 2$, which implies that both margins are apriori heavy-tailed but have a finite expected value. Finally, we have assigned the prior $D \sim \text{Beta}(0.5, 0.5)$ to the discount parameter for all instances of the

model. Keeping in mind space constraints, and the preference for stable process scale mixtures noted in Section 4.2.2, here we mainly concentrate on assessing the performance of the latter. The posterior inference algorithms available from the supplementary materials A can however be used for fitting multivariate as well as conditional heavy-tailed PYP shape mixtures, and some instances of the latter are available from the `pityoR` package.

One-Shot Experiments: One-shot experiments for the bivariate and conditional scenarios suggest that the proposed methods accurately recover the true distribution for all scenarios being examined. Such findings should of course be regarded as tentative, as they are the outcome of a single run experiment and will be subject to the scrutiny of the Monte Carlo simulation study in the next section.

4.4.2 Monte Carlo Simulation Study

We now report the main findings of a Monte Carlo simulation study. For each scenario from Section 4.4.1, we simulated 100 data sets each containing $n = 1\,000$ observations. All models have been fitted using stable process scale mixture models with the same specifications and MCMC settings as described in Section 4.4.1. We start with the univariate unit Pareto scenario, which will reinforce our preference for stable process scale mixture models.

In Figure 4.3 we present the posterior Monte Carlo means of the log-survival estimates for the tail of the univariate scenario and compare it with the corresponding Monte Carlo mean for a Dirichlet process mixture based on the same kernel. As can be seen from Figure 4.3, stable process mixing accurately estimates the tail, whereas Dirichlet process mixing markedly underestimates it; this numerical evidence showcases that the proposed stable process scale mixtures are a natural option for modelling risk and extremes in a heavy-tailed framework. Such numerical performance of the proposed methods finding is not surprising in light of Theorem 11 a); the performance over the bulk (not shown) is comparable for both forms of mixing. Interestingly, Figure 4.3 also reveals that the Monte Carlo mean based on fitting Dirichlet process mixture with a Pareto kernel and a Gamma centering distribution overestimates the tail of the distribution. Such numerical finding is not surprising, keeping in mind Theorem 11 b), given that the left endpoint of the Gamma centering distribution function is 0 and hence the resulting mixture is super heavy-tailed (i.e. $\alpha(F) = 0$). Next, we move to the bivariate and conditional scenarios from Section 4.4.1. Figure 4.4 shows 100 posterior estimated contours for the three bivariate scenarios. As it can be seen from the latter figure, the proposed stable process scale mixture model can capture the true contours over different levels of dependence (Scenarios 1–2) and even with challenging marginals such as mixtures (Scenario 3).

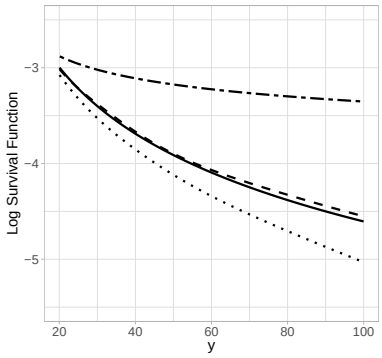


Figure 4.3: Mean of the Monte Carlo fits (dashed line) for the log-survival function for the univariate scenario obtained using the stable process scale mixture model from Section 4.2.2, from the 95% to the 99% quantile, plotted against the true (solid line). The dotted line shows the Monte Carlo mean of the fits from a DPM with the same Erlang kernel, whereas the dashed–dotted line shows the Monte Carlo mean of the fits from a DPM with a Pareto kernel and a Gamma centering distribution.

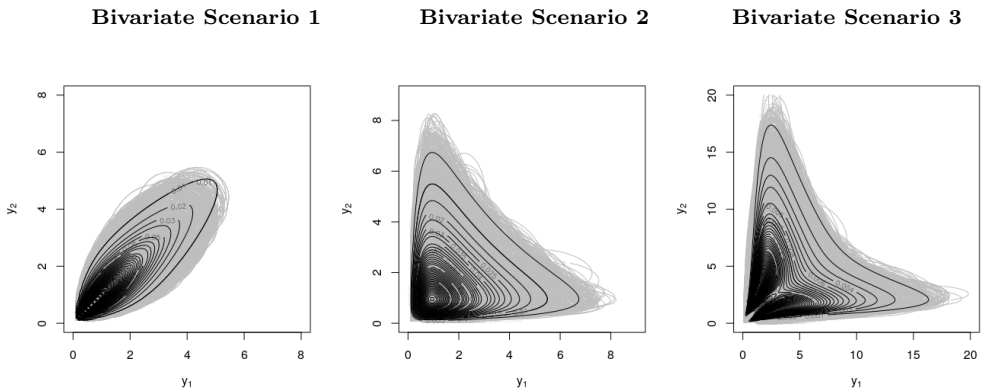


Figure 4.4: Monte Carlo Simulation for the Bivariate Scenarios 1–3: Contours of the joint density estimates (gray) obtained with proposed stable process scale mixture model from Section 4.3.1, for the 100 simulated data sets, plotted against the true (black).

4.4.3 Additional Numerical Results

This section provides additional numerical evidence. The one-shot experiment and Monte Carlo means of the fits for the bivariate and conditional scenarios from Table 2 in the chapter are reported in Figures 4.5–4.7. Overall, the estimates accurately recover the true joint densities.

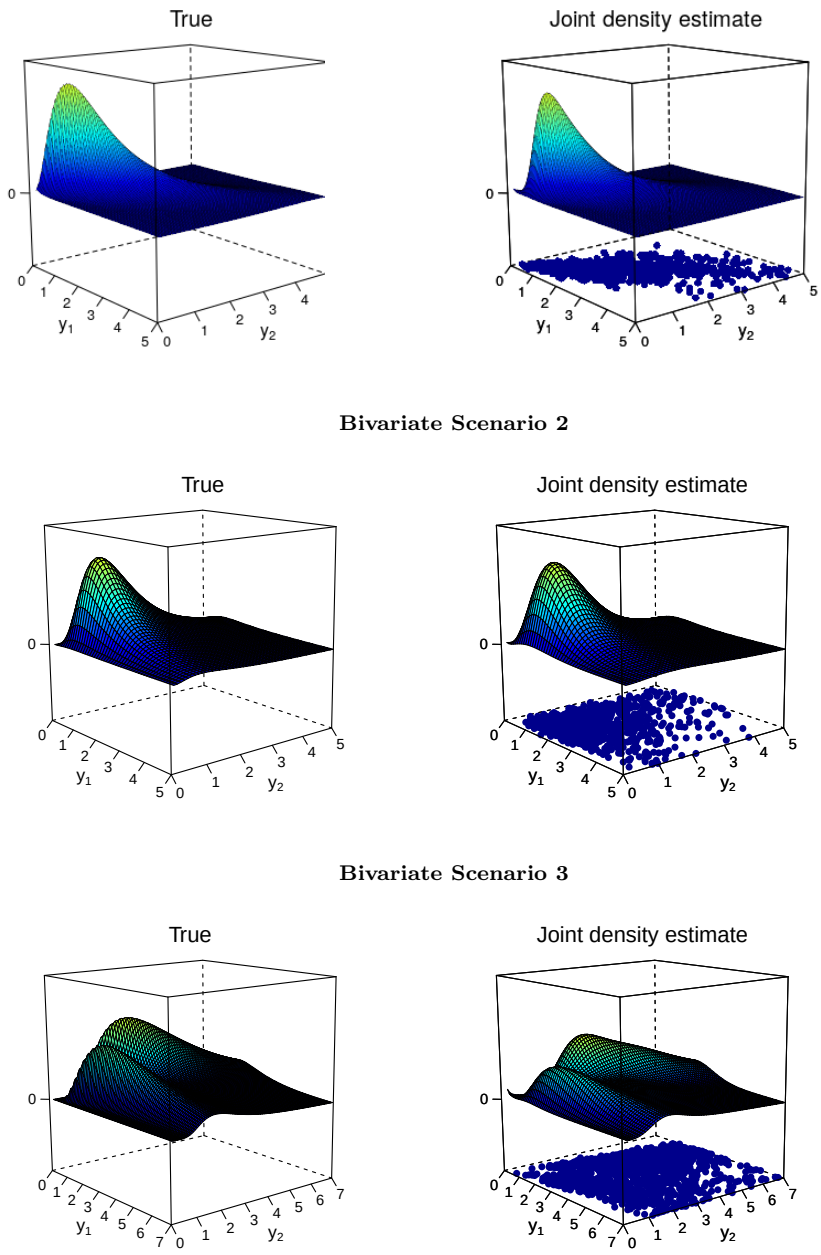


Figure 4.5: True (left) and estimated posterior densities (right) for Bivariate Scenarios 1–3 (top to bottom) obtained using the proposed stable process scale mixture model from Section Section 4.3.1 for a one-shot experiment. The simulated data for each scenario are overlaid at the bottom of the box.

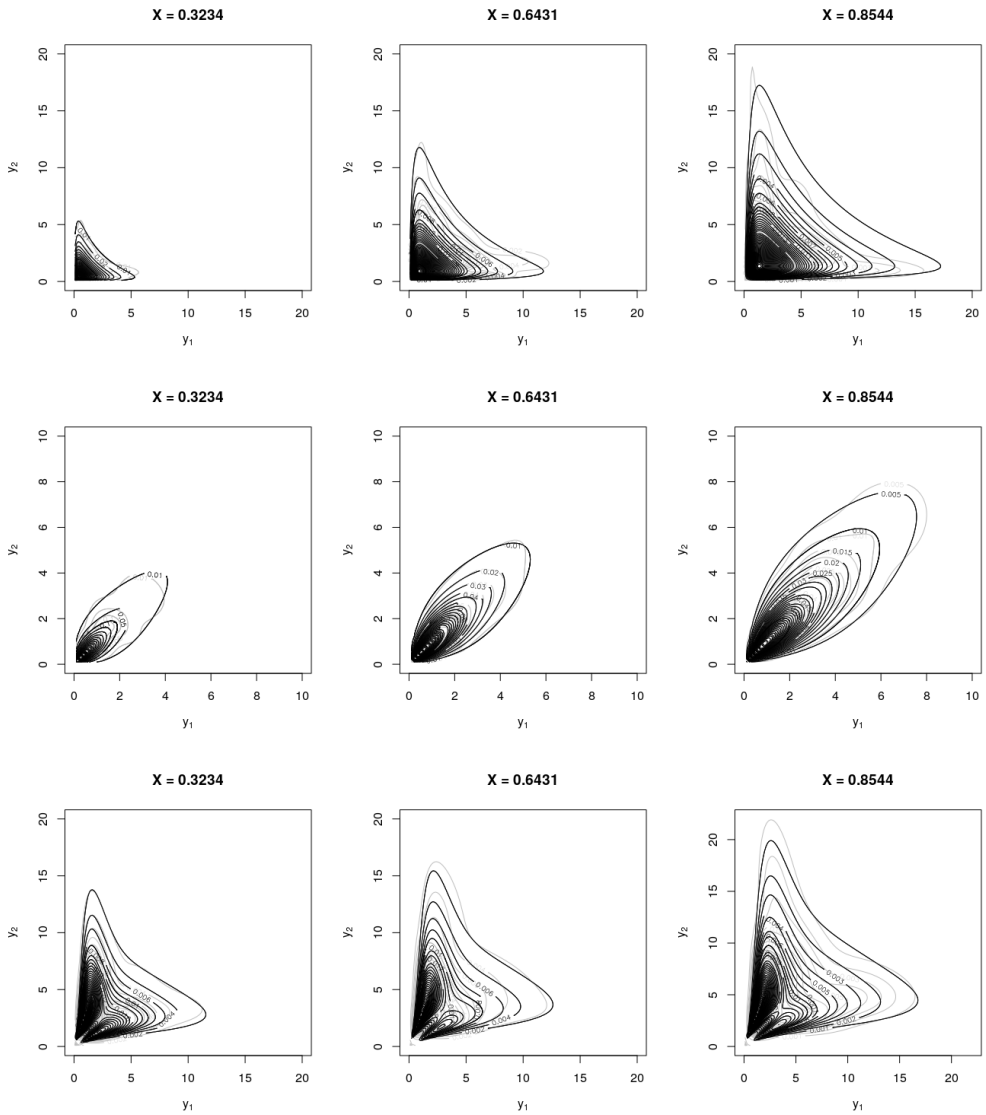


Figure 4.6: One-Shot Experiment for the Conditional Bivariate Scenarios 1–3: Contours of the conditional joint density estimates (gray)—given three levels of the covariate—obtained with proposed conditional stable process scale mixture model from Section 4.3.2 in the chapter, for a one-shot experiment, plotted against the true (black).

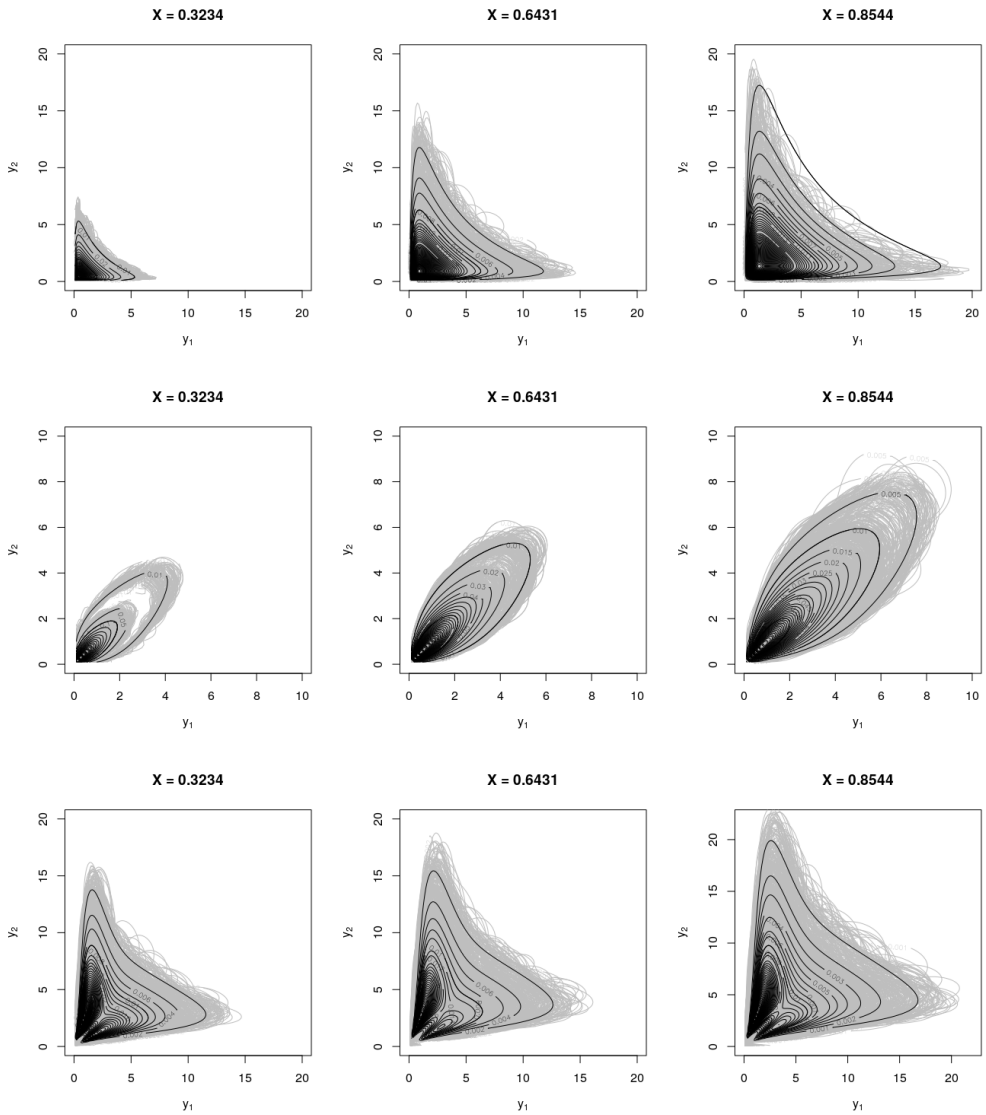


Figure 4.7: Monte Carlo Simulation for the Conditional Bivariate Scenarios 1–3: Contours of the conditional joint density estimates (gray)—given three levels of the covariate—obtained with proposed conditional stable process scale mixture model from Section 4.3.2 in the chapter, for the 100 simulated datasets, plotted against the true (black).

4.5 Application to Heavy-tailed Brain Data

4.5.1 Applied Context and Data Description

We now showcase the application of the proposed methods to a neuroscience case study. Brain rhythm signals are key for understanding how the human brain works; loosely speaking, they consist of patterns of neuronal activity that are believed to be linked with certain behaviors,

arousal intensity, and sleep states (Frank, 2009). These signals are typically measured using an electroencephalogram (EEG), which records electrical activity in the brain via electrodes attached to the scalp. An EEG signal tracks the activity of billions of neurons, and such signals cover a broad spectrum of frequency bands. Say, the alpha band typically refers to 8–13Hz, while beta refers to 13–20Hz; for a primer on brain rhythms and EEG signals, see for instance Buzsaki (2006) and Ombao et al. (2016, Ch. 7).

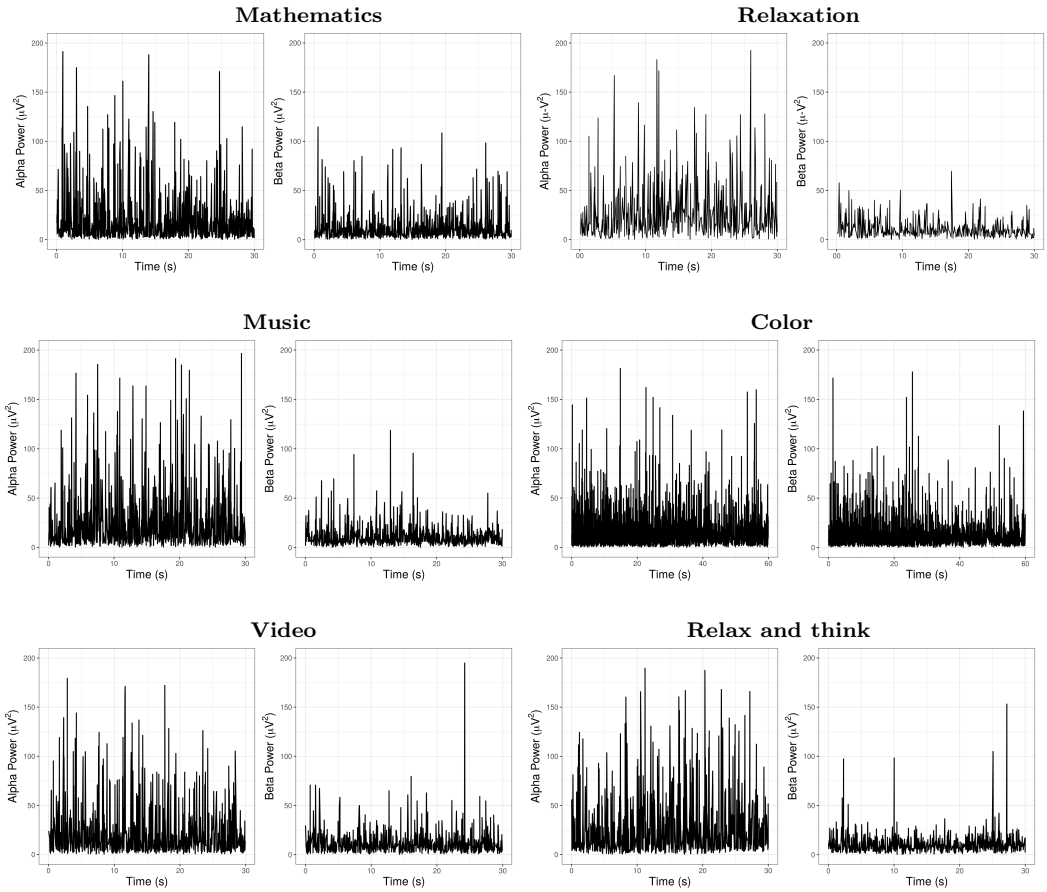


Figure 4.8: Spectral power (μV^2) of alpha and beta brainwave data plotted against time in seconds (s).

Alpha and beta rhythms are believed to be heavy-tailed (e.g. Roberts et al., 2015), and hence the main goal of our analysis will be to learn about the marginal and joint distribution of these heavy-tailed oscillations, given a variety of stimuli to be described below. In Section 4.5.4, we report evidence supporting the claim that in line with Roberts et al. (2015) our alpha and beta brainwave data are indeed heavy-tailed. We assess this by learning from data about the so-called extreme value index of a generalized Pareto distribution—which is known to be positive for heavy-tailed data (Coles, 2001, Section 4). The data to be analyzed are available from the R package `pityoR`, and were gathered from a UC Berkeley study that involved 30 participants who were subject to several audio-visual activities and stimuli, namely: *mathematics*, *relaxation*, *music*, *color*, *video* as well as *relax and think*. Figure 4.8 shows the spectral power in (micro)-

Volts squared (μV^2) for alpha and beta waves for all participants; roughly speaking, higher peaks indicate higher neural activity at a certain point in time on a frequency band of interest. The Ljung–Box test results reported in Section 4.5.4 suggest that the recorded trajectories of spectral power can be regarded as independent over time across all stimuli. Figure 4.8 illustrates some aspects of alpha and beta bands that help to build intuition on their signatures for the different stimuli. For instance, when the stimulus is *mathematics* we can notice high activity for both alpha and beta waves, similar behavior to watching a *video* or finding the *color*—all these being tasks that relate to immediate attention. And indeed, it has been suggested that alpha bands tend to be associated with 'attention' as well as with 'information processing' (Klimesch, 2012). Interestingly, the patterns of alpha and beta waves for both *mathematics* and *music* are reasonably similar—which might not be surprising in light of what has been claimed elsewhere (e.g., Boettcher et al., 1994, and references therein). Whether that similarity of *mathematics* and *music* also holds for the *joint* distribution of alpha and beta waves is something to be examined below in Section 4.5.3. Next, we learn from the heavy-tailed brainwave data discussed above using the methods proposed in Sections 4.2–4.3. Except where mentioned otherwise, all fits have been conducted using the same model specifications and MCMC settings as in Section 4.4.

4.5.2 Marginal Brainwave Analysis

Figure 4.10 shows the marginal density estimates of alpha and beta power pooling all subjects for each stimulus that were obtained using the proposed stable process scale mixture model from Section 4.2.2. Specifically, the fits from Figure 4.10 were obtained using the specification in (4.11) along with an uninformative Gamma prior and an Erlang kernel. To assess the quality of the obtained fits, we depict in Figure 4.10 q-q boxplots of random quantile residuals (Dunn and Smyth, 1996). The q-q boxplots introduced by Rodu and Kafadar (2022) contains the traditional information of a boxplot, it means, the sample median and quartiles. However, it also provides extra information about the tails. The interpretation is as follows, the middle box of the q-q boxplot, is the same as in the traditional boxplot, and the whiskers are modified to display the difference between the tail distribution of the data and the tail distribution of the reference. If the deviation is to the right, it would indicate a heavier-tail than the tail of the reference distribution and if the deviation is to the left, it would indicate a lighter-tail than the tail of the reference distribution. The obtained q-q boxplots provide evidence that the stable process scale mixture adjusts well both the bulk and the right tail of the data. We have also compared the fitted stable process scale mixture model against the DP shape mixtures in (4.9) with a Pareto kernel and a Gamma centering distribution. In line with the findings from the univariate scenario of the simulation study in Section 4.4.2, we again found evidence in favor of a far more sensible behavior of the stable process scale mixture in comparison with the shape mixture of heavy-tailed kernels. The comparison of the q-q boxplots in Figure 4.10—for

the stable process scale mixture—against those available on Section 4.5.4—for the DP shape mixture of Pareto kernels—clearly indicates a better performance of the former over the latter over both the left and right tails.

4.5.3 Stimulus-Specific Joint Brainwave Analysis

While Section 4.5.2 offered a one-dimensional snapshot across different stimuli, we now apply the proposed methods to learn about the joint distribution of the power of brainwaves on alpha and frequency bands, conditional on the activities and stimuli discussed in Section 4.5.1. To put it differently, we now apply the multivariate regression framework from Sections 4.3.1–4.3.2 so to learn about the conditional dependence structure governing alpha and beta rhythms, and to borrow strength across stimuli, rather than just fitting each density individually as in Section 4.5.2.

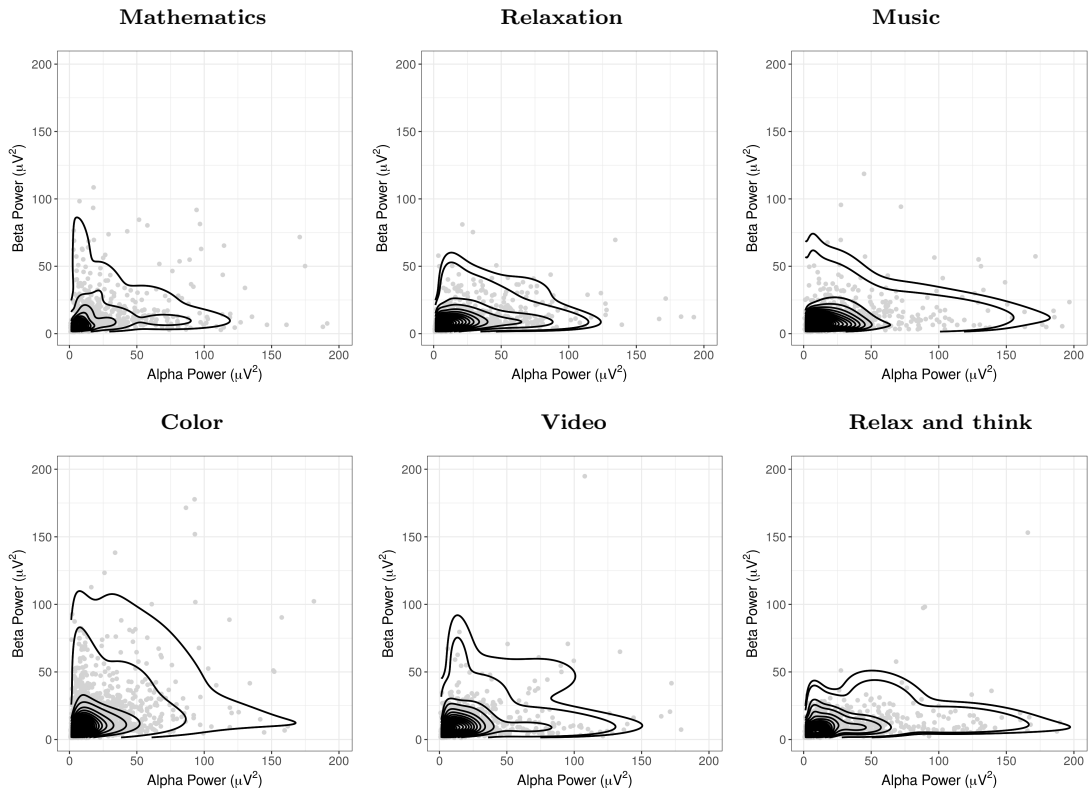


Figure 4.9: Contours of the posterior conditional joint density estimate of alpha and beta power for each specific stimulus along with raw data; the fit was obtained using the stable process scale mixture from Section 4.3.2.

Figure 4.9 shows the contours of the fitted conditional joint densities, given the stimulus under analysis, and it sheds light on the dynamics governing the joint behavior of the alpha and beta brain rhythms. First, the joint densities for some stimuli look similar—such as, for example, *music* and *relax and think*—which suggests a similar joint behavior of the rhythms of alpha and beta bands for these stimuli. Second, *mathematics* and *music*—which looked similar just by

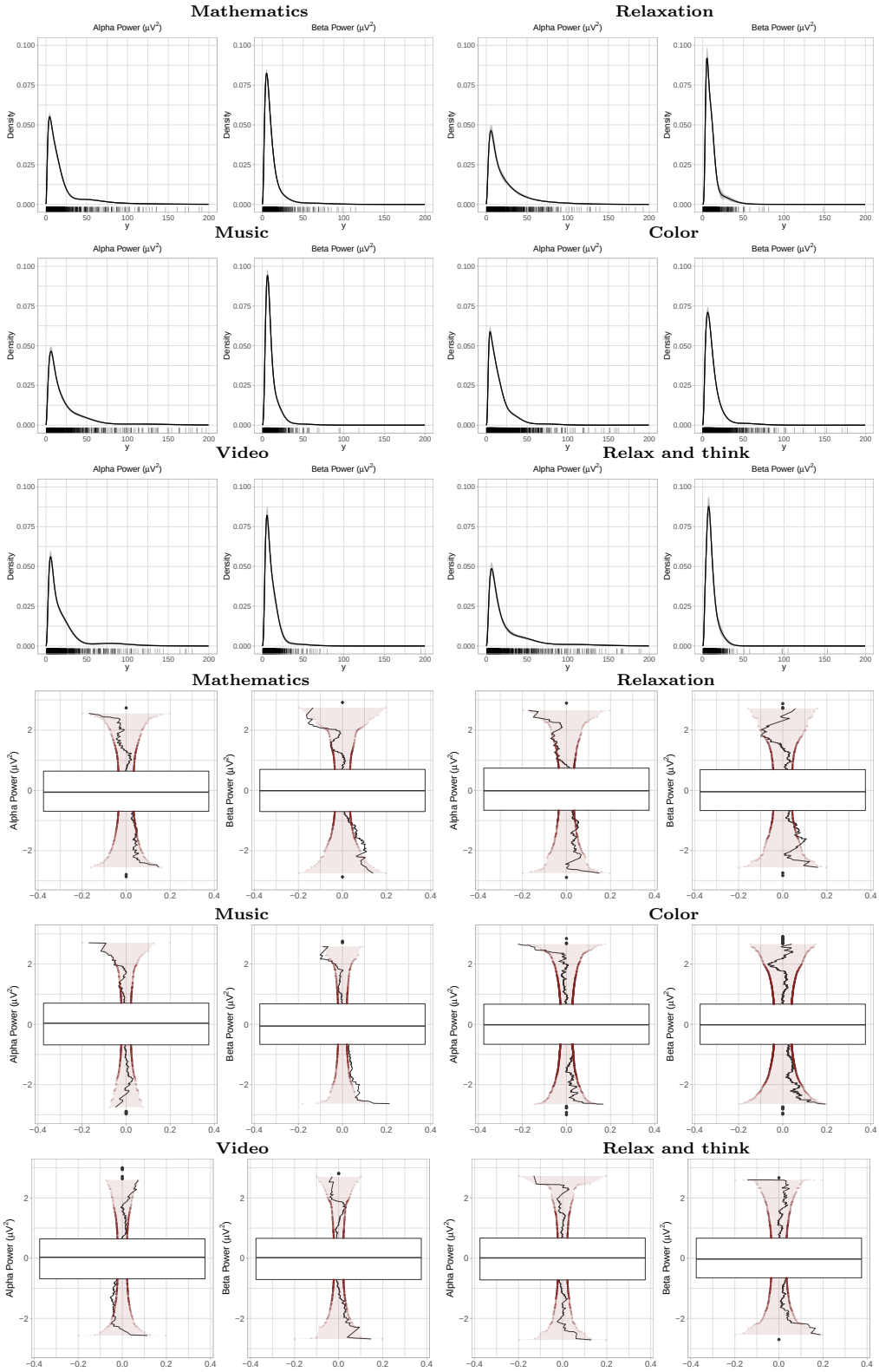


Figure 4.10: Top: Marginal density estimates of alpha and beta power for each stimulus, obtained with proposed stable process scale mixture model from Section 4.2.2, along with 95% credible bands. Bottom: Corresponding q-q boxplot of randomized quantile residuals.

examining the raw data in Figure 4.8 and the marginal fits in Figure 4.8—have a clearly different dependence structure as can be seen from Figure 4.9. In other words, while marginally the alpha- and beta-band oscillations for *mathematics* and *music* do look similar, their 'synchronization' or joint behavior looks markedly different.

4.5.4 Additional Empirical Results

This section reports additional empirical analyses. Table 4.3 reports posterior mean estimates of the extreme value index of the generalized Pareto distribution (Coles, 2001, Section 4), for each stimuli. All estimates are above zero, confirming that the brainwave data from Section 4.5 in the chapter are indeed heavy-tailed. Such estimates are fairly in line with the ones reported in Tables 4.4 and Table 4.5, which respectively contain the posterior mean EVI obtained using the unconditional and conditional stable process scale mixture models. Table 4.6 presents the p-values for the Ljung–Box test for independence; according to the Ljung–Box test, the observations for all the stimuli can be regarded as independent over time. Figure 4.11 depicts the qq-plots of the randomized quantile residuals Dunn and Smyth (1996), of the stable process scale mixture fitted in Section 4.2.2. Finally, Figure 4.12 shows the q-q boxplot (Rodu and Kafadar, 2022) of the randomized quantile residuals for the shape mixture model of a Pareto Type II kernel, which should be compared against Figure 5 in the chapter.

Stimulus	**Alpha Power**	**Beta Power**
Mathemathics	0.7938 (0.6819, 0.9015)	0.5167 (0.4355, 0.6034)
Relaxation	0.6140 (0.5083, 0.7226)	0.3306 (0.2581, 0.4078)
Music	0.7230 (0.6231, 0.8292)	0.3732 (0.3025, 0.4463)
Color	0.5160 (0.4551, 0.5977)	0.4138 (0.3598, 0.4705)
Video	0.6871 (0.5856, 0.7954)	0.3262 (0.2512, 0.4056)
Relax	0.7591 (0.6598, 0.8669)	0.2046 (0.1389, 0.2760)

Table 4.3: Posterior mean extreme value index of a generalized Pareto distribution along 95% with credible interval.

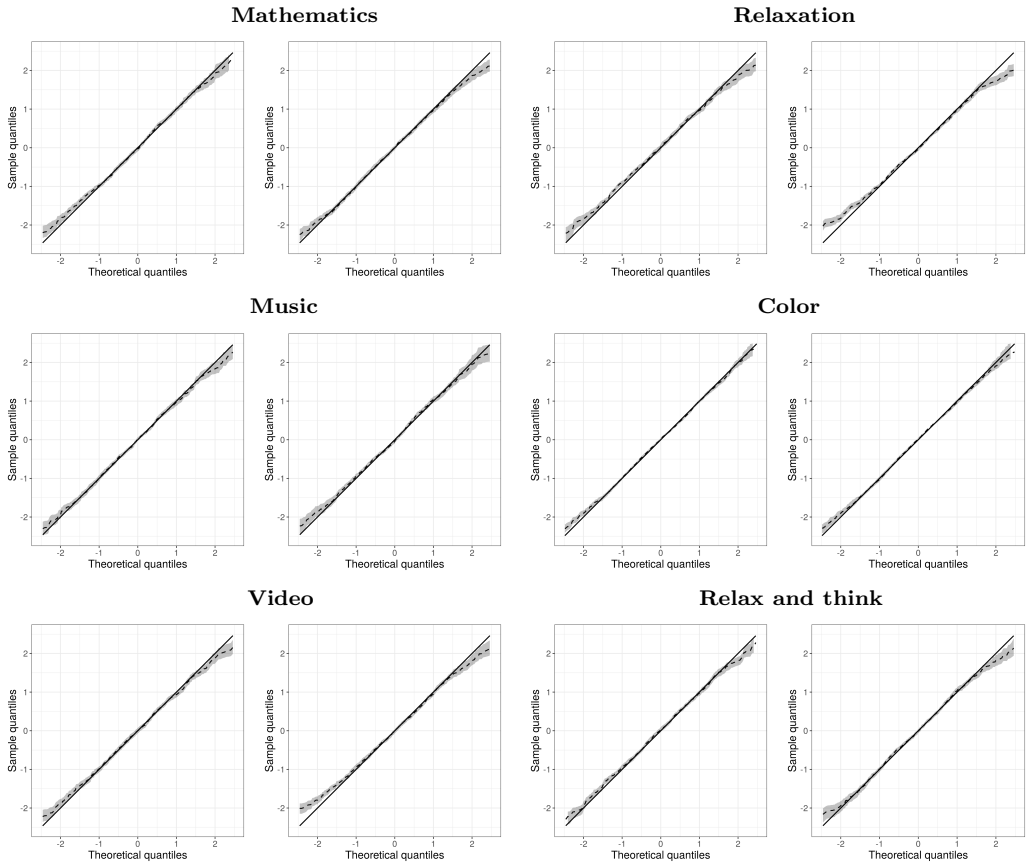


Figure 4.11: q-q plot of randomized quantile residuals for the proposed stable process scale mixture model from Section 4.2.2 of the chapter, along with credible bands, against the Normal theoretical quantiles.

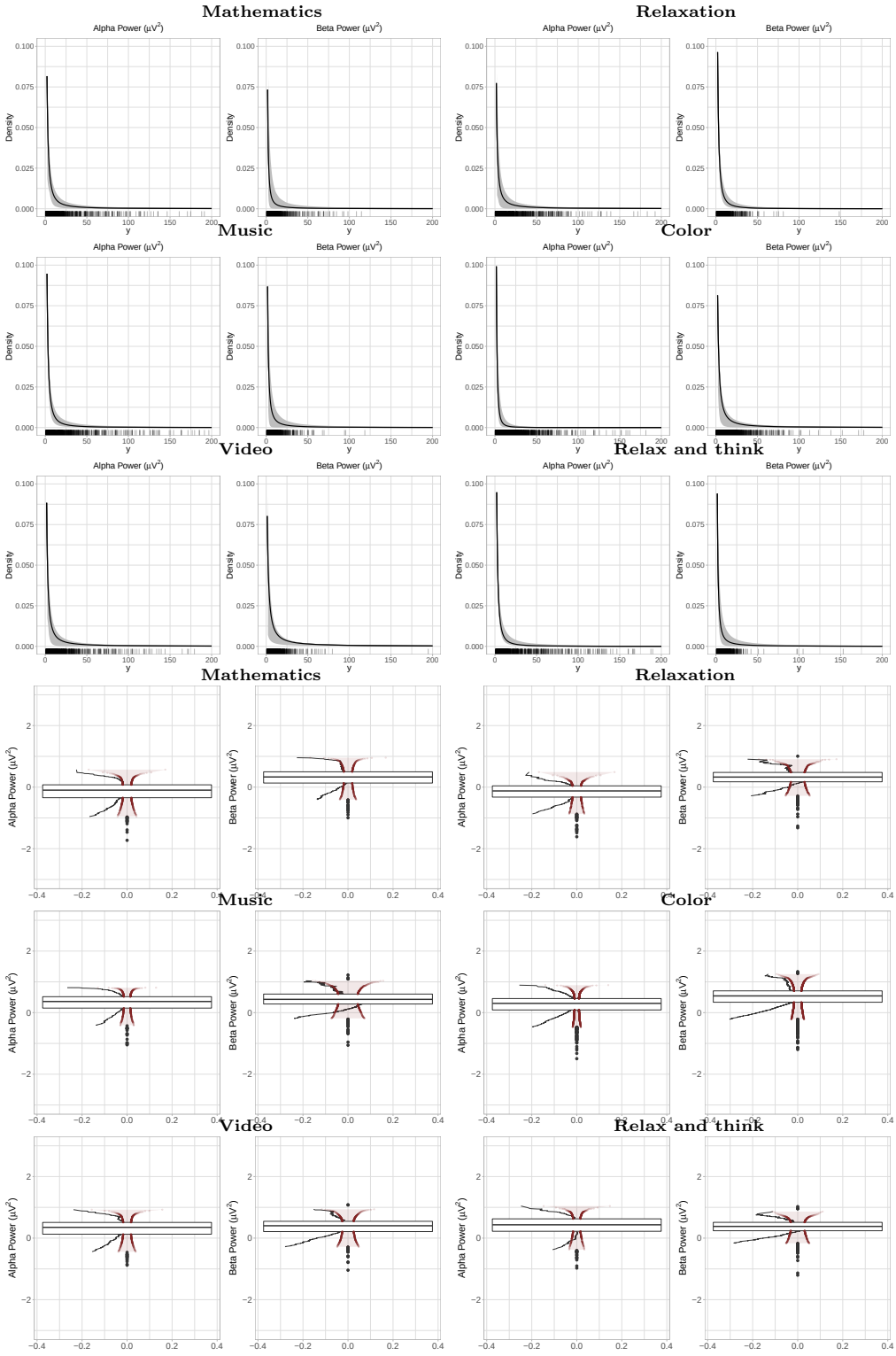


Figure 4.12: Top: Marginal density estimates of alpha and beta power for each stimulus, obtained with the shape mixture model from Section 4.2.2 of the chapter, along with 95% credible bands. Bottom: Corresponding q-q boxplot of randomized quantile residuals.

Stimulus	Alpha Power	Beta Power
Mathemathics	0.8234 (0.6951, 0.9613)	0.4892 (0.3961, 0.5942)
Relaxation	0.6076 (0.4899, 0.7297)	0.3306 (0.2422, 0.4216)
Music	0.7419 (0.6292, 0.8614)	0.4056 (0.3222, 0.4976)
Color	0.4813 (0.4040, 0.5607)	0.4276 (0.3631, 0.4970)
Video	0.6953 (0.5773, 0.8109)	0.3284 (0.2432, 0.4172)
Relax and think	0.7715 (0.6628, 0.8987)	0.2215 (0.1445, 0.3025)

Table 4.4: Posterior mean extreme value index under the unconditional stable process scale mixture model from Section 4.3.1 in the chapter along 95% with credible interval.

Stimulus	Alpha Power	Beta Power
Mathemathics	0.7498 (0.6708, 0.8541)	0.4820 (0.4218, 0.5679)
Relaxation	0.6068 (0.5174, 0.7017)	0.3224 (0.2581, 0.3862)
Music	0.7040 (0.6162, 0.7916)	0.3503 (0.2801, 0.4021)
Color	0.4826 (0.4388, 0.5269)	0.3822 (0.3434, 0.4219)
Video	0.6648 (0.5772, 0.7557)	0.3531 (0.2862, 0.4213)
Relax and think	0.7317 (0.6481, 0.8184)	0.2208 (0.1645 , 0.2782)

Table 4.5: Posterior mean extreme value index under the conditional stable process scale mixture model from Section 4.3.2 in the chapter along 95% with credible interval.

Stimulus	Ljung–Box p-value (Alpha Power)	Ljung–Box p-value (Beta Power)
Mathemathics	0.9362	0.3965
Relaxation	0.5891	0.5017
Music	0.0979	0.7899
Color	0.3651	0.3180
Video	0.2189	0.1446
Relax and think	0.0751	0.4035

Table 4.6: Ljung–Box test for all the stimuli.

Estimates of alpha power and beta power using model (12) are presented in Figure 4.13. We can see that the estimates using the conditional version of our model, model (14), are better, as it allows to borrow strength information between different predictor levels.

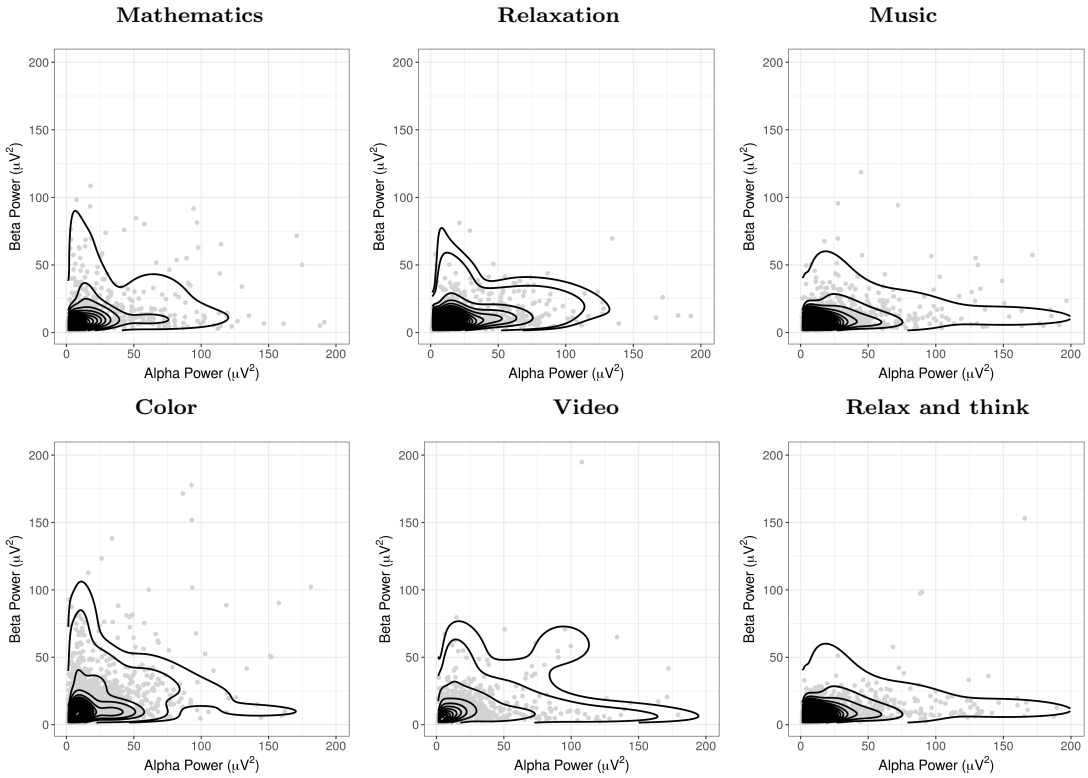


Figure 4.13: Contours of the posterior conditional joint density estimate of alpha and beta power for each specific stimulus along with raw data; the fit was obtained using the unconditional stable process scale mixture from Section Section 4.3.1.

4.6 Concluding Remarks

This chapter studied the tails of some prominent Pitman–Yor processes, and it has shown in particular that the tail of the stable law process is tantamount to that of the baseline. This result is in clear contrast to what is known to hold for the DP, whose tails are exponentially much thinner than those of the centering; in addition, we have also derived for the first time envelopes on which the tail of the stable process must lie. We then devised two classes of heavy-tailed Pitman–Yor mixture models, along with their extensions to a multivariate heavy-tailed setting—as well as to a regression framework. Equipped with the above-mentioned characterization of the tails of the PYP, we have shown that not all heavy-tailed Pitman–Yor process mixture models are alike. To put it differently, our theoretical and numerical analyses pinpoint a clear preference for stable process scale mixtures over shape mixtures of heavy-tailed kernels. Particularly, we have shown that shape mixtures of Pareto-type kernels can be super heavy-tailed even though the centering is 'only' heavy-tailed; this implies that a naïve application of the latter mixture models might lead to an overestimation of the mass at the tail—along with poor inferences at the bulk of the distribution. On the contrary, we have found stable process scale mixtures to obey natural properties—such as the stability of the heavy-tail from Theorem 10—and to perform well numerically in both the bulk as well as in the right tail. Keeping in mind the scope of our case study, we have concentrated the numerical illustrations on the right tail as well as on kernels supported over the positive real line. Yet, we underscore that the theory and methodologies from Sections 4.2–4.3 hold more generally over the entire real line as well as for left tails. Finally, the fact that other statistical functionals, such as tail indices, can be readily inferred from the proposed methods implies that, as a byproduct, the conditional version from Section 4.3 may be used as a tail index regression model in the same vein as Wang and Tsai (2009).

We close the chapter with some final comments on open challenges for future analysis. It seems reasonable to conjecture that the tail of the PYP for the case $M > -D$ with $0 < D < 1$ might suffer from similar issues like that of the DP; the analysis of the latter case is however as non-trivial as those presented herein, and hence we leave it as an open problem for future research. As we have shown here, stable law processes obey the stability of the heavy-tail property, and it would be interesting to have a broader understanding on how large is the class of random probability measures obeying that property. Finally, by keeping in mind the importance of modelling rare but catastrophic events in a variety of fields—such as climatology, geology, insurance, risk analysis, and extreme value theory—the methodologies proposed herein may pave the way for further applications and developments at the interface between heavy-tails and Bayesian nonparametrics.

4.7 Technical Details

4.7.1 Technical Details and Auxiliary Lemmata

In addition to the auxiliary results below, we first recall a basic fact on subordinators that will be handy for the proof of Theorem 9. If $S(t)$ is a subordinator, then $S(t+h)-S(t)$ has the same distribution as $S(h)$, for every $t, h \geqslant 0$ (Bertoin, 1999, p. 5). This implies that $\{S(1)-S(t)\}_{t\in[0,1]}$ is equal in distribution to $\{S(1-t)\}_{t\in[0,1]}$, and hence the following subordinator representation holds for the tail of $G \sim \mathrm{PYP}(\sigma, 0, G_0)$:

$$1-G(y) \underset{(2)}{=} 1-\frac{S\{G_0(y)\}}{S(1)} = \frac{S(1)-S\{G_0(y)\}}{S(1)} \overset{\mathrm{d}}{=} \frac{S\{1-G_0(y)\}}{S(1)}. \tag{4.22}$$

Lemma 1 gathers two well-known results on lower and upper envelopes of stochastic processes over the short-run which can be found in Bertoin (1999, Theorem 11) and Sato (1999, Proposition 47.16). Lemma 2 is a well-known result in regular variation Embrechts et al. (e.g. 1997, Theorem A.33) and for the extended Breiman's lemma see Denisov and Zwart (2007, Proposition 2.1).

Lemma 1. *The following results hold:*

a) If $\{S(t) : t \geqslant 0\}$ is a subordinator with Laplace exponent $\Phi \in RV_D$, with $D \in (0,1)$, then $\liminf_{t\to 0^+} |S(t)|/l(t) = D(1-D)^{(1-D)/D}$, a.s., where $l(t) = \log|\log t|/\Phi^{-1}(t^{-1}\log|\log t|)$ for $0 < t < e^{-1}$, and Φ^{-1} is the inverse function of Φ.

b) Let $\{S(t)\}$ be a D-stable process on $\mathbb{R}$ with $0 < D < 2$ and $\nu(0,\infty) > 0$. If $u(t)$ is a real-valued function that is positive, continuous, and increasing on some $(0,\delta)$, and for which $u(t)/\{t\log\log(1/t)\}^{1/2} \to 0$, as $t \to 0^+$, then

$$\limsup_{t\to 0_+} \frac{S(t)}{u(t)} = \begin{cases} 0, & \int_0^\delta u^{-D}(t)\,\mathrm{d}t < \infty, \\ \infty, & \int_0^\delta u^{-D}(t)\,\mathrm{d}t = \infty, \end{cases} \quad a.s.$$

Lemma 2 (Representation theorem)**.** *If $h \in RV_\alpha$ for some $\alpha \in \mathbb{R}$, then*

$$h(y) = c(y)\exp\left\{\int_z^y \frac{a(u)}{u}\,\mathrm{d}u\right\}, \quad y \geqslant z,$$

for some $z > 0$ with $c(y) \to c \in (0,\infty)$, $a(u) \to \alpha$ as $y \to \infty$. The converse also holds.

Lemma 3 (Extended Breiman's lemma)**.** *Let X and Y be random variables and suppose X has a regularly varying tail, $P(X > x) = x^{-\alpha}\mathscr{L}(x)$, with tail index $\alpha \geqslant 0$, and $Y \geqslant 0$ with $E(Y^\alpha) < \infty$. Then, if $\liminf_{x\to\infty}\mathscr{L}(x) > 0$ and $P(Y > x) = o\{P(X > x)\}$, it follows that XY has regularly varying tail with tail index α.*

4.7.2 Proofs of Main Results

Proof of Theorem 9: We start with the lower envelope. The Laplace exponent of a D stable process, $\Phi(\lambda) = \lambda^D$, is regularly varying at ∞ with index $D \in (0,1)$, and note also that $\Phi^{-1}(y) = y^{1/D}$. Hence, Lemma 1 a) implies that

$$\liminf_{t\to 0^+} \frac{S(t)}{l(t)} = D(1-D)^{(1-D)/D}, \quad \text{with } l(t) = t^{1/D}\{\log|\log t|\}^{1-1/D}. \tag{4.23}$$

Combining (4.23) with the representation of the stable law process in Example 6 yields

$$\liminf_{G_0(y)\to 0^+} \frac{S\{G_0(y)\}/S(1)}{l\{G_0(y)\}} = \liminf_{G_0(y)\to 0^+} \frac{G(y)}{l\{G_0(y)\}} = D(1-D)^{(1-D)/D}/S(1).$$

Hence, (4.22) yields

$$\liminf_{G_0(y)\to 1^-} \frac{1-G(y)}{u_r\{1-G_0(y)\}} = D(1-D)^{(1-D)/D}/S(1),$$

from where the final result follows. Next, we focus on the upper envelope. Consider the following family of functions for $r > 0$,

$$u_r(t) = t^{1/D}|\log t|^{r/D} = t^{1/D}\{\log(1/t)\}^{r/D}, \qquad t \in (0, e^{-r}), \tag{4.24}$$

where $D \in (0,1)$. We start by checking if $u_r(t)$ verifies the assumptions of Lemma 1 b). It follows that $u_r(t)$ is positive and nondecreasing on $(0,\delta)$, with $\delta = \mathrm{e}^{-r}$. Indeed,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\{u_r(t)\} &= D^{-1}t^{1/D-1}\{\log(1/t)\}^{r/D} + t^{1/D}rD^{-1}\{\log(1/t)\}^{r/D-1}(-1/t) \\ &= D^{-1}t^{1/D-1}[\{\log(1/t)\}^{r/D} - r\{\log(1/t)\}^{r/D-1}] > 0, \quad t \in (0,\delta).\end{aligned} \tag{4.25}$$

Additionally, it follows that

$$\lim_{t\to 0^+} \frac{u_r(t)}{\{t\log\log(1/t)\}^{1/2}} = \lim_{t\to 0^+} \frac{t^{1/D}\{\log(1/t)\}^{r/D}}{\{t\log\log(1/t)\}^{1/2}} = \lim_{t\to 0^+} t^{1/D-1/2} \times \lim_{t\to 0^+} \frac{\{\log(1/t)\}^{r/D}}{\{\log\log(1/t)\}^{1/2}} \to 0, \tag{4.26}$$

as $t \to 0^+$ given that $1/D - 1/2 > 0$ (recall that $D < 1$), and hence $u_r(t)$ obeys the assumptions of Lemma 1(2). Hence applying Lemma 1(2), with (4.24), in the representation of the stable law process from Example 6 yields

$$\limsup_{G_0(t)\to 0_+} \frac{G(t)}{u_r\{G_0(t)\}} = \begin{cases} 0, & r > 1, \\ \infty, & 0 < r \leqslant 1, \end{cases} \quad a.s. \tag{4.27}$$

Finally, Equation (4.22) implies the final result. □

Proof of Theorem 10: It follows from Theorem 9 that as $y \to y_+$, then

$$1 - G(y) = \{1 - G_0(y)\}^{\{1+o(1)\}/D}, \quad \text{a.s.} \tag{4.28}$$

By assumption, $1 - G_0 \in \text{RV}_{-\alpha_0}$ and hence it follows by the representation theorem that $1 - G_0(y) = c(y) \exp\{\int_z^y b(u)/u \, \mathrm{d}u\}$, for some $z > 0$ with $c(y) \to c \in (0, \infty)$, $a(u) \to -\alpha_0$ as $y \to \infty$. This combined with (4.28) yields that

$$1 - G(y) = \{1 - G_0(y)\}^{\{1+o(1)\}/D} = c^*(y) \exp\left\{ \int_z^y \frac{a^*(u)}{u} \right\}, \quad y \geqslant z,$$

for some $z > 0$, with

$$c^*(y) = \{c(y)\}^{\{1+o(1)\}/D} \to c^{1/D} \in (0, \infty), \quad a^*(u) = \left(\frac{1 + o(1)}{D} \right) b(u) \to -\alpha_0/D,$$

as $y \to \infty$. The final result follows from the representation theorem. □

Proof of Theorem 11

a) Let $U \mid V = \sigma \sim K(\,\cdot\,; \eta_\sigma)$ and $V \sim G$, and consider the decomposition $U = U_+ - U_-$, where $U_+ = \max(U, 0)$ and $U_- = \max(-U, 0)$. Since the focus is on the right tail, we concentrate on U_+, and note below that for $y > 0$ the scale mixture in (4.8) can be written as the density of the product of U_+ and V. In detail, it follows from Rohatgis well-known result on the product of random variables (e.g. Glen et al., 2004), that

$$f_{U_+V}(y) = \int_0^\infty f_{U_+,V}\left(\sigma, \frac{y}{\sigma}\right) \frac{1}{\sigma} \, \mathrm{d}\sigma. \tag{4.29}$$

Hence, combining the fact that $U_+ \mid V = \sigma \sim K(y; \eta_\sigma) I(y > 0) + P(U_+ = 0 \mid V = \sigma) I(y = 0)$ along with Bayes theorem implies that (4.29) can be rewritten as follows

$$f_{U_+V}(y) = \int_0^\infty f_{U_+|V}\left(\frac{y}{\sigma}, \sigma\right) \frac{\mathrm{d}G(\sigma)}{\mathrm{d}\sigma} \frac{1}{\sigma} \, \mathrm{d}\sigma = \int_0^\infty K\left(\frac{y}{\sigma}; \eta_\sigma\right) \frac{1}{\sigma} \, \mathrm{d}G(\sigma) = f(y), \tag{4.30}$$

for $y > 0$. Now, since by assumption G_0 has a regularly varying tail with tail index α_0, it follows from Theorem 10 that V has a regularly varying tail with tail index α_0/D. Now, let $V_0 \sim G_0$ and $\mathscr{L}^*(\sigma) = \{\mathscr{L}(\sigma)\}^{1/D}$ and note that the assumptions along with Theorem 10 and the representation theorem imply $E(U_+^{\alpha_0}) < \infty$, $P(V > \sigma) = \sigma^{-\alpha_0/D} \mathscr{L}^*(\sigma)$ and $P(U_+ > \sigma) = o\{P(V_0 > \sigma)\}^{1/D} = o\{P(V > \sigma)\}$ as well as that $\lim_{\sigma \to \infty} \mathscr{L}^*(\sigma) = \{\lim_{\sigma \to \infty} \mathscr{L}(\sigma)\}^{1/D} > 0$. In other words, the assumptions of the extended Breiman's lemma apply from where it readily follows that U_+V is regularly varying at infinity with tail index α_0/D, and hence the same claim can be made about $f_{U_+V}(y) = f(y)$. This proves the result.

□

b) Let $i(h)$ be a permutation such that $\inf\{\alpha : G_0(\alpha) > 0\} \equiv \alpha_{i(1)} \leqslant \alpha_{i(2)} \leqslant \cdots$. Then,

$$1 - F(y) = \sum_{h=1}^{\infty} \pi_h \frac{\mathscr{L}(y)}{y^{\alpha_h}} = \sum_{j=1}^{\infty} \frac{\pi_{i(j)} \mathscr{L}(y)}{y^{\alpha_{i(j)}}} = \frac{\mathscr{L}^*(y)}{y^{\alpha_{i(1)}}},$$

and it can be easily shown that $\mathscr{L}^*(y) = \mathscr{L}(y)\{\pi_{i(1)} + \sum_{j=2}^{\infty} \pi_{i(j)}/(y^{\alpha_{i(j)} - \alpha_{i(1)}})\}$ is a slowly varying function, from where the final result follows. □

Proof of Theorem 12: We only present the proof of Theorem 12 a) as that of claim b) follows a similar line of attack. We start by showing that the marginal distributions F_k are univariate Pitman–Yor mixtures, and then using Theorem 11 it follows that their tails, $1 - F_k$, are regularly varying, for $k = 1, \ldots, d$. Let $\mathrm{d}\mathbf{y}_{-k} = \mathrm{d}y_1 \ldots \mathrm{d}y_{k-1}\mathrm{d}y_{k+1} \ldots \mathrm{d}y_d$ and note that (4.12) combined with Fubini's theorem implies that

$$\begin{aligned}
f_k(y_k) = \int_{\mathbb{R}_+^{d-1}} f(\mathbf{y})\, \mathrm{d}\mathbf{y}_{-k} &= \int_{\mathbb{R}_+^{d-1}} \left\{ \int_{\mathbb{R}_+^d} \prod_{j=1}^{d} K_{\sigma_j}(y_j; \eta_{\sigma_j})\, \mathrm{d}G(\boldsymbol{\sigma}) \right\} \mathrm{d}\mathbf{y}_{-k} \\
&= \int_{\mathbb{R}_+^d} \left\{ \int_{\mathbb{R}_+^{d-1}} \prod_{j=1}^{d} K_{\sigma_j}(y_j; \eta_{\sigma_j}) \mathrm{d}\mathbf{y}_{-k} \right\} \mathrm{d}G(\boldsymbol{\sigma}) \\
&= \int_{\mathbb{R}_+^d} K_{\sigma_k}(y_k; \eta_{\sigma_k}) \left\{ \prod_{j=1, j\neq k}^{d} \int_0^{\infty} K_{\sigma_j}(y_j; \eta_{\sigma_j}) \mathrm{d}y_j \right\} \mathrm{d}G(\boldsymbol{\sigma}) \\
&= \int_{\mathbb{R}_+^d} K_{\sigma_k}(y_k; \eta_{\sigma_k})\, \mathrm{d}G(\boldsymbol{\sigma}) = \int_0^{\infty} K_{\sigma_k}(y_k; \eta_{\sigma_k})\, \mathrm{d}G_k(\sigma_k).
\end{aligned}$$

Since by assumption $G_{0,k}(\sigma_k)$ has a regularly varying tail with tail index $\alpha_{0,k}$, it follows from Theorem 11 a) that $1 - F_k$ is regularly varying with tail index $\alpha(F_k) = \alpha_{0,k}/D$, for $k = 1, \ldots, d$, from where the final result follows. □

Chapter 5

Interfacing Nonstationary Multivariate Extremes and Heteroscedastic Extremes

We devise a framework for learning about the frequency and magnitude of extreme values in the joint tail as well as in other related regions induced by aggregating variables on a common scale. The proposed framework reveals explicit links between nonstationary multivariate extremes and heteroscedastic extremes, and hence can be used to track the dynamics governing the degree of association between the extremes of a random vector over time. Another contribution of this chapter rests on development of Bayesian semiparametric inference methods for the two targets of interest: the structure scedasis function and the structure coefficient of tail dependence. To learn about the structure scedasis function from data we resort to a Bayesian nonparametric approach that defines a prior in the space of structure scedasis functions. For aggregation rules for which the structure coefficient of tail dependence is shown to be non-degenerate, we learn about it from data using a prior supported over its parameter space. The simulation study shows that the proposed methods are able to recover the true frequency and magnitude of extremes in a failure region defined by different aggregation rules. An application of the proposed methodology to the FAANG stocks (Facebook, Apple, Amazon, Netflix, and Alphabets Google) sheds light on some interesting features on the dynamics governing their extreme joint losses over time. This chapter is based on a paper of the same name, under review for The Electronic Journal of Statistics, with co-author Miguel de Carvalho. The methodology is implemented and added to the R package `extremis`*, available in CRAN, with creator Miguel de Carvalho and Vianey Palacios as co-author.*

5.1 Introduction

The widely felt impact of record-breaking extreme events—such as stock market crashes, earthquakes, heatwaves, or widespread flooding—calls for an urgent need for a better understanding and quantification of their risk. Whereas classical statistics is mostly concerned about inferences surrounding the bulk of a distribution, statistics of extremes (Beirlant et al., 2004; Coles, 2001; Davison and Huser, 2015; de Haan and Ferreira, 2006) deals with the rather challenging situation of conducting inference at the tails. In a multivariate context, the degree of association between the extreme values of a random vector with common margins, (X, Y), is often evaluated by,

$$\chi = \lim_{z \to \infty} P(X > z \mid Y > z). \tag{5.1}$$

The measure χ in (5.1) quantifies the probability of X being extreme, given that Y is extreme. If $\chi = 0$ the variables are said to be asymptotically independent, whereas if $0 < \chi \leqslant 1$ the variables are asymptotically dependent.

Given the time-evolving nature of multivariate extreme values, a recent body of literature has been devoted to nonstationary multivariate extremes (Castro et al., 2018; de Carvalho, 2016b; Escobar-Bach et al., 2018; Gong and Huser, 2019; Mhalla et al., 2017, 2019). One of the main targets of this recent field is to track how the dependence between the extreme values evolves over time—or according to another covariate—and thus to shed light on the dynamics governing extremal dependence. This has been achieved by indexing the parameter of the so-called multivariate extreme value distribution over time and then to learn about it from data.

Assume we want to model the extreme joint losses of a set of Big Tech stocks known as FAANG (Facebook, Apple, Amazon, Netflix and Alphabet's Google). These stocks trade on the NASDAQ stock market and have attracted both personal investors and some prominent money managers and professional investors. Then, if the magnitude of extremes is changing over time, we can have an extreme loss—leading to a market crash. However, another possible outcome is that the frequency of extremes is changing over time, in this later case we can see a constant period of losses—leading to a market crash. For this reason, the type of nonstationarity we are interested in, is the one arising from changes in the frequency of extremes rather than the magnitude. Note that this nonstationary compares only the distribution tails and does not impose any assumption on the central parts of the distributions.

In this chapter we contribute to the recent literatures on nonstationary multivariate extremes (references above) and on heteroscedastic extremes (de Haan, 2015; Einmahl et al., 2016, 2021; Mefleh et al., 2020) as well as to that on Bayesian modelling of multivariate extremes (Dombry et al., 2017; Guillotte et al., 2011; Hanson et al., 2017; Sabourin and Naveau, 2014). The focus of statistical methods for heteroscedastic extremes lies in modelling univariate extremes

of independent but not necessarily identically distributed data. A main contribution of this chapter will be to shed light on links between the nonstationary multivariate extremes and heteroscedasdic extremes.

First, we observe in Section 5.2 that by definition the scedasis function of the joint tail can be regarded as a standardized time-varying version of the coefficient of extremal dependence in (5.1). As we show below (Sections 5.2–5.3), this implies that the scedasis based on the aggregation rule $\phi_m(x, y) = \min(x, y)$ can be regarded as a measure of extremal dependence over time. Second, while examining a general class of aggregation rules for extreme values, we show that a constant scedasis corresponds to the case of stationary multivariate extremes, whereas peaks in a structure scedasis correspond to an increase in the mass of the corresponding risk set; see Proposition 6. A consequence of the latter result is that a constant structure scedasis corresponds to a stationary extremal dependence structure, whereas a peak in a structure scedasis indicates a higher (or lower) degree of association between the extremes of the basis process over that period (depending on the aggregation rule). The coefficient of tail dependence for different aggregation rules is also examined here as well as the relation between the coefficients underlying such rules.

Another contribution of this chapter rests on Bayesian inference for the structure scedasis functions via finite mixtures of Polya trees (MPT). Polya trees are one of most well-known priors on spaces of functions in the fast-evolving field of Bayesian nonparametrics (Ghosal and Van der Vaart, 2015), and the theory on MPT was introduced and studied by Christensen et al. (2008); Hanson and Johnson (2002); Hanson (2006); Lavine et al. (1992, 1994), among others. The proposed inferences for the structure scedasis function are fully supported on the unit interval, do not suffer from boundary bias—as do, for example, kernel-based approaches (Wand and Jones, 1995, Section 5.5)—nor entail the challenging choice of a smoothness parameter. Finally, for aggregation rules for which the structure coefficient of tail dependence is shown to be non-degenerated, we learn about it from data using a prior supported over its parameter space.

The masterplan of this chapter is as follows. Section 5.2 puts forward the first link between multivariate nonstationary extremes and heteroscedastic extremes, with a focus on the joint tail and on the bivariate setting. Section 5.3 moves beyond the joint tail and the bivariate setting, and it examines the interface between nonstationary multivariate extremes and heteroscedastic extremes on a more general framework. Bayesian inference is discussed in Section 5.4. Section 5.5 reports the main findings of a Monte Carlo simulation study, and Section 5.6 showcases an application of the proposed methods to some leading US tech-stocks. Finally, Section 5.8 discusses the main results and concludes the chapter. The Supporting Materials B contain further numerical outputs as well as notes on the R package *extremis* which can be used to implement the proposed methods.

5.2 Linking χ and the Scedasis

5.2.1 Starting Point

We open the chapter with a simple yet enlightening link between the scedasis function for the joint tail and a time-varying version of (5.1). Let $\{(X_t, Y_t)\}_{t=1}^{T}$ be a sequence of independent random vectors, and following standard practice in multivariate extreme value theory suppose that $\{X_t\}$ and $\{Y_t\}$ are unit Fréchet distributed, i.e., $P(X_t < z) = P(Y_t < z) = \exp(-1/z)$, with $z > 0$ for all t. Let $Z_t = \min(X_t, Y_t)$, with survivor function

$$1 - F_t(z) = P(Z_t > z) = P(X_t > z, Y_t > z)$$

for all t. The *structure scedasis for the joint tail* is defined as the scedasis of Z_t, that is,

$$c^*\left(\frac{t}{T}\right) = \lim_{z\to\infty} \frac{1 - F_t(z)}{1 - F(z)}, \tag{5.2}$$

for some unknown baseline distribution function $F(z)$ with an infinite upper endpoint. It follows directly from (5.2) that

$$c^*\left(\frac{t}{T}\right) = \lim_{z\to\infty} \frac{1 - F_t(z)}{1 - F(z)} = \lim_{z\to\infty} \frac{P(X_t > z, Y_t > z)}{F(z)} \equiv \chi_t. \tag{5.3}$$

In other words, Equation (5.3) reveals the first link between nonstationary multivariate extremes and heteroscedastic extremes. Specifically, (5.3) implies that the structure scedasis of the joint tail is a time-varying version of the tail-dependence coefficient in (5.1), and hence it can be understood as a measure of nonstationary extremal dependence. Whenever $\int_0^1 c^*(s)\,\mathrm{d}s > 0$, a standardized version of the structure scedasis can be defined as

$$c\,(s) = \frac{c^*\,(s)}{\int_0^1 c^*\,(s)\;\mathrm{d}s}, \tag{5.4}$$

for $s \in [0, 1]$. The structure scedasis function c carries information on the dynamics of the frequency of extremes in the joint tail over time. A uniform structure scedasis $c(s) = 1$ indicates a constant frequency of joint extremes over time, whereas if the structure scedasis peaks at some period, it provides an indication of an higher frequency of joint extremes at that period. In other words, a uniform structure scedasis implies a pattern of stationary bivariate extremes, whereas a peak in the scedasis implies an increase in the level of extremal dependence at that period.

Under the assumption that F is in the domain of attraction of a generalized extreme value (GEV) distribution with extreme value index $0 < \gamma \leqslant 1$, and under (5.2) it follows from de

Haan and Ferreira (2006, Theorem 1.2.1), that for all $z > 0$,

$$\lim_{w\to\infty} \frac{1 - F_t(wz)}{1 - F_t(w)} = z^{-1/\gamma}. \tag{5.5}$$

Thus, all F_t have the same extreme value index as F. Following Ledford and Tawn (1996), we will refer to γ in (5.5) as the *coefficient of tail dependence*.

5.2.2 Examples based on Time-varying Extreme Value Copulas

Next, we illustrate some instances of structure scedasis functions for the joint tail based on time-varying extreme value copulas (Castro et al., 2018; de Carvalho, 2016b). Recall that the time-varying bivariate extreme value copula is defined as

$$\mathbb{C}_t(u, v) = G_t\left(\frac{-1}{\log(1-u)}, \frac{-1}{\log(1-v)}\right), \quad (u, v) \in [0, 1]^2.$$

Here, $G_t(x, y)$ is a time-varying bivariate extreme value distribution. That is,

$$G_t(x, y) = \exp\left\{-\ell_t\left(\frac{1}{x}, \frac{1}{y}\right)\right\},$$

for $x, y > 0$, where the time-varying tail dependence function is

$$\ell_t(x, y) = \int_0^1 2\max\{wx, (1-w)y\}\, H_t(\mathrm{d}w), \tag{5.6}$$

and H_t is a time-varying angular measure that obeys the following moment constraint for all t,

$$\int_0^1 w\, H_t(\mathrm{d}w) = \frac{1}{2}.$$

Some parametric instances will be considered below so to illustrate the relation between the structure scedasis and χ_t given by (5.3) as well as to shed light on the interpretation of the structure scedasis for the joint tail. For Examples 10–12 below, we suppose that the baseline is unit Fréchet distributed; we focus on the structure scedasis density as the coefficient of tail dependence of all models below is $\gamma = 1$; for details see Heffernan and Tawn (2004).

Example 10 (Logistic)**.** The tail dependence function for the time-varying logistic extreme value copula is

$$\ell_t(x, y) = (x^{1/\alpha_t} + y^{1/\alpha_t})^{\alpha_t},$$

for $x, y > 0$, where $0 < \alpha_t \leqslant 1$. The true structure scedasis function under this model is the standardized version of

$$c^*(t/T) = 2 - 2^{\alpha_t}. \tag{5.7}$$

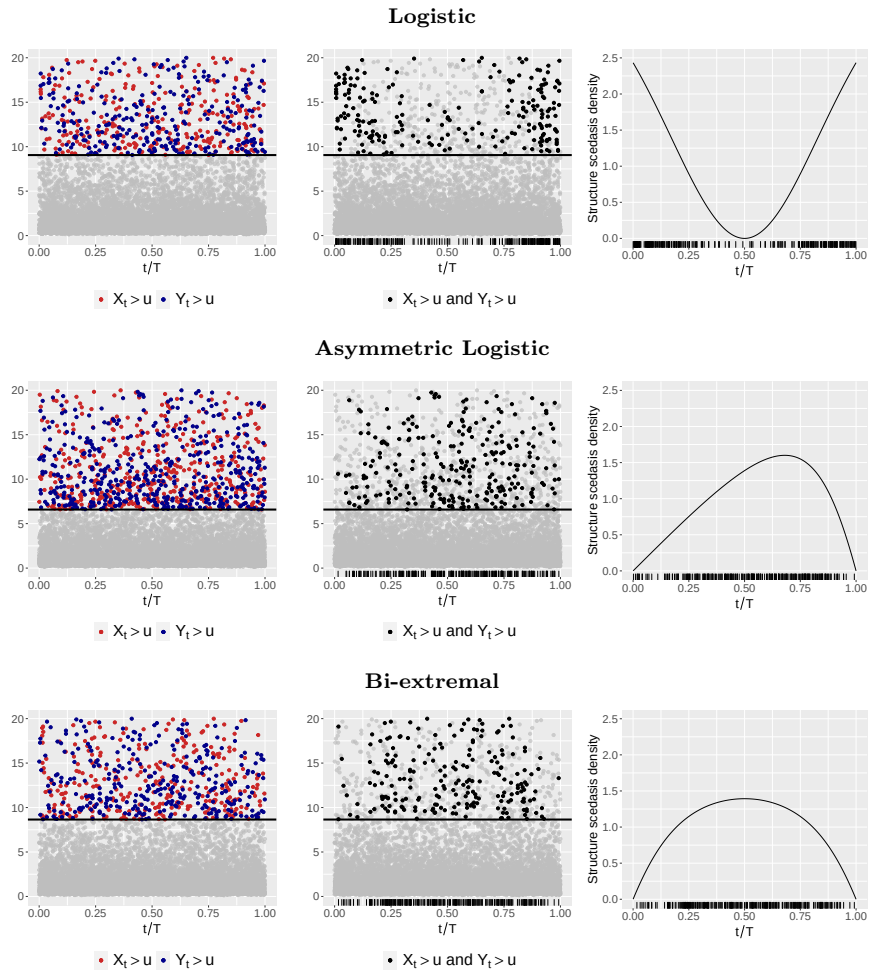


Figure 5.1: Structure scedasis function for time-varying logistic extreme value copulas from Examples 10–11 along with simulated data ($T = 5000$). (Left) Data above threshold. (Middle) Rug of times of exceedances of Z_t above threshold and corresponding exceedances. (Right) Structure scedasis function.

A more general framework is provided by the following setup.

Example 11 (Bi-extremal)**.** The tail dependence function for the time-varying bi-extremal extreme value copula is

$$\ell_t(x, y) = (1 - \psi_t)x + \{(\psi_t x)^{1/\alpha_t} + y^{1/\alpha_t}\}^{\alpha_t},$$

for $x, y > 0$, where $0 < \alpha_t \leqslant 1$ and $0 \leqslant \psi_t \leqslant 1$. The true structure scedasis function is the standardized version of

$$c^*(t/T) = 1 + \psi_t - \{(\psi_t)^{1/\alpha_t} + 1\}^{\alpha_t}. \tag{5.8}$$

Examples 10 and 11 can be nested in the more general framework of a time-varying asymmetric logistic extreme value copula.

Example 12 (Asymmetric logistic)**.** The tail dependence function for the time-varying asymmetric logistic extreme value copula is

$$\ell_t(x,y) = (1-\psi_{1,t})x + (1-\psi_{2,t})y + \{(\psi_{1,t}x)^{1/\alpha_t} + (\psi_{2,t}y)^{1/\alpha_t}\}^{\alpha_t},$$

for $x, y > 0$, where $0 < \alpha_t \leqslant 1$ and $0 \leqslant \psi_{j,t} \leqslant 1$, for $j = 1, 2$. The structure scedasis function under this model is the standardized version of

$$c^*(t/T) = 2 + \psi_{1,t} + \psi_{2,t} - \{(\psi_{1,t})^{1/\alpha_t} + (\psi_{2,t})^{1/\alpha_t}\}^{\alpha_t}. \tag{5.9}$$

Figure 5.1 shows the structure scedasis function underlying the three examples above along with simulated data. We set $\alpha_t = \sin\{\pi(t/T)\}$ for Example 10; also, we set $\alpha_t = 0.5$ and $\psi_t = \sin\{\pi(t/T)\}$ for Example 11; finally, we take $\alpha_t = 0.5$, $\psi_{1,t} = t/T$ and $\psi_{2,t} = \sin\{\pi(t/T)\}$ for Example 12. As it can be seen from Figure 5.1, the more mass the structure scedasis allocates to a period, the higher the degree of extremal dependence between X_t and Y_t.

5.3 On Further Links Over a General Framework

5.3.1 Structure Scedasis and Nonstationary Multivariate Extremes

The interfaces between nonstationary multivariate extremes and heteroscedastic extremes extend more generally beyond the process $Z_t = \min\{X_t, Y_t\}$ and beyond the bivariate case. To see this, consider a more general framework that extends Section 5.2 where $Z_t \equiv \phi(\mathbf{Y}_t)$, ϕ is an aggregation rule, that is $\phi : \mathcal{Y} \subseteq \mathbb{R}^d \to \mathbb{R}$, and $\mathbf{Y}_t = (Y_{1,t}, \ldots, Y_{d,t})$. Below, we focus on the scedasis of $\phi(\mathbf{Y}_t)$, and comment on the connection with the extremal dependence of $\mathbf{Y}_t$. The setup is as follows. Let $\{\mathbf{Y}_t\}$ be a stochastic process in $\mathcal{Y} \subseteq \mathbb{R}^d$ with unit Fréchet margins, and let $\{Z_t\} = \{\phi(\mathbf{Y}_t)\}$. We assume that $F_t(z) = P(Z_t \leqslant z)$ have a common upper endpoint, and hence there exists a function

$$c^*_\phi\left(\frac{t}{T}\right) = \lim_{z\to\infty} \frac{1-F_t(z)}{1-F(z)}, \tag{5.10}$$

to which we refer to as the *structure scedasis*. If the baseline distribution function is in the domain of attraction of a GEV distribution with extreme value index $\gamma_\phi > 0$, then it can be easily shown that all $F_t(z)$ are also in the domain of attraction of a GEV distribution with the same extreme value index. Throughout, we refer to $\{Z_t\} = \{\phi(\mathbf{Y}_t)\}$ as the *structure process*, $U_z = \{\mathbf{y} : \phi(\mathbf{y}) > z\}$ as the *risk set*, $\{\mathbf{Y}_t\}$ as the *basis process*, and γ_ϕ as the *structure coefficient of tail dependence*.

Let L_t be the law of $\mathbf{Y}_t$ and let $\mu_t(A) = \int_A L_t(\mathrm{d}\mathbf{y})$, where A is a Borel set in $\mathbb{R}^d$. As it will be seen below, a constant structure scedasis corresponds to a form of stationarity in the extremal dependence structure; but before we are able to appreciate this we need to clarify

what is meant by stationarity within the current framework.

Definition 12 (ϕ-stationary extremal dependence)**.** *Let $\{Z_t\} = \{\phi(\mathbf{Y}_t)\}$ be a structure process. The extremal dependence structure of $\{\mathbf{Y}_t\}$ is said to be ϕ-stationary, if $\mu_t(U_z)$ is not a function of t, as $z \to \infty$.*

The following proposition elucidates what the scedasis of the structure process ($\{Z_t\}$) has to say on the extremal dependence of the basis process ($\{\mathbf{Y}_t\}$).

Proposition 6. *Let $\{Z_t\} = \{\phi(\mathbf{Y}_t)\}$ be a structure process. The structure scedasis in (5.10) obeys the following properties:*

1. *If $\mathbf{Y}_t$ is an asymptotically dependent random vector for all t, then the extremal dependence structure of $\{\mathbf{Y}_t\}$ is ϕ-stationary if and only if the structure scedasis density is uniform.*

2. *The mass of the risk set U_z increases (or remains constant) from time period $t-1$ to t as $z \to \infty$, if and only if the structure scedasis density increases (or remains constant) from $(t-1)/T$ to t/T.*

Proposition 6 implies that a constant structure scedasis corresponds to a stationary extremal dependence structure, whereas a peak on the structure scedasis should be understood as an indication of a higher / lower degree of association between the extremes of the basis process $\{\mathbf{Y}_t\}$ over that period (depending on ϕ).

5.3.2 Examples Based on a Class of Canonical Aggregation Rules

Simple Aggregation Rules and Kolmogorov-type

Next, we illustrate the framework from Section 5.3.1. All examples to be examined below fit into the following general class of aggregation rules which borrows inspiration from Kolmogorov's celebrated notion of regular mean de Carvalho (2016a); Kolmogorov (1930).

Definition 13 (Aggregation rules)**.** *A family of maps $\phi^{(d)} : \mathcal{Y} \subseteq \mathbb{R}^d \to \mathbb{R}$ is a simple aggregation rule if it continuous and increasing in each variable, for every $d \in \mathbb{N}$. A simple aggregation rule is said to be of Kolmogorov-type if in addition it obeys the following properties, for every $\mathbf{y} = (y_1, \ldots, y_d) \in \mathcal{Y}$:*

1. *$\phi^{(d)}(\mathbf{y})$ is symmetric.*

2. *If we aggregate components of $\mathbf{y}$, $a = \phi^{(q)}(y_1, \ldots, y_q)$, the overall aggregated value should remain unaffected; that is, for every $1 \leqslant q < d$ and $d \geqslant 2$,*

$$\phi^{(d)}(y_1, \ldots, y_q, y_{q+1}, \ldots, y_d) = \phi^{(d-q+1)}(a, y_{q+1}, \ldots, y_d).$$

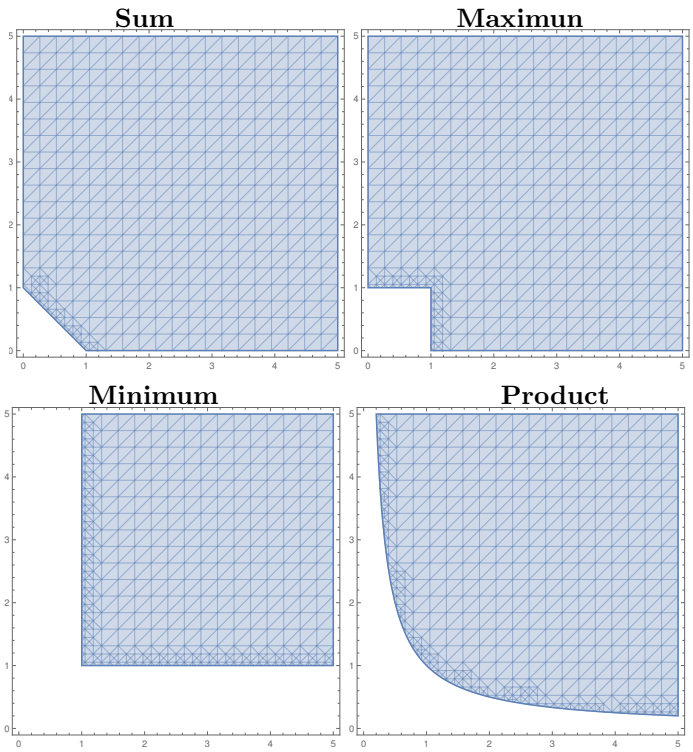


Figure 5.2: Example of the risk sets over different Kolmogorov aggregation rules.

To ease notation below we drop the superscript in $\phi^{(d)}$ unless required. Examples of *Kolmogorov aggregation rules* include $\phi_m(\mathbf{y}) = \bigwedge_{i=1}^d y_i$, $\phi_M(\mathbf{y}) = \bigvee_{i=1}^d y_i$, $\phi_R(\mathbf{y}) = \sum_{i=1}^d y_i$, and $\phi_P(\mathbf{y}) = \prod_{i=1}^d y_i$. Examples of *simple aggregation rules* include the Bruun–Tawn norm (Bruun and Tawn, 1998) and positive linear combinations, respectively given by $\phi_{BT}(\mathbf{y}) = (d^{-1}\sum_{i=1}^d y_i^\beta)^{1/\beta}$ with $\beta \in \mathbb{R}$, and $\phi_{LC}(\mathbf{y}) = \sum_{i=1}^d \beta_i y_i$, with $\beta_i > 0$ for all i, and where $\mathbf{y} \in \mathcal{Y} = (0,\infty)^d$. Examples of Kolmogorov structure processes include the time-varying minimum, maximum, and radius

$$Z_t^m \equiv \phi_m(\mathbf{Y}_t) = \bigwedge_{i=1}^d Y_{i,t}, \quad Z_t^M \equiv \phi_M(\mathbf{Y}_t) = \bigvee_{i=1}^d Y_{i,t}, \quad Z_t^R \equiv \phi_R(\mathbf{Y}_t) = \sum_{i=1}^d Y_{i,t},$$

where $\mathbf{Y}_t = (Y_{1,t}, \ldots, Y_{d,t})$, all of which can be examined within the setup proposed in this chapter. And indeed, the heteroscedastic extremes of Z_t^m have already been pre-examined in Section 5.2 in the case $d = 2$. Figure 5.2 shows the risk sets over different Kolmogorov aggregation rules.

Connecting the Scedasis of Different Kolmogorov Aggregation Rules

Connections between structure scedasis yield from different aggregation rules can be established under some assumptions. For example, it can be shown that the structure scedasis of the time-

varying maximum, c_M^*, and time-varying minimum, c_m^*, for $d = 2$ relate as follows

$$c_M^*\left(\frac{t}{T}\right) = 2 - c_m^*\left(\frac{t}{T}\right). \tag{5.11}$$

As another example, note that for a general positive integer $d \geqslant 2$, the scedasis of the radius and that of the maximum coincide, i.e.,

$$c_R^*\left(\frac{t}{T}\right) = c_M^*\left(\frac{t}{T}\right),$$

under a similar assumption as that of Bae and Ko (2017) that for every $j = 2, \dots d$,

$$\frac{P(Y_{1,t} + \cdots + Y_{j-1,t} > y - x \mid Y_{j,t} = x)}{P(Y_{1,t} + \cdots + Y_{j-1,t} > y - x)} = O(1)$$

holds uniformly for all $x \in [y_0, y]$, for every t. Well known-results on the sum of heavy-tailed dependent risks Albrecher et al. (2006) can be used to characterize the scedasis of the time-varying radius, c_R^*. For example, for $d = 2$, it follows that

$$c_R^*\left(\frac{t}{T}\right) = 2,$$

provided that the copula density can be bounded for all $(a, b) \in [x_0, 1]^2$, for a certain constant $x_0 < 1$; see (Albrecher et al., 2006, Lemma 2.7) for details. In other words, the time-varying radius structure.

Structure Coefficient of Tail Dependence for Kolmogorov Aggregation Rules

Let $\mathscr{G}_\phi \subseteq (0, \infty)$ be the parameter space for γ_ϕ. Different aggregation rules will lead to different parameter spaces. For example, it can be shown that the parameter space of γ_ϕ for the time-varying maximum and radius is degenerated as $\mathscr{G}_M = \mathscr{G}_R = \{1\}$. On the other hand, the parameter space of γ_ϕ for the time-varying minimum is $\mathscr{G}_m = (0, 1]$ and the parameter space for the time-varying product is $\mathscr{G}_P = (0, d]$. The specific value of γ_ϕ will depend on the extremal dependence of the basis process, and it may vary for the cases of asymptotic dependence (AD) and asymptotic independence (AI). For instance, $\gamma_P = d$ for AI and $\gamma_P < d$ for AD, whereas $\gamma_m < 1$ for AD and $\gamma_m = 1$ for AI. For details, see Section 5.9.2.

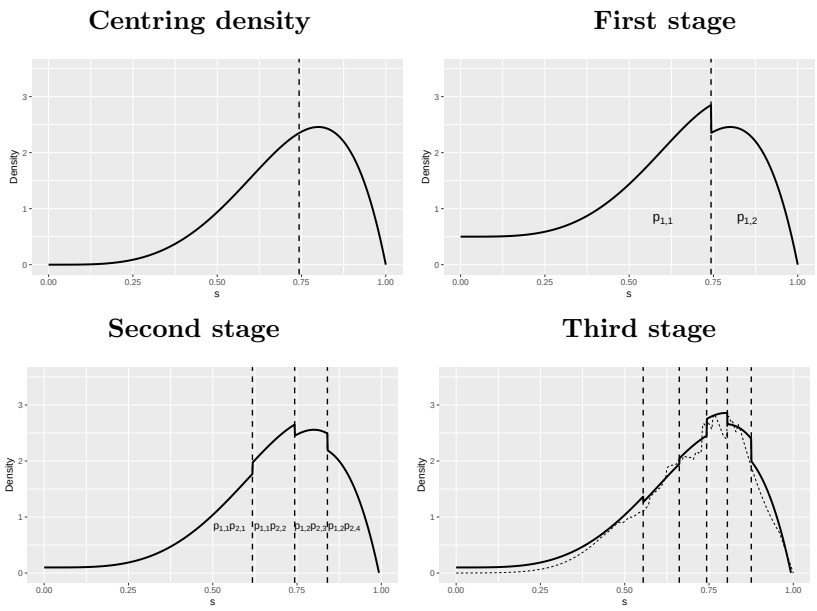


Figure 5.3: Example of Polya tree densities centred at a Beta$(5, 2)$ density over stages 1–3; the third stage is mixed over $a \sim \text{LN}(\log 2, .05)$ and $b \sim \text{LN}(\log 5, .05)$.

5.4 Semiparametric Bayesian Inference for Heteroscedastic Extremes

In this section we develop estimators for the two targets of interest, the structure scedasis function (c_ϕ) and the coefficient of tail dependence (γ_ϕ). Note that we only aim to learn about γ_ϕ for aggregation rules for which $\mathscr{G}_\phi$ is non-degenerated. Our Bayesian approach will be semiparametric, as it entails setting a prior on a function space (for c_ϕ) as well as a prior over a scalar (for γ_ϕ). The proposed model can be completely characterized by the parameters $(c_\phi, \gamma_\phi) \in \mathscr{C} \times \mathscr{G}_\phi$ where $\mathscr{C}$ is the space of all continuous density functions supported over the unit interval.

Here and below—whenever typographically convenient and there is no ambiguity—we omit the dependence of c_ϕ and γ_ϕ on the subscript ϕ, and define the integrated structure scedasis as $C(s) := \int_0^s c(u)\mathrm{d}u$, for $s \in [0, 1]$. Similarly to Poon et al. (2003) we suggest to first estimate γ and only if there is evidence in favor of asymptotic dependence to estimate c.

5.4.1 Inference for the Structure Coefficient of Tail Dependence

We now move to the parametric part of the model. Again, we recall that we only aim to learn about the structure coefficient of tail dependence whenever $\mathscr{G}_\phi$ is not a singleton; see Section 5.3.2 for details. From (5.5) it follows that $Z_t = \phi(\mathbf{Y}_t)$ is in the domain of attraction of a GEV distribution with extreme value index $\gamma_\phi > 0$, for all t. Consider the relative excesses

$E = \{E_1, \ldots, E_k\} = \{Z_t/u : Z_t > u\}$, for a sufficiently large threshold u. Then, it follows that $P(E_j > z \mid Z_j > u) \approx z^{-1/\gamma_\phi}$, for $j = 1, \ldots, k$. To conduct Bayesian inference we need to set a prior $\mathsf{p}(\gamma_\phi)$ supported over $\mathscr{G}_\phi$ of the structure coefficient of tail dependence γ_ϕ. The idea is to define a prior as a mixture of a distribution with support on the open interval of $\mathscr{G}_\phi$ and a point mass at the upper limit. That is, our prior for the structure coefficient of tail dependence includes a point mass at the upper limit so to induce shrinkage—in case there is evidence in favor of asymptotic dependence. To make matters concrete, below we focus on ϕ_m, in which case $0 < \gamma \leqslant 1$, but the same principles extend immediately for the other aggregation rules such as ϕ_P. Consider the following hierarchical model,

$$\mathsf{p}(E_1, \ldots, E_k \mid \gamma, \pi) = \gamma^{-k} \prod_{j=1}^{k} E_j^{-(1+1/\gamma)},$$

$$\mathsf{p}(\gamma \mid \pi) = \pi + (1-\pi)\beta(\gamma; a_\gamma, b_\gamma), \quad \mathsf{p}(\pi) = \beta(\pi; a_\pi, b_\pi).$$

where $\beta(\cdot; a, b)$ is the density of a Beta distribution with shape $a, b > 0$. Thus, the full conditionals of γ and π are,

$$\mathsf{p}(\gamma \mid \pi, E) \propto \gamma^{-k} \prod_{j=1}^{k} E_j^{-(1+1/\gamma)} \left\{\pi + (1-\pi)\beta(\gamma; a_\gamma, b_\gamma)\right\},$$

$$\mathsf{p}(\pi \mid \gamma) = \beta(\gamma; a_\gamma + |r_i|, b_\gamma + i - |r_i|)$$

where $r_i = \{j \leqslant i : \gamma^{(j)} = 1\}$ for a sample $\gamma^{(1)}, \ldots, \gamma^{(i)}$. The full conditional of γ has no closed-form expression, and hence we update this parameter using a Metropolis–Hastings step. For details on the algorithm, and comments on extensions to other aggregation rules see Section 5.9.3.

5.4.2 Polya Tree-based Inference for the Structure Scedasis Function

Setting a Prior on the Space of Structure Scedasis Functions

Our approach for learning from data involves ideas, concepts, and methods from Bayesian nonparametrics (Ghosal and Van der Vaart, 2015). Most Bayesian nonparametric approaches, such as Polya trees (Lavine et al., 1992, 1994) and Dirichlet processes (Ferguson, 1973), can be understood as extensions of standard parametric methods in the sense that they are centered a priori around a parametric model, $\{C_{0,\theta} : \theta \in \Theta \subseteq \mathbb{R}^p\}$, but assign positive mass to a variety of alternatives. Hence, a recurring theme in the Bayesian nonparametric literature is to consider a parametric approach as a reference, while allowing for other alternative models to 'take over' when data suggests that the parametric model is inappropriate. To do Bayesian nonparametrics in the sense described above we need prior probability models over probability distributions.

Bayesian inference for the structure scedasis function involves defining a prior over $\mathscr{C}$. Our

prior will be based on Polya trees, and thus we start with some preparations on the latter. To illustrate how Polya trees provide a natural extension of parametric statistical models, below we consider the process of generalizing a Beta(a, b) family via a Polya tree, with $a, b > 0$; other statistical models can be generalized in a similar fashion, and keeping in mind that scedasis densities are supported on the unit interval here we focus on distributions supported in $(0, 1)$. A Polya tree entails a number of stages (J), a centering distribution $(C_{0,\theta})$, and each stage involves a certain number of parameters. In the first stage, we partition the unit interval into two bins, that is $B_{1,1} = (0, m)$ and $B_{1,2} = (m, 1)$, where $m = \mathbb{I}^{-1}_{1/2}(a, b)$ is the median and $\mathbb{I}_x(a, b)$ is denoting the regularized Beta function. See Figure 5.3 for the case $a = 5$ and $b = 2$. Let T_1 follow the first stage distribution; the parameters of the first stage quantify the amount of mass of T_1 that lies below and above the median of the centring distribution, that is

$$p_{1,1} \equiv P(T_1 \in B_{1,1}), \quad p_{1,2} \equiv P(T_1 \in B_{1,2}) = 1 - p_{1,1}.$$

Let's now move to the second stage; let T_2 follow the second stage distribution. We proceed as in the first stage but break the unit interval into four pieces with equal mass $B_{2,1} = (0, q_1]$, $B_{2,2} = (q_1, m]$, $B_{2,3} = (m, q_3]$, $B_{2,4} = (q_3, 1)$, where q_1 and q_3 are respectively the first and third quartiles of the centring Beta distribution. The parameters of the second stage are conditional probabilities relative to the partition of the first stage, that is,

$$\begin{aligned} p_{2,1} &\equiv P(T_2 \in B_{2,1} \mid T_2 \in B_{1,1}), \quad p_{2,2} \equiv P(T_2 \in B_{2,2} \mid T_2 \in B_{1,1}), \\ p_{2,3} &\equiv P(T_2 \in B_{2,3} \mid T_2 \in B_{1,2}), \quad p_{2,4} \equiv P(T_2 \in B_{2,4} \mid T_2 \in B_{1,2}). \end{aligned}$$

The probability of each set in the partition for the second stage is then

$$\begin{aligned} P(T_2 \in B_{2,1}) &= p_{1,1}p_{2,1}, \quad P(T_2 \in B_{2,2}) = p_{1,1}p_{2,2}, \\ P(T_2 \in B_{2,3}) &= p_{1,2}p_{2,3}, \quad P(T_2 \in B_{2,4}) = p_{1,2}p_{2,4}. \end{aligned}$$

See Figure 5.3 for a blueprint of these stages; the subsequent stages extend analogously. If we consider the parameters a and b as random—and average over a prior on (a, b)—then we obtain a so-called finite mixture of Polya trees (MPT).

Data

To learn about the structure integrated scedasis, C, via a mixture of Polya tree priors we proceed as follows. Let $I = \{t/T : Z_t > u\} = \{\tau_1, \ldots, \tau_k\}$ be the standardized times of the $k = o(T)$ exceedances and define $\eta_t = 1(Z_t > u)$. Under Equation (5.10), where $1 - F_t(z) \approx$

$c(t/T)\{1-F(z)\}$ for $z > u$, the likelihood is given by

$$L(c(t/T),\gamma) = \prod_{t=1}^{T} \{c(t/T)f(z)\}^{\eta_t} \left[1 - c(t/T)\{1-F(z)\}\right]^{1-\eta_t}, \tag{5.12}$$

for large u, where $f(z)$ is the density of the baseline distribution in (5.2).

Inference for Structure Scedasis

The hierarchical representation of the model for the structure scedasis function is as follows

$$I \mid C \sim C, \quad C \mid \theta, \Pi \sim \mathrm{PT}^J(\alpha, C_{0,\theta}), \quad \theta \sim \mathsf{p}(\theta).$$

Here, $\Pi = \{\Pi^j; j = 1,\ldots,J\}$ is a sequence of nested partitions, such that the jth level $\Pi^j = \{B_{j,l} : l = 1,\ldots,2^j\}$ is a partition of the unit interval and $\mathrm{PT}^J(\alpha, C_{0,\theta})$ is a mixture of finite Polya trees with two main parameters: A centring distributions function on the unit interval $\{C_{0,\theta}\}$; a precision parameter ($\alpha > 0$). The parameter α controls how much deviations from the centering will be allowed, in the sense that the smaller the α the less credit is given to the centering distribution.

A tail-free process defines a random probability measure C satisfying the tree additivity property, that is, $C(B_{j-1,k}) = C(B_{j,k+1}) + C(B_{j,k})$. Let $p_{j,2l-1} = C(B_{j,2l-1} \mid B_{j-1,l})$ and $p_{j,2l} = C(B_{j,2l} \mid B_{j-1,l})$ be the probability of the sets $B_{j,2l-1}$ and $B_{j,2l}$ conditional to the set $B_{j-1,l}$ under the measure C, at each level $j \in \{1,\ldots,J\}$. These conditional probabilities are relative to the partition of the previous stage, and the tree additivity property implies that $p_{j,k+1} + p_{j,k} = 1$, allowing us to calculate the probability of every set $B_{j,l}$ as the product of the conditional probabilities. For example, the probability of the set $B_{3,1}$ is $C(B_{3,1}) = p_{3,1}p_{2,1}p_{1,1}$ as by construction $B_{3,1} \subset B_{2,1} \subset B_{1,1}$(see Figure 5.3). More generally, the probability of a set l in the partition J is given by the product of the subsets corresponding to the branch of the tree, that is,

$$C(B_{J,l}) = \prod_{j=1}^{J} p_{j,\lceil l2^{j-J}\rceil}. \tag{5.13}$$

A Polya tree process is a tail-free process if the conditional probabilities have independent Beta priors, such that $p_{j,2l-1} \sim \mathrm{Beta}(\alpha_{j,2l-1}, \alpha_{j,2l})$. The posterior of the Polya tree process C given the sample I is also a Polya tree process, where the parameters of the beta are updated as $p^*_{j,2l-1} \sim \mathrm{Beta}(\alpha_{j,2l-1} + k_{j,2l-1}, \alpha_{j,2l} + k_{j,2l})$, where $k_{j,l}$ is the number of exceedances from I that lie in the interval $B_{j,l}$. For more details, see Ghosal and Van der Vaart (2015). The probability of a set l in the partition J is given by (5.13), using the updated probabilities $p^*_{j,l}$,

$$C(B_{J,l} \mid \text{else}) = \prod_{j=1}^{J} p^*_{j,\lceil l2^{j-J}\rceil}. \tag{5.14}$$

Then, following Hanson (2006), the posterior density

$$c(\tau_i \mid \Pi, \theta) = 2^J C(B_{J,l(\tau_i)} \mid \text{else}) c_{0,\theta}(\tau_i), \quad i = 1, \ldots, k, \tag{5.15}$$

where $c_{0,\theta} = \mathrm{d}C_{0,\theta}/\mathrm{d}s$, and where $l(\tau_i)$ is the set l at level J containing τ_i, for all i. Note that (5.15) implies that the posterior structure scedasis is a suitably 'tilted' version of the centering $C_{0,\theta}$.

In terms of the choice of the parameters, we consider the rule of thumb $J = \log_2(k)$ for the number of levels (Hanson and Johnson, 2002), and set $\alpha_{j,l} = \alpha j^2$ as this guarantees an absolutely continuous C with probability one in an infinite tree (Kraft, 1964).

5.5 Simulation Study

5.5.1 Simulation Setup and Preliminary Analysis

We now assess the performance of the proposed methods using simulated data; below we focus on the aggregation rule $\phi_m(x, y) = \min(x, y)$, but in the Supporting Materials B we report a twin numerical study for the structure processes generated from the aggregation rule $\phi_R(x, y) = x+y$; we do not examine performance for $\phi_M(x, y) = \max(x, y)$ as it is straightforward to obtain the structure scedasis of the time-varying maximum from that of the time-varying minimum as can be seen from (5.11). First, we concentrate on illustrating the methods on a one shot experiment and on describing the scenarios from which the data have been simulated from; a Monte Carlo simulation study will be presented in Section 5.5.2. The scenarios under which data are generated stem from Examples 10–12. Specifically, $T = 1000$ observations are simulated from:

- **Scenario A**: Logistic extreme value copula from Example 10, with $\alpha_t = \sin\{\pi(t/T)\}$.
- **Scenario B**: Bi-extremal copula from Example 11, with $\alpha_t = 0.5$, and $\psi_t = \sin\{\pi(t/T)\}$.
- **Scenario C**: Asymmetric logistic extreme value copula from Example 12, with $\alpha = 0.5$, $\psi_{1,t} = t/T$, and $\psi_{2,t} = \sin\{\pi(t/T)\}$.

We have transformed the simulated data to unit Fréchet margins using the empirical distribution function, so to assess the effects of the transformation to common margins on the proposed methods. Some comments on learning about the structure scedasis function via a mixture of Polya trees are in order. We use a Beta distribution as the baseline distribution, that is $C_{0,\theta} \sim \text{Beta}(a, b)$, and we set $a \sim \text{Log-normal}(m_0, s_0)$ and $b \sim \text{Log-normal}(\tau_1, \tau_2)$; finally, we set $\alpha \sim \text{Gamma}(a_0, b_0)$. In terms of hyperparameters, we set $a_0 = 0.1$, $b_0 = .1$, $M = 8$ for the levels of the tree, $m_0 = 0.5$, $s_0 = 1$, $\tau_1 = .01$, and $\tau_2 = .01$ for the structure scedasis and $a_\gamma = 1$, $b_\gamma = 1$, $a_w = 1$ and $b_w = 1$ for the coefficient of tail dependence. For the threshold

we set the $u = 0.95\%$ quantile and run a burn-in period of 5000 iterates, after which we saved 5000 posterior iterates.

The outcome of a single-run experiment conducted according to the settings above is presented in Figure 5.4. Such experiment allows us to anticipate some strengths and limitations, with the proposed method. As it can be seen from this figure, our estimator is overall close to the true structure scedasis function and it thus captures the joint frequency of extremes over time. Also, the proposed Polya tree-based method is fully supported on the unit interval, and it does not suffer from boundary bias. Parenthetically, Figure B.1 in the Supporting Materials B underscores how poor the fit of a plain vanilla kernel-based method would be over Scenario A given the issue of boundary bias. While bias correction methods have been proposed to address the latter issue (e.g. Jones, 1993), we regard the facts that our Polya tree-based estimator does not require such correction, nor it entails the choice of a bandwidth, as favourable aspects.

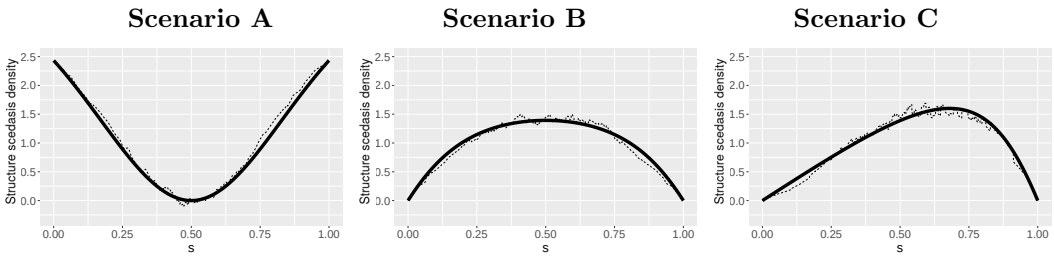


Figure 5.4: Structure scedasis density estimate obtained via the posterior mean of a mixture of finite Polya trees over a single-run experiment (dashed) plotted against true (solid). Here, the structure process is the time-varying minimum.

5.5.2 Monte Carlo Simulation Study

Here we report the main numerical findings from a Monte Carlo simulation study. We simulated 1 000 time series of length $T = 500, 1\,000, 5\,000$, and 10 000 from Scenarios A–C introduced in Section 5.5.1. Figure 5.5 show the structure scedasis density estimates obtained with the proposed methods for Scenarios A–C over this Monte Carlo simulation study; we used the same prior specification as in Section 5.5.1, and have also transformed once more the simulated data to unit Fréchet margins using the empirical distribution function. We start with the structure scedasis density. Figure 5.5 suggests that the proposed Polya tree-based estimator for the structure scedasis density perform well over Scenarios A–C, in line with the preliminary experiments from Section 5.5.1. Next, we move to the Bayesian estimator of the coefficient of tail dependence from Section 5.4. The Monte Carlo posterior mean estimates reported in Table 5.1 suggest an overall good accuracy of the proposed Bayesian approach for learning about the coefficient of tail dependence. In an attempt to examine the frequentist properties of our Bayesian method from a numerical viewpoint, we also report in Table 5.1 the coverage probabilities, i.e. the number of times the true value γ was contained the credible interval of

probability $1-\alpha$. As it can be seen from Table 5.1 the coverage probabilities reasonably follow the significance levels, especially for higher T, thus suggesting a good frequentist behavior the proposed approach. For Scenarios A–C the true coefficient of tail dependence is $\gamma = 1$.

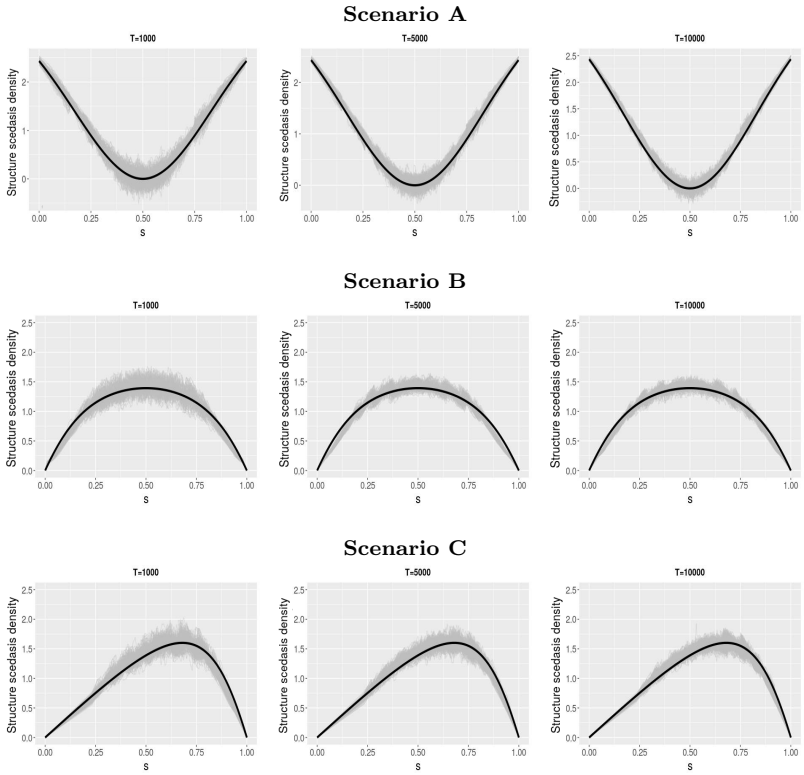


Figure 5.5: First 150 trajectories of structure scedasis density estimates obtained via the posterior mean of a mixture of finite Polya trees over the Monte Carlo simulation study (gray) plotted against true (black). Here, the structure process is the time-varying minimum.

5.5.3 Nonstationary Margins

In this subsection we present the same analysis from Section 6 of the chapter but allowing for the margins to be nonstationary. We transform the bivariate returns (R_t^X, R_t^Y) to unit Fréchet margins (X_t, Y_t) using the transformation:

$$(X_t, Y_t) = (-1/\log F_{R_t^X}(R_t^X), -1/\log F_{R_t^Y}(R_t^Y)),$$

where $F_{R_t^X}$ and $F_{R_t^Y}$ are the respective marginal time-varying distribution functions for R_t^X and R_t^Y. To estimate $F_{R_t^X}$ and $F_{R_t^Y}$ we compute the time-varying distribution function estimator

Table 5.1: Monte Carlo posterior mean coefficient of tail dependence and coverage probabilities I_α at levels $\alpha = 0.90, 0.95$. For all cases the true coefficient of tail dependence is $\gamma = 1$.

Scenario	**Sample size** (T)	**Monte Carlo posterior mean**	$I_{.90}$	$I_{.95}$
A	500	0.950	0.997	1.000
B	500	0.956	0.998	1.000
C	500	0.936	0.994	0.994
A	1 000	0.966	1.000	1.000
B	1 000	0.973	1.000	1.000
C	1 000	0.945	0.988	0.996
A	5 000	0.987	1.000	1.000
B	5 000	0.990	1.000	1.000
C	5 000	0.958	0.972	0.990
A	10 000	0.989	1.000	1.000
B	10 000	0.968	0.924	0.967
C	10 000	0.947	0.930	0.986

of Harvey and Oryshchenko (2012). That is

$$\widehat{F}_{R_t^X}(x) = \sum_{i=1}^{T} H\left(\frac{x - R_i^Y}{h}\right) w_{t,i}, \quad t = 1, \ldots, T,$$

where $H(\cdot)$ is a kernel distribution function, and $\sum_{i=1}^{T} w_{t,i} = 1$ for all t. We use the Gaussian kernel and the weights are computed using the algorithm of Koopman and Harvey (2003), where it follows that $w_{t,i} \approx (1-w)/(1+w)w^{|t-i|}$ for $i = 1, \ldots, T$. and $0 \leqslant w < 1$. We tried different values of the parameter w and obtain similar results, the results report here are using $w = 6$. Let $\widehat{F}_{R_t^Y}$ being analogously defined. Figure 5.6 shows the scedasis function pairwise and Figure 5.7 shows the estimation for the 5-dimensional case. Table 5.2 shows the posterior mean and credible bands of the coefficient of extreme value dependence. The posterior mean coefficient of tail dependence for this multivariate analysis is 0.67 (CI = $(0.54, 0.85)$). Thus, in both cases the results are tantamount to the ones obtained in Section 5 of the chapter.

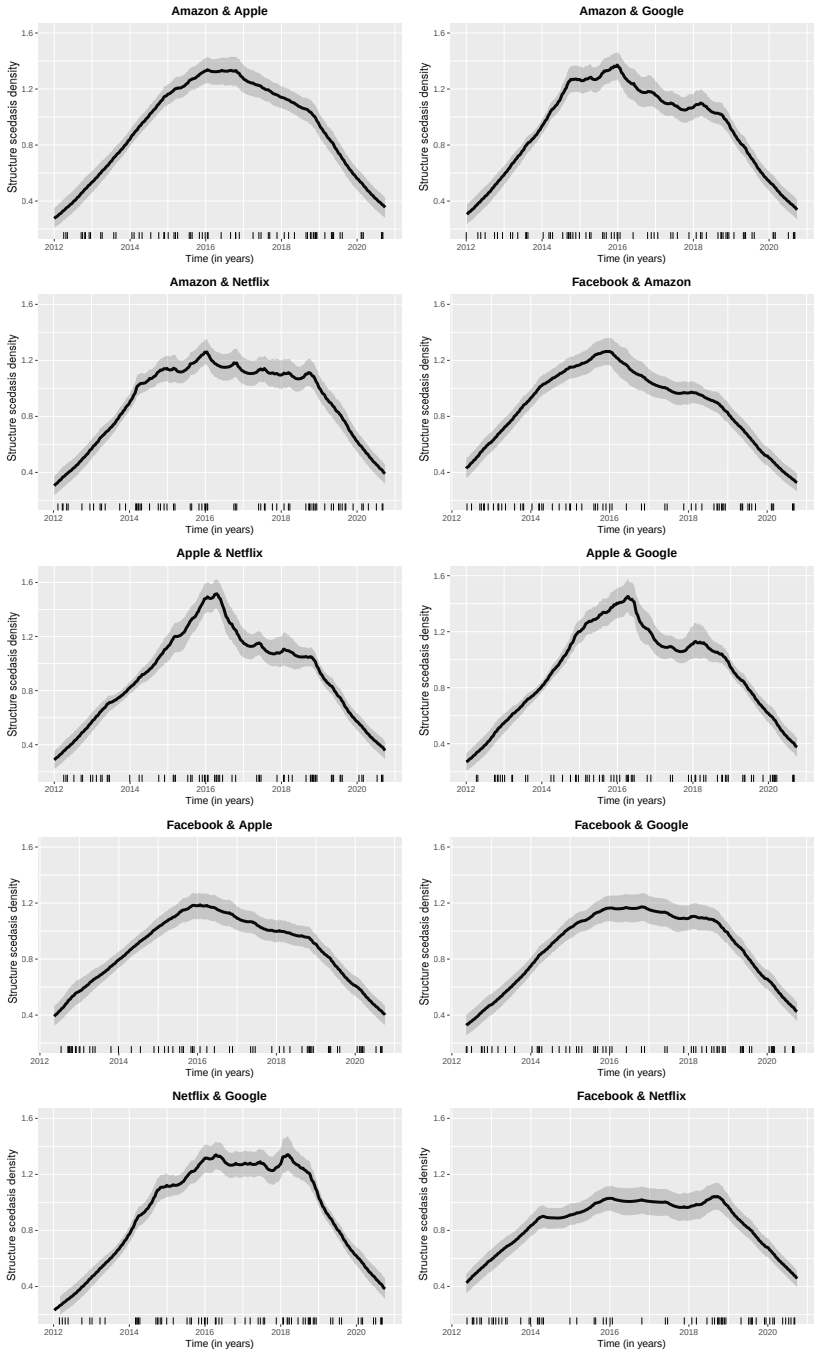


Figure 5.6: Structure scedasis (for the pairwise setting) posterior mean along with pointwise credible bands using finite mixtures of Polya trees for FAANG stocks over 2012–2020; here, the structure process is the time-varying minimum and the marginal transformation is using the time-varying version.

Table 5.2: Posterior median and 95% credible intervals of the extreme value index of the structure process of the time-varying minimum and the marginal transformation is using the time-varying version.

Pair of FAANG stocks	Lower limit	Posterior median	Upper limit
Facebook–Amazon	0.967	1.000	1.000
Facebook–Apple	0.996	1.000	1.000
Facebook–Netflix	0.986	1.000	1.000
Facebook–Google	1.000	1.000	1.000
Amazon–Apple	1.000	1.000	1.000
Amazon–Netflix	0.987	1.000	1.000
Amazon–Google	0.992	1.000	1.000
Apple–Netflix	0.997	1.000	1.000
Apple–Google	1.000	1.000	1.000
Netflix–Google	0.995	1.000	1.000

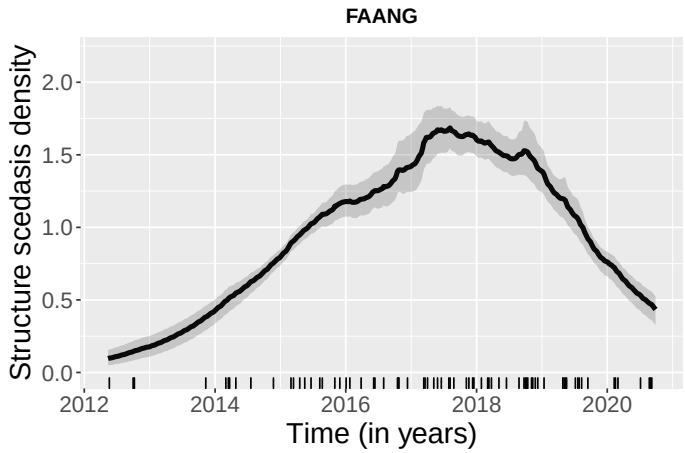


Figure 5.7: Structure scedasis (for the multivariate setting) posterior mean along with pointwise credible bands using finite mixtures of Polya trees for FAANG stocks over 2012–2020; here, the structure process the time-varying minimum and the marginal transformation is using the time-varying version.

5.5.4 Structure Scedasis of Minimum

This section supplements the Monte Carlo study in the chapter by considering: i) a case with a smaller sample size ($T = 500$); ii) studying two additional scenarios with $\gamma < 1$ (as those in Section 5.5.2 were all for $\gamma = 1$). In terms of i), we report the simulation study for Scenarios A–C when $T = 500$ in Figure B.2, where it can be seen a good overall performance of the proposed Polya tree inference for the structure scedasis. In terms of ii), we construct simulation scenarios with $\gamma < 1$ using a the time-varying Pareto-type model, that is, $F(Z_t > z) = z^{-1/\gamma}\mathcal{L}_t(z)$, where $\mathcal{L}_t(z)$ is a time-varying slowly varying function (i.e. $\mathcal{L}_t(wz)/\mathcal{L}_t(z) \to 1$, as $z \to \infty$, for any $w > 0$ and for all t), and set:

- **Scenario D**: $\mathcal{L}_t(z) = t/T$ and $\gamma = 0.3$.
- **Scenario E**: $\mathcal{L}_t(z) = \sqrt{t/T}$ and $\gamma = 0.5$.

In both scenarios we set the baseline F to be unit Pareto. The Monte Carlo posterior mean estimates of γ are reported in Table 5.3 and provide an overall good fit; the posterior scedasis estimates for the minimum are presented in Figure 5.8 and also closely follow the target.

Table 5.3: Monte Carlo posterior mean coefficient of tail dependence and coverage probabilities I_α at levels $\alpha = 0.90, 0.95$. For scenario D and E, the true coefficient of tail dependence is $\gamma = 0.3$, 0.5, respectively.

Scenario	**Sample size** (T)	**Monte Carlo posterior mean**	$I_{.90}$	$I_{.95}$
D	500	0.3048	0.824	0.930
E	500	0.5057	0.802	0.876
D	1 000	0.3043	0.860	0.952
E	1 000	0.5048	0.853	0.958
D	5 000	0.3000	0.899	0.950
E	5 000	0.5001	0.907	0.951
D	10 000	0.3000	0.901	0.948
E	10 000	0.4999	0.900	0.951

5.6 Tracking Extreme Joint Losses of Leading US Tech-Stocks

5.6.1 Data Description, Financial Rationale, and Preprocessing

In this section we apply the proposed methods so to model the magnitude and frequency of extreme joint losses of a set of Big Tech stocks known as FAANG (Facebook, Apple, Amazon, Netflix and Alphabet's Google). The data were gathered from Yahoo Finance and consist of weekly closing prices from 1 Jan. 2012 to 11 Oct. 2020; this is mostly a period of sustained growth, with a few sharp sell-offs at the end of the observation period, and over which test T_4 from Einmahl et al. (2016, p. 38) suggests that the extreme value index is constant over time. The period under analysis also includes the beginning of the Covid-19 era, which some speculate will get these Big Tech companies to become even bigger (Wigglesworth, 2020).

Figure 5.9 depicts the raw data. Since the goal will be to focus on extreme losses, we use weekly negative returns as a unit of analysis. Weekly negative returns are computed by taking

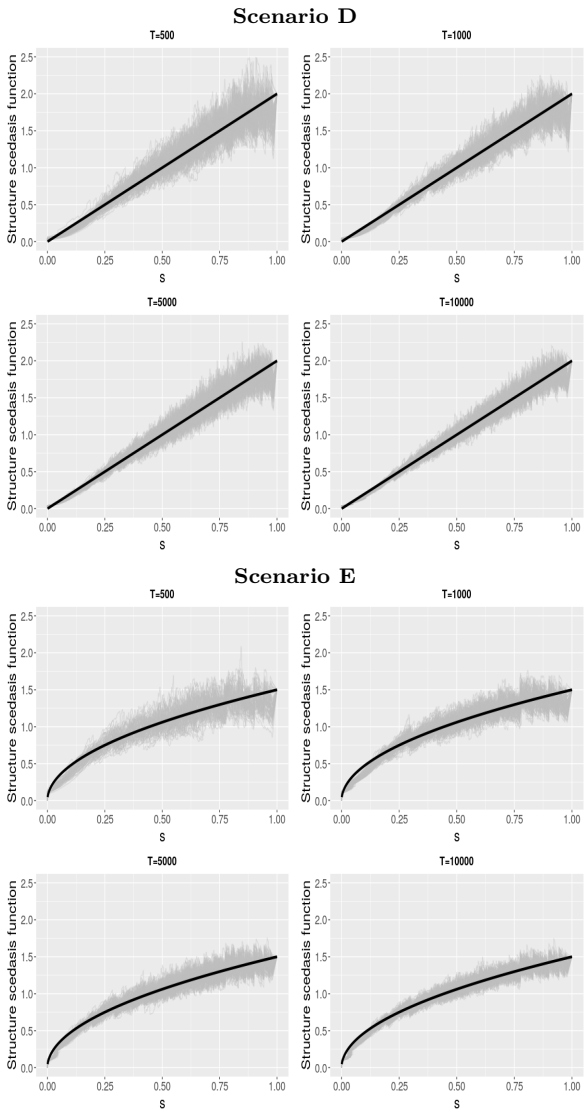


Figure 5.8: First 150 trajectories of structure scedasis density estimates obtained via the posterior mean of a mixture of finite Polya trees over the Monte Carlo simulation study (gray) plotted against true (black).; here the structure process is the time-varying minimum.

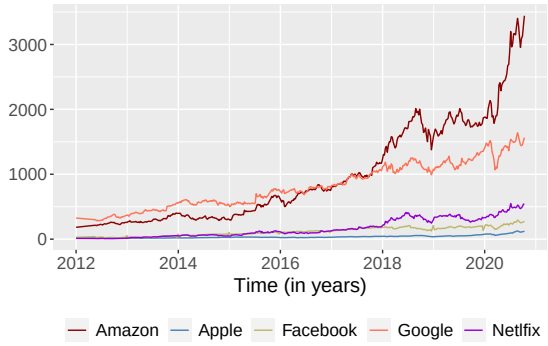


Figure 5.9: FAANG prices over 2012–2020.

the negative of the first differences of the logarithmic prices. Some comments on preprocessing are in order. We transform the bivariate returns (R_t^X, R_t^Y) to unit Fréchet marginals (X_t, Y_t) using the transformation:

$$(X_t, Y_t) = -(1/\log\{\mathbb{F}_{R^X}(R_t^X)\}, 1/\log\{\mathbb{F}_{R^Y}(R_t^Y)\}),$$

where $\mathbb{F}_{R^X}$ and $\mathbb{F}_{R^Y}$ are the respective marginal distribution functions for R_t^X and R_t^Y. We then work with exceedances of $Z_t = \min(X_t, Y_t)$ above the 0.95 quantile. We estimate $\mathbb{F}_{R^X}$ and $\mathbb{F}_{R^Y}$ using a suitably rescaled empirical distribution function, that is $\widehat{\mathbb{F}}_{R^X}(x) = 1/(T+1)\sum_{t=1}^{T} I(R_t^X \leqslant x)$ with $\widehat{\mathbb{F}}_{R^Y}$ being analogously defined. Ljung–Box tests (Tsay, 2002, Ch. 2) were applied to the Z_t and no evidence in favor of Z_t being dependent was found, with the exception of the pair Facebook & Netflix; the results below for the latter pair should keep in mind this disclaimer.

5.6.2 Learning About the Dynamics and Magnitudes of Pairwise Extreme Losses

As noted in Section 5.4, similarly to Poon et al. (2003), we first estimate γ_m and only if there is evidence in favor of asymptotic dependence we estimate c_m. We recall that our prior for the structure coefficient of tail dependence includes a point mass at 1 so to induce shrinkage—in case there is evidence in favor of asymptotic dependence. The obtained posterior inferences for the coefficient of tail dependence for all pair of stocks are presented in Table 5.4 and suggest evidence in favor of asymptotic dependence for each pair of stocks. Given this we next proceed and learn about the structure scedasis. Bayesian inference for the structure scedasis function for all pairs of FAANG stocks is presented in Figure 5.10. Some comments on implementation are in order. As in Section 5.5 we assume the number of levels to be $M = 8$. In terms of prior information, for the parameter $\alpha \sim \text{Gamma}(a_0, b_0)$ we use a non-informative prior $(a_0, b_0) = (0.1, 0.1)$, whereas for the parameter of the centring Beta distribution we set $a \sim \text{Log-normal}(\hat{\mu}_z, \hat{\sigma}_z)$ and $b \sim \text{Log-normal}(0.1, 0.1)$, where $\hat{\mu}_z$ and $\hat{\sigma}_z$ are respectively the sample mean and standard deviation. In terms of MCMC, we run a burn-in period of 5000 iterates, after which we saved 5000 posterior iterates.

As it can be seen from Figure 5.10, most structure scedasis functions tend to peak around 2016–2019, thus indicating that extreme joint losses have occurred mostly around that time. From a financial outlook the dynamics portrayed by the structure scedasis in Figure 5.10 may look surprising at first, keeping in mind the well-known fact that the 2020 pandemic crisis has led to some sharp sell-offs worldwide (Zhang et al., 2020). And in fact economists have been painting a doomsday scenario for the real economy in the short-run since early 2020 (McKibbin and Fernando, 2020). It should be noticed however that Figure 5.10 simply claims that the

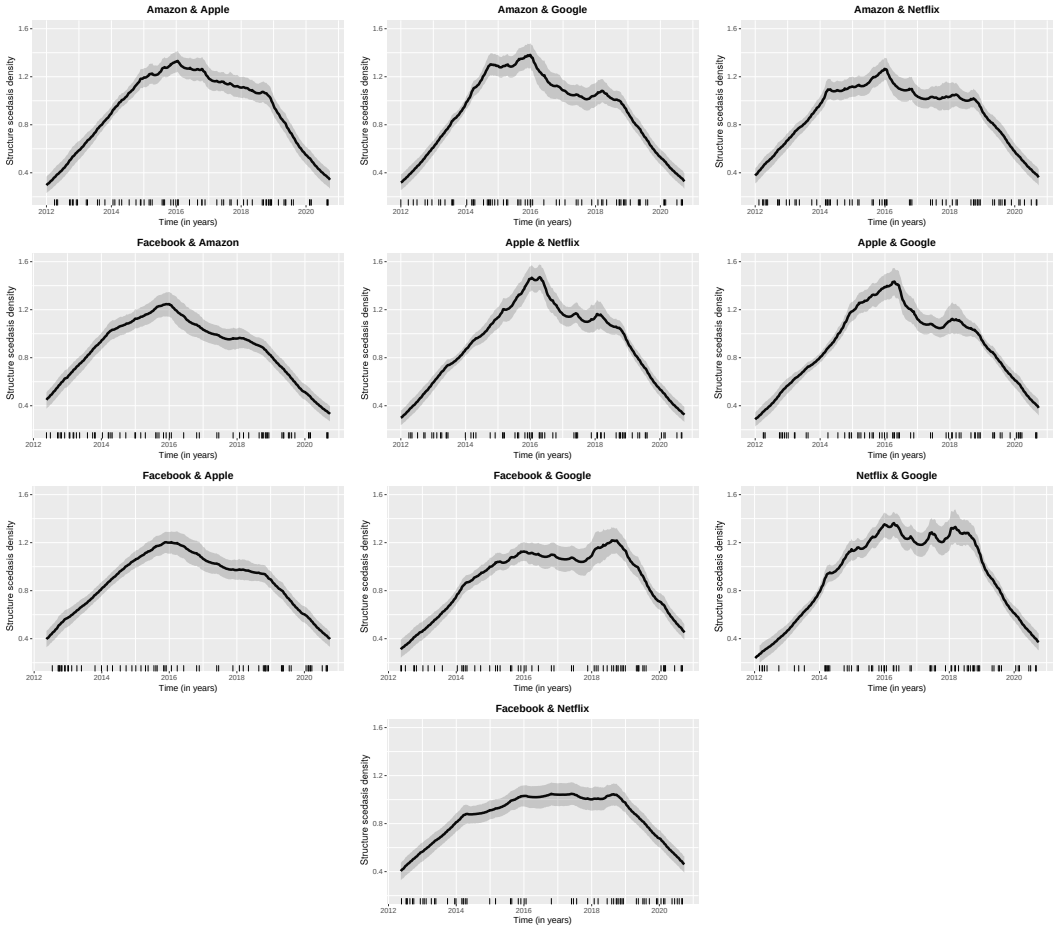


Figure 5.10: Structure scedasis (for the pairwise setting) posterior mean along with pointwise credible bands using finite mixtures of Polya trees for FAANG stocks over 2012–2020; here, the structure process the time-varying minimum.

relative frequency of extreme joint losses has been higher over 2016–2019, than over the 2020 pandemic outbreak.

Parenthetically, we note that a probabilistic interpretation can be given to Figure 5.10, in the sense that the structure scedasis in the joint tail region (5.2) is proportional to the probability of extreme joint losses over time. To see this, note that that for a fixed large z,

$$P(X_t > z, Y_t > z) \approx \frac{c_m^*(t/T)}{z^{1/\gamma_m}} \propto c_m^*(t/T).$$

5.6.3 Multivariate Analysis

Section 5.6.2 offered a pairwise analysis but in practice the interest often lies in more than two stocks. Keeping this in mind, below we conduct an analysis that focuses on the time-varying minimum structure process $Z_t = \bigwedge_{i=1}^{d} Y_{i,t}$, where $\{Y_{1,t}\}, \ldots, \{Y_{d,t}\}$ are d time series (say, $d = 5$ FAANG stocks). The reported analysis used the same prior specification and MCMC setup as

Table 5.4: Posterior median and 95% credible intervals of the extreme value index of the structure process of the time-varying minimum.

Pair of FAANG stocks	Lower limit	Posterior median	Upper limit
Facebook–Amazon	0.951	1.000	1.000
Facebook–Apple	0.984	1.000	1.000
Facebook–Netflix	0.943	1.000	1.000
Facebook–Google	0.988	1.000	1.000
Amazon–Apple	1.000	1.000	1.000
Amazon–Netflix	0.971	1.000	1.000
Amazon–Google	0.982	1.000	1.000
Apple–Netflix	0.954	1.000	1.000
Apple–Google	0.998	1.000	1.000
Netflix–Google	0.959	1.000	1.000

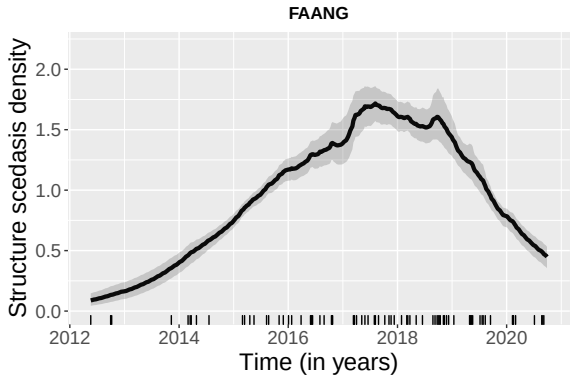


Figure 5.11: Structure scedasis (for the multivariate setting) posterior mean along with pointwise credible bands using finite mixtures of Polya trees for FAANG stocks over 2012–2020; here, the structure process the time-varying minimum.

in Section 5.6.2. Figure 5.11 depicts the structure scedasis associated with such time-varying minimum computed for all FAANG stocks. The structure scedasis of this Z_t again showcases that from 2017–2019 the frequency of extreme joint losses was actually higher over 2016–2019 than over the recent COVID-19 pandemic. Again, we underscore that Figure 5.10 simply claims that the relative frequency of extreme joint losses has been higher over 2016–2019 than over the recent pandemic outbreak. Many geopolitical issues—such as the US–China trade war (Liu and Woo, 2018; Li et al., 2018)—and US policy issues—such as President Trump impeachment (Jackman, 2017)—may have been the drivers for some of these joint sell-offs. The posterior mean coefficient of tail dependence for this multivariate analysis is 0.70 (CI $= (0.57, 0.89)$), thus suggesting that despite the sturdy growth of FAANG stocks over time the comovements of their extreme losses is substantial; this matches the intuition from Figure 5.9 where it can be seen that sharp dips for these stocks tend to be synchronized.

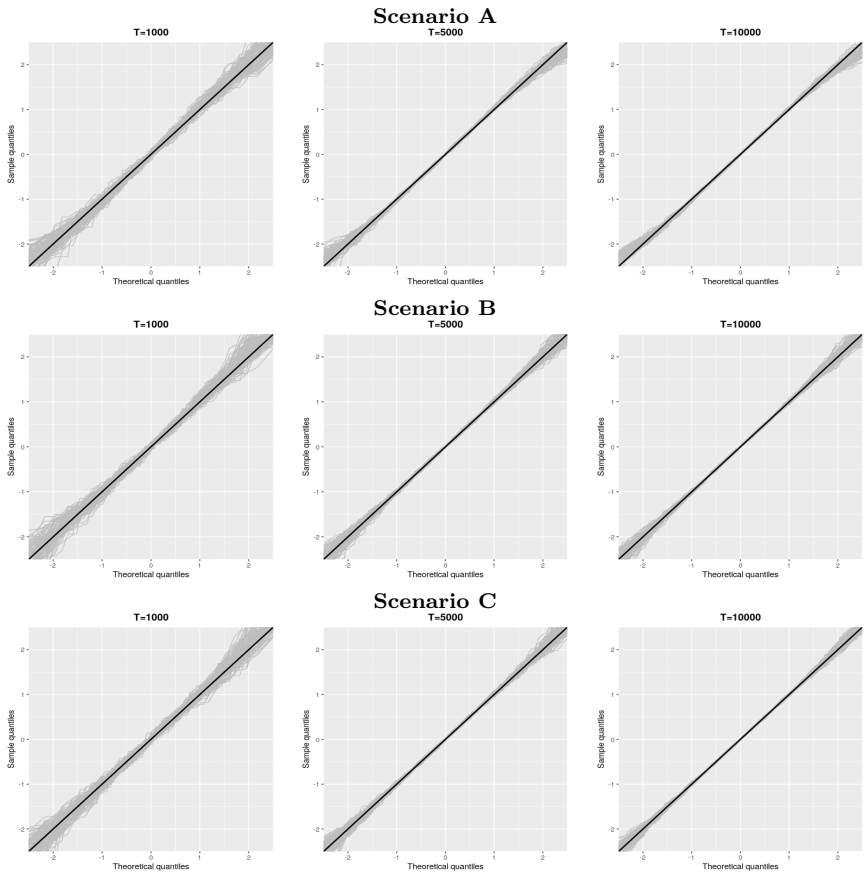


Figure 5.12: QQ-plots of randomized quantile residuals for the first 150 trajectories of structure scedasis density estimates obtained via the posterior mean of a mixture of finite Polya trees over the Monte Carlo simulation study from Section 4 in the chapter, here the structure process is the time-varying miminum.

5.7 QQ-plots from Randomized Quantile Residuals

To assess the fit of the proposed methods we resort to a version of randomized quantile residuals (Dunn and Smyth, 1996), where residuals are defined as $\varepsilon_i = \Phi^{-1}\{C(\tau_i)\}$, for $i = 1, \ldots, k$; here, C is the integrated scedasis function and τ_i the standardized time of exceedances. If C is the true distribution of the time of exceedances, it means that the assumption $\tau_i \mid C \overset{\text{iid}}{\sim} C$ is true, then $C(\tau_i)$ should be Uniform and thus ε_i should be Normal distributed, for all i. Figures 5.12–5.13 depict QQ-plots of randomized quantile residuals plotted against the theoretical standard Normal quantiles, and suggest acceptably good fits of the proposed model—both in the Monte Carlo simulation study from Section 5.5.2 and in the real data analysis from Section 5.6.

5.8 Discussion

This chapter interfaces nonstationary multivariate extremes with heteroscedastic extremes, and it introduces a flexible Bayesian approach for modelling the frequency and magnitude of the

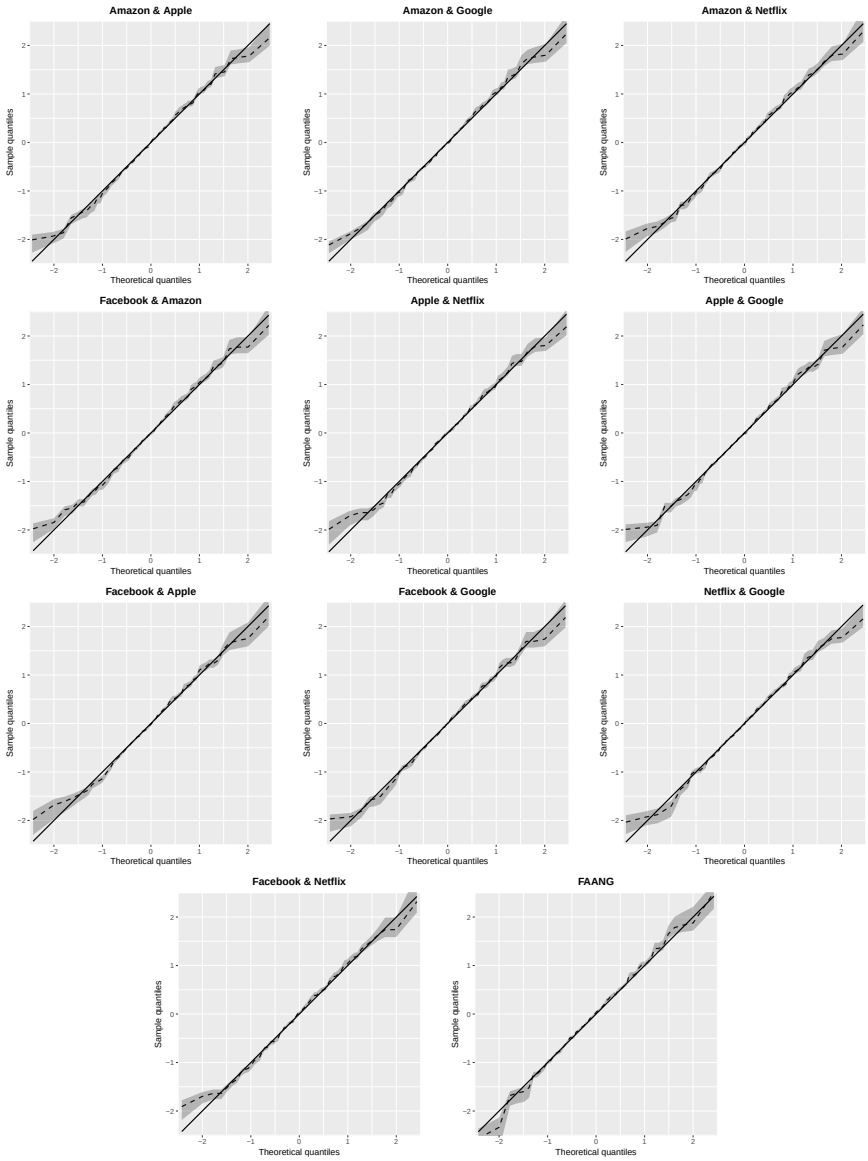


Figure 5.13: QQ-plots of randomized quantile residuals (for the pairwise setting and the 5-dimensional setting) posterior inference using finite mixtures of Polya trees for FAANG stocks over 2012–2020; here, the structure process is the time-varying minimum.

extremes of a structure process. In particular, we show that the structure scedasis summarizes key features of the extremal dependence structure of the basis process. The proposed framework is suitable for modelling nonstationary multivariate extremes, as it has been designed for tracking the dynamics governing the frequency of extreme values in the joint tail or over other risk sets defined by other rules for aggregating extremes.

We develop Bayesian estimators for the two targets of interest—the structure scedasis function and the extreme value index. The nonparametric part of the model (i.e. the structure scedasis) describes the frequency of extreme values in a risk set over time, and it is thus a measure of the level of association of the extremes over time. For learning about the behavior of the frequency of extremes in a risk set, we define a prior in the space of all structure scedasis via finite mixtures of Polya trees. Our Polya tree-based approach relies on a parametric approach as a baseline model (say, a Beta centring distribution), while allowing for deviations from it whenever the data provide evidence for it. The application of our model to the so-called FAANG stocks revealed some interesting dynamics on the frequency of joint extremes over time, especially the fact the relative frequency of extreme joint losses has been higher over 2016–2019, than over the 2020 pandemic outbreak.

For some applications the extreme value index may not be constant over time, thus questioning how to proceed in that case. We argue that even in that case, learning about the law of the standardized times of exceedances—i.e. learning about the structure scedasis—is the same as inferring about the dynamics governing extremal dependence of a basis process, and thus of the utmost interest in practice.

A natural avenue for future research entails modelling changepoints or structure changes over time in the frequency of extremes in a risk set. While it is known that breaks in tail behavior are key in applications (see, for instance, Quintos et al., 2001; Lin and Kao, 2008, and references therein), most attention has been focused on modelling structural changes in the magnitude of extremes rather than on frequency, and with the exception of de Carvalho et al. (2020) all prior developments are for the univariate setting. modelling changepoints in the frequency of extremes in a risk set would require setting a prior on the space of discontinuous scedasis functions, and with the times of the breaks being themselves treated as a parameter. Whether Leadbetter's $D(u_n)$ mixing conditions (Leadbetter, 1974) could justify the use of the proposed model for dependent sequences of extremes in a risk set is another open problem to be addressed in future research. Finally, it would be elegant from a statistical viewpoint to jointly model nonstationarities in margins and dependence but to our knowledge little has been developed on this direction.

5.9 Technical Details

5.9.1 Proof of Proposition 6

The proofs are as follows:

1. Note first that (5.10) implies that

$$c_\phi^*\left(\frac{t}{T}\right) = \lim_{z\to\infty}\frac{P(Z_t > z)}{1-F(z)} = \lim_{z\to\infty}\frac{\mu_t(U_z)}{1-F(z)}. \tag{5.16}$$

   It follows from Definition 12 that $\mu_t(U_z)$ does not depend on t, as $z \to \infty$. Thus, $c_\phi^*(t/T) = \kappa$ for a positive constant $\kappa < \infty$, and hence the standardized structure scedasis is $c(s) = 1$, for $s \in [0,1]$. The reverse implication follows analogously, from where the final result follows.

2. We only present the proof for the direct implication as the reverse implication follows analogously. By assumption the mass of the risk set increases (or remains constant) from $t-1$ to t, that is $\mu_t(U_z) \geqslant \mu_{t-1}(U_z)$, as $z \to \infty$, and hence it follows from (5.16) that

$$c_\phi^*\left(\frac{t}{T}\right) = \lim_{z\to\infty}\frac{\mu_t(U_z)}{1-F(z)} \geqslant \lim_{z\to\infty}\frac{\mu_{t-1}(U_z)}{1-F(z)} \geqslant c_\phi^*\left(\frac{t-1}{T}\right). \tag{5.17}$$

   In turn (5.17) implies that

$$c_\phi\left(\frac{t}{T}\right) = \frac{c_\phi^*(t/T)}{\int_0^1 c(s)\,\mathrm{d}s} \geqslant \frac{c_\phi^*((t-1)/T)}{\int_0^1 c(s)\,\mathrm{d}s} \geqslant c_\phi\left(\frac{t-1}{T}\right),$$

   which proves the final result. □

5.9.2 Parameter Space of Structure Coefficients of Tail Dependence

In this section we derive the support of the coefficients of tail dependence constraints for the Kolmogorov-type aggregation rules covered in Section 5.3.2. Below $\mathcal{L}$ denotes a slowly-varying function (i.e., $\mathcal{L}(az)/\mathcal{L}(z) \to 1$ as $z \to \infty$, for all $a > 0$) that may be indexed over time but to ease notation we drop its dependence on it.

Time-varying Minimum

Note that $P\{\phi_m(\mathbf{Y}_t) > z\} \sim z^{-1/\gamma_m}\mathcal{L}_m(z)$ and since

$$P\{\phi_m(\mathbf{Y}_t) > z\} \leqslant 1 - \exp(1/z) \sim 1/z,$$

we must have $0 < \gamma_m \leqslant 1$, and hence $\mathscr{G}_m = (0,1]$.

Time-varying Maximum

Note that $P\{\phi_M(\mathbf{Y}_t) > z\} \sim z^{-1/\gamma_M}\mathcal{L}_M(z)$ and since

$$\begin{aligned} P\{\phi_M(\mathbf{Y}_t) > z\} &= 1 - P\{\phi_M(\mathbf{Y}_t) \leqslant z\} \\ &= 1 - P(Y_{1,t} \leqslant z, \ldots, Y_{d,t} \leqslant z) \\ &\geqslant 1 - P(Y_{1,t} \leqslant z) \\ &= 1 - \exp(1/z) \\ &\sim 1/z \end{aligned}$$

we must have $1 \leqslant \gamma_M$, and hence $\mathscr{G}_M = [1, \infty)$.

Time-varying Radius

Note that $P\{\phi_R(\mathbf{Y}_t) > z\} \sim z^{-1/\gamma_R}\mathcal{L}_R(z)$. Since $\{\phi_R(\mathbf{Y}_t) > z\} \supset \bigcup_{i=1}^d \{Y_{i,t} > z\}$ and $\{\phi_R(\mathbf{Y}_t) > z\} \subset \bigcup_{i=1}^d \{Y_{i,t} > z/d\}$, by the principle of inclusion-exclusion it follows that

$$\begin{aligned} P\{\phi_R(\mathbf{Y}_t) > z\} &\geqslant \sum_{i=1}^d P\{Y_{i,t} > z\} + \cdots + (-1)^{d+1} P\{Y_{1,t} > z, \ldots, Y_{d,t} > z\} \\ &= P\{\phi_M(\mathbf{Y}_t) > z\} \\ &\geqslant 1 - \exp(1/z) \\ &\sim 1/z. \end{aligned} \tag{5.18}$$

and

$$\begin{aligned} P\{\phi_R(\mathbf{Y}_t) > z\} &\leqslant \sum_{i=1}^d P\{Y_{i,t} > z/d\} + \cdots + (-1)^{d+1} P\{Y_{1,t} > z/d, + \ldots, Y_{d,t} > z/d\} \\ &= P\{\phi_M(\mathbf{Y}_t) > z/d\} \\ &\leqslant P\{Y_{i,t} > z/d\} d \\ &\sim 1/z. \end{aligned} \tag{5.19}$$

Then, combining (5.18) and (5.19)

$$P\{\phi_M(\mathbf{Y}_t) > z\} \leqslant P\{\phi_R(\mathbf{Y}_t) > z\} \leqslant P\{\phi_M(\mathbf{Y}_t) > z/d\},$$

hence $\gamma_R = \gamma_M = 1$, that is $\mathscr{G}_M = \mathscr{G}_R = \{1\}$.

Time-varying Product

Note that $P\{\phi_P(\mathbf{Y}_t) > z\} \sim z^{-1/\gamma_P}\mathcal{L}_P(z)$. Since $\{\phi_P(\mathbf{Y}_t) > z\} \subset \{Y_{1,t} > z^{1/d}, \ldots, Y_{d,t} > z^{1/d}\}$, it follows that

$$P\{\phi_P(\mathbf{Y}_t) > z\} \geqslant P\{Y_{1,t} > z^{1/d}, \ldots, Y_{d,t} > z^{1/d}\} \sim z^{-1/d\gamma_m}\mathcal{L}_m(z).$$

And since $\{\phi_P(\mathbf{Y}_t) > z\} \subset \{\phi_M(\mathbf{Y}_t) > z^{1/d}\}$, we have

$$P\{\phi_P(\mathbf{Y}_t) > z\} \leqslant P\{\phi_M(\mathbf{Y}_t) > z^{1/d}\} \sim z^{-1/d\gamma_M}\mathcal{L}_M(z),$$

then $d\gamma_M \geqslant \gamma_P \geqslant d\gamma_m$ and hence $\mathscr{G}_P = (0, d]$.

5.9.3 Posterior Inference for the Structure Coefficient of Tail Dependence

Time-varying Minimum

Here we develop the posterior inference algorithm for learning about the structure coefficient of tail dependence of the time-varying minimum; we comment on how to adapt the proposed algorithm for the time-varying product as well as for other structure processes. Following the hierarchical model from Section 5.4.1, the full conditional of $\gamma \in \mathscr{G}_m = (0, 1]$ is

$$\begin{aligned}
\mathsf{p}(\gamma \mid E) &= \mathsf{p}(E \mid \gamma)\,\mathsf{p}(\gamma \mid \pi)\,\mathsf{p}(\pi) \\
&= \gamma^{-k}\prod_{j=1}^{k} E_j^{-(1+1/\gamma)}\,\{\pi + (1-\pi)\beta(\gamma; a, b)\}\,\mathsf{p}(\pi) \\
&\propto \gamma^{-k}\prod_{j=1}^{k} E_j^{-(1+1/\gamma)}\,\{\pi + (1-\pi)\beta(\gamma; a, b)\}\,.
\end{aligned}$$

To update γ we use a Metropolis–Hastings step where the proposal distribution is a mixture, $q(\gamma^* \mid \gamma) \propto \omega + (1-\omega)\mathrm{TN}(\gamma^* \mid \gamma, 1)$; here, $\mathrm{TN}(\cdot \mid \gamma, 1)$ is the density of the truncated Normal distribution on $(0, 1)$, with mean γ and variance 1. At each step, we use the values of γ to compute the full conditional of π, which can be augmented using a latent variable u,

$$\mathsf{p}(\gamma \mid \pi) = \pi + (1-\pi)\beta(\gamma; a, b), \quad \mathsf{p}(\gamma, u \mid \pi) = 1_{(u<\pi)} + 1_{(u>\pi)}\beta(\gamma; a, b).$$

If we integrate over γ we get $\mathsf{p}(u \mid \pi) = \{1_{(u<\pi)} + 1_{(u>\pi)}\} \sim \mathrm{Bern}(\pi)$, and then the full conditional of π is

$$\mathsf{p}(\pi \mid \text{else}) = \beta(\gamma; a + |r_i|, b + i - |r_i|)$$

where $r_i = \{j \leqslant i : \gamma^{(j)} = 1\}$. Algorithm 1 summarizes the inference procedure based on the derivations above.

Algorithm 1: Inference for structure coefficient of tail dependence

1 Initialize $(\gamma^{(1)}, \pi^{(1)})$.
2 Sample $\gamma^* \sim \pi^{(i)} + (1 - \pi^{(i)})\text{TN}(\gamma^{(i)}, 1)$.
3 Compute the ratio $R = \mathsf{p}(\gamma^*)q(\gamma \mid \gamma^*)/\mathsf{p}(\gamma)q(\gamma^* \mid \gamma)$, and accept $\gamma^{(i+1)} = \gamma^*$ with probability $\min\{R, 1\}$, else $\gamma^{(i+1)} = \gamma^{(i)}$.
4 Sample $\pi^{(i)}$ from $\text{Beta}(a + |r_i|, b + i - |r_i|)$, where $r_i = \{j \leqslant i : \gamma^{(j)} = 1\}$.
5 Repeat Steps 2 through 4 until reaching stationarity.

Time-varying Product

Here we describe the posterior inference algorithm to learn about the structure coefficient of tail dependence, when the structure process is the time-varying product and the parameter space of γ is $\mathscr{G}_P = (0, d]$. The same line of attack as in Section 5.9.3 can once more be used provided that we change the support of $\mathsf{p}(\gamma)$ and $q(\gamma^* \mid \gamma)$ to $(0, d)$, and set

$$\mathsf{p}(\gamma \mid \pi) = \pi + (1 - \pi)\,\mathsf{p}(\gamma), \quad \mathsf{p}(\pi) = \beta(\pi; a_\pi, b_\pi).$$

That is, the posterior inference algorithm in this case is tantamount to Algorithm 1 but in Step 2 the truncated Normal (TN) would be over $(0, d)$, and Step 3 requires that $\mathsf{p}(\gamma)$ is supported over $(0, d)$. For other structure processes with a different parameter space Steps 2–3 of Algorithm 1 can also be adapted in a similar fashion.

Part III—Final Observations

Chapter 6

Closing Remarks and Discussion

In this chapter, we will recapitulate the main contributions made on Chapter 4 and Chapter 5, as well as opportunities for future research.

6.1 Discussion

I have argued throughout this work, the importance of modelling heavy-tailed data using appropriate random probability measures, and how characterizing their tails is crucial to doing so. In Chapter 4 we developed new theory regarding the tail of an important type of Pitman–Yor process, the so-called stable law process. The main novelty of this chapter was to shown that the stable law process will be heavy-tailed, in the sense of regular variation, if and only if, the center measure is also heavy-tailed. Another important contribution of this chapter was the derivation of the exact envelopes where the tail of the stable process lies. To our understanding, this is the first time there is a result concerning the exact rate of decay of this random measure, as well as envelopes for its tail. Exploding the theoretical findings, we were able to compare two big classes of mixture models for heavy-tailed distributions, one class based on shape mixtures and the other class based on scale mixtures. Namely, we identify a clear preference for stable process scale mixtures over shape mixtures of heavy-tailed kernels by showing that shape mixtures of Pareto-type kernels can be super heavy-tailed even though the centering is only heavy-tailed; in other words, that it can lead to an overestimation of the mass at the tail. Outside the principal contribution of this chapter, smaller contributions were made. For instance, tail indices or other statistical functionals can be readily inferred from the proposed methods, which provides yet another possible application of the methodologies.

In this thesis, I have mentioned the challenges of modelling heavy-tailed data, specially when data is nonstationary. In particular, the difficulty of tracking how the dependence between the extreme values evolves over time. In Chapter 5 we interfaced two main topics on extreme value theory, *heteroscedastic extremes* and *nonstationary multivariate extremes*, to addressed this

difficult problem. We proposed the scedasis structure function for different *risk sets* induced by an aggregation rule of common scale heavy-tailed random variables. We show how the proposed framework is suitable for modelling *nonstationary multivariate extremes*, as it has been designed for tracking the dynamics governing the frequency of extreme values in a risk set. Whilst the work presented in this chapter was mainly motivated by the problem of tracking the dependence, we found interesting links between the structure scedasis and tail index of different aggregation rules, which shed light on the regularly variation property of different aggregation rules. We developed Bayesian nonparametric inference for the structure scedasis, based on Mixture of Polya trees centering on a distribution with support on the unit interval and show how this model do not suffer from boundary bias. Another important contribution is the construction of a public package called `extremis`, available on CRAN, were the proposed methods were implemented.

6.2 Future work

Heavy-tailed analysis is an ever-evolving field and there is always room for future developments, so I will comment on potential opportunities for future research.

Contribution 1: Heavy-Tailed Pitman–Yor Mixture Models

There are some open questions that are of interest for future analysis. For instance, the characterization of the tail of the Pitman–Yor process for the case when $M \neq 0$ and $0 < D < 1$. We conjecture that it will not be heavy-tailed as the subordinator process have an exponential term, however this answer is not simply due to the complexity of the subordinator-type representation. On a similar line of thought, another open question is, how large is the class of random measures that posses a heavy-tailed behavior and if we can characterized the rate of decay. Having an answer to these questions will provide flexibility to construct heavy-tailed models where we can provide the actual rate of decay of the tail and compare how adequate they are for heavy-tailed modelling. This opens a big area of intersection between Bayesian nonparametric theory and extreme value theory, and a potential area of development.

Contribution 2: Interfacing Nonstationary Multivariate Extremes and Heteroscedastic Extremes

The methodology proposed on this chapter were made under the assumption that the extreme value index is constant over time, it would be interesting to see what happened when we remove this assumption. Another assumption is the one of independence, which we think will be more complicated to remove as the extension of heteroscedastic extremes to the dependent case is still an open question. Another open question related to the scedasis function is the ability to model

the changepoints in this function, which will require to set a prior on the space of discontinuous scedasis functions and model the breaks as extra parameters. Finally, other future possible research is to provide methodology to model nonstationarities in margins and dependence at the same time, which to our knowledge there is potential development in this area.

Part IV—Supporting Materials

Appendix A

Supporting Materials for Chapter 4

A.1 Supporting Information for Section 2

A.1.1 Auxiliary Results for the Envelopes of Example 3

On the Domain of Attraction of Example 3

This section shows that the lower and upper envelopes in Example 3 based on a standard Pareto baseline are in the Gumbel domain of attraction—in the case of the DP—and in the Fréchet domain of attraction—in the case of the stable law process. The main result of this section is the following.

Proposition 7. *Suppose G_0 is a standard Pareto distribution, $1 - G_0(y) = y^{-1}$, for $y > 1$. For large y:*

(a) If $G \sim DP(M, G_0)$, then $1 - G$ is in the Gumbel domain of attraction, almost surely.

(b) If $G \sim SP(D, G_0)$, then $1 - G$ is in the Fréchet domain of attraction, almost surely.

The proof of Proposition 7 will be presented below, and it uses the following elementary lemma, which for completeness, we also prove below.

Lemma 4. *If $1 - F(y) = \exp\{-y^\tau \mathscr{L}(y)\}$ where $\tau > 0$ and $\mathscr{L}$ is a slowly varying function, then F belongs to the Gumbel domain of attraction.*

Proofs of Proposition 1 and Lemma 1

Proof of Lemma 4. Consider the auxiliary function $b(y) = y^{1-\tau}\mathscr{L}(y)^{-1}\tau^{-1}$, and note that

$$
\begin{aligned}
\lim_{y\to\infty}\frac{b(y)}{b(y+vb(y))} &= \lim_{y\to\infty}\frac{y^{1-\tau}\mathscr{L}(y)^{-1}\tau^{-1}}{\{y+vb(y)\}^{1-\tau}\mathscr{L}\{y+vb(y)\}^{-1}\tau^{-1}} \\
&= \lim_{y\to\infty}\frac{y^{1-\tau}\mathscr{L}(y)^{-1}}{[y+v\{y^{1-\tau}\mathscr{L}(y)^{-1}\tau^{-1}\}]^{1-\tau}\mathscr{L}\{y+vb(y)\}^{-1}} \\
&= \lim_{y\to\infty}\frac{y^{1-\tau}\mathscr{L}(y)^{-1}}{y\{1+y^{-\tau}\mathscr{L}(y)^{-1}\tau^{-1}\}^{1-\tau}\mathscr{L}\{y+vb(y)\}^{-1}} \\
&= \lim_{y\to\infty}\frac{y^{1-\tau}\mathscr{L}(y)^{-1}}{y^{1-\tau}\{1+y^{-\tau}\mathscr{L}(y)^{-1}\tau^{-1}\}^{1-\tau}\mathscr{L}\{y+vb(y)\}^{-1}} \\
&= \lim_{y\to\infty}\frac{\mathscr{L}(y)^{-1}}{\{1+y^{-\tau}\mathscr{L}(y)^{-1}\tau^{-1}\}^{1-\tau}\mathscr{L}\{y+vb(y)\}^{-1}} \\
&= 1,
\end{aligned}
$$

since,

$$
\lim_{y\to\infty}\frac{\mathscr{L}(y)^{-1}}{\mathscr{L}\{y+vb(y)\}^{-1}} = \lim_{y\to\infty}\frac{\mathscr{L}[y+v\{y^{1-\tau}\mathscr{L}(y)^{-1}\tau^{-1}\}]}{\mathscr{L}(y)} = \lim_{y\to\infty}\frac{\mathscr{L}[y\{1+vy^{-\tau}\mathscr{L}(y)^{-1}\tau^{-1}\}]}{\mathscr{L}(y)} = 1.
$$

Then, we can write the limit of interest as follows,

$$
\lim_{y\to\infty}\frac{1-F(y+b(y)t)}{1-F(y)} = \lim_{y\to\infty}\exp\left\{-\int_y^{y+tb(y)}\frac{1}{b(u)}\mathrm{d}u\right\}.
$$

Let $v=(u-y)/b(y)$ so that $u=y+vb(y)$; then, the integral above can be rewritten with the latter change of variables, and since the integrand converges locally uniformly to 1, $b(y)/b(y+vb(y))\to 1$, it follows that

$$
\begin{aligned}
\lim_{y\to\infty}\frac{1-F(y+b(y)t)}{1-F(y)} &= \lim_{y\to\infty}\exp\left\{-\int_0^t\frac{b(y)}{b(y+vb(y))}\mathrm{d}v\right\} \\
&= \exp\left\{-\int_0^t\lim_{y\to\infty}\frac{b(y)}{b(y+vb(y))}\mathrm{d}v\right\} \\
&= \exp\left\{-\int_0^t 1\,\mathrm{d}v\right\} \\
&= \exp\{-t\}.
\end{aligned}
$$

This proves that F is in the Gumbel domain of attraction. □

Proof of Proposition 7.

(a) If $G\sim \mathrm{DP}(M,G_0)$, then for large y:

$$
\exp\{-ys/M\log|\log(M/y)|\}\leqslant 1-G(y)\leqslant\exp\left[-y/\{M|\log(M/y)|^r\}\right],\quad \text{a.s.}
$$

For the upper bound, note that $\mathscr{L}_U(y)=1/\{M|\log(M/y)|^r\}$ and $\mathscr{L}_L(y)=s/M\log|\log(M/y)|$

are slowly varying functions. We can also see that for very large y, $y > M$, $|\log(M/y)| = -\log(M/y) = \log(y/M)$, then

$$\begin{aligned}\lim_{y\to\infty}\frac{\mathscr{L}_U(ty)}{\mathscr{L}_U(y)} &= \lim_{y\to\infty}\frac{1/\{M\log(yt/M)^r\}}{1/\{M\log(y/M)^r\}}\\ &= \lim_{y\to\infty}\frac{\log(y/M)^r}{\log(yt/M)^r}.\end{aligned}$$

Let $\ell_U(y) = \log(y/M)$, then

$$\begin{aligned}\lim_{y\to\infty}\frac{\ell_U(ty)}{\ell_U(y)} &= \lim_{y\to\infty}\frac{\log(yt/M)}{\log(y/M)} = \frac{\infty}{\infty}, \quad \text{L'hopital,}\\ &= \lim_{y\to\infty}\frac{1/y}{1/y}\\ &= 1.\end{aligned}$$

Recall that if $\mathscr{L}$ is a slowly varying function, then $\mathscr{L}^\alpha$ is slowly varying as well for all $\alpha \in \mathbb{R}$, and in particular for $\alpha = r$. In addition, note that $\ell_U(y) = \log(y/M)$ is a slowly varying function, and hence $\lim_{y\to\infty}\mathscr{L}_U(ty)/\mathscr{L}_U(y) = 1$. For the lower bound consider,

$$\lim_{y\to\infty}\frac{\mathscr{L}_L(yt)}{\mathscr{L}_L(y)} = \lim_{y\to\infty}\frac{\log\{\log(y/M)\}}{\log\{\log(yt/M)\}} = \frac{\infty}{\infty}, \quad \text{L'hopital,} \tag{A.1}$$

$$= \lim_{y\to\infty}\frac{\mathscr{L}_L'(ty)}{\mathscr{L}_L'(y)} \tag{A.2}$$

$$= \lim_{y\to\infty}\frac{\log(yt/M)}{\log(y/M)} = \frac{\infty}{\infty}, \quad \text{L'hopital,} \tag{A.3}$$

$$= \lim_{y\to\infty}\frac{1/y}{1/y} \tag{A.4}$$

$$= 1. \tag{A.5}$$

Note that (A.1) is satisfied for any values of r and s. Then, for large y, we can rewrite the bounds in terms of upper and lower slowly varying functions as follows,

$$\exp\{-y\mathscr{L}_L(y)\} \leqslant 1 - G(y) \leqslant \exp\{-y\mathscr{L}_U(y)\} \quad \text{a.s.}$$

Thus, the tail $1 - G(y)$ is bounded from both sides by $1 - F_l(y) = \exp\{-y\mathscr{L}_L(y)\}$ and $1 - F_u(y) = \exp\{-y\mathscr{L}_U(y)\}$, where by Lemma 4, F_l and F_u are in the Gumbel domain of attraction. Hence, there exist auxiliary functions $u(y)$ and $l(y)$ such that,

$$\lim_{y\to\infty}\frac{1 - G(y + u(y)t)}{1 - F_u(y)} \leqslant e^{-t},$$

and,

$$\lim_{y\to\infty}\frac{1 - G(y + l(y)t)}{1 - F_l(y)} \geqslant e^{-t},$$

where $u(y) = \mathscr{L}_U(y)^{-1}$ and $l(y) = \mathscr{L}_L(y)^{-1}$. Thus, we have that almost surely,

$$1 - G(y) = \exp\{-y^{1+o(1)}\}.$$

Hence, by defining an auxiliary function $b(y) = y^{-o(1)}$, satisfying $b(y)/b(y + vb(y)) \to 1$, it follows that

$$\lim_{y\to\infty} \frac{1 - G(y + b(y)t)}{1 - G(y)} = e^{-t},$$

and thus it finally follows from Proposition 2.1 in Beirlant et al. (2004) that $1 - G$ is in the Gumbel domain of attraction, almost surely.

(b) The result follows directly from Theorem 3 in the main chapter, as G_0 has a regularly varying tail. Indeed, G has a regularly varying tail with tail index $1/D$, and hence $1 - G$ is in the Fréchet domain of attraction.

□

A.2 Posterior Inference Algorithms

Here we develop the posterior inference algorithm to learn about the proposed heavy-tailed mixture models. For generality, we will focus on the versions from Section 3, but trivially this algorithm can be updated to fit the other models on the chapter. The numerical procedure to be discussed below is based on Gibbs sampling. Specifically, we propose to use the slice sampler algorithm for infinite mixtures proposed by Walker (2007), and further developed by Kalli et al. (2011). The proposal in Walker (2007) adapts the number of components in the mixture according to data complexity, and—conditional on the number of components at each iteration—the posterior inference is straightforward.

For concreteness, below we focus on the model in Equation (12), but similar comments readily apply to the multivariate heavy-tailed shape mixtures in Equation (13) of the main chapter. Let $\{\mathbf{y}_i\}_{i=1}^n$ be a random sample, with $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,d})^{\mathrm{T}}$. The joint likelihood for $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ is

$$f_{\pi,\sigma,\eta_\sigma}(\mathbf{y}) = \prod_{i=1}^{n} \sum_{h=1}^{\infty} \pi_h \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,h}}), \tag{A.6}$$

where $\pi = \{\pi_j\}_{j\geqslant 1}$ and $\sigma = \{\boldsymbol{\sigma}_j\}_{j\geqslant 1}$ are respectively the infinite collections of weights and d-dimensional atoms, and η_σ denotes the remainder parameters. Let s_i be a latent variable such that

$$(\mathbf{y}_i \mid s_i = j) \sim \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,j}}),$$

where $j = 1, 2, \ldots$ and $i = 1, \ldots, n$. The joint likelihood of $\mathbf{y}$ and $\mathbf{s} = (s_1, \ldots, s_n)$ is

$$f_{\pi,\sigma,\eta_\sigma}(\mathbf{y}, \mathbf{s}) = \prod_{j=1}^{\infty} \pi_j^{n_j} \prod_{\{i:s_i=j\}} \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,j}}), \tag{A.7}$$

where $\pi_j = P(s_i = j)$ and $n_j = \sum_{i=1}^{n} I(s_i = j)$. To avoid the computation of infinite terms in (A.7), Walker (2007) proposes to use the latent variables $\{u_i\}_{i=1}^{n}$, such that:

$$f_{\pi,\sigma,\eta_\sigma}(\mathbf{y}_i, u_i) = \sum_{j=1}^{\infty} I(u_i < \pi_j) \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,j}}). \tag{A.8}$$

In (A.8) only a finite number N of π_j's satisfies the condition $\{u_i < \pi_j\}$. In particular, $N = \max_i\{N_i\}$ and N_i is the smallest integer such that $\sum_{j=1}^{N_i} \pi_j > 1 - u_i$, see Walker (2007) and Kalli et al. (2011) for more details. Note that this implies that now $\pi = \{\pi_j\}_{j=1}^{N}$ and $\sigma = \{\boldsymbol{\sigma}_j\}_{j=1}^{N}$. Considering the latent variables $\mathbf{s}$ and $\mathbf{u} = (u_1, \ldots, u_n)$, the likelihood for one observation is

$$f_{\pi,\sigma,\eta_\sigma}(\mathbf{y}_i, s_i, u_i) = I(u_i < \pi_{s_i}) \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,s_i}}),$$

and hence the likelihood for the n observations can be expressed as

$$f_{\pi,\sigma,\eta_\sigma}(\mathbf{y}, \mathbf{s}, \mathbf{u}) = \prod_{i=1}^{n} I(u_i < \pi_{s_i}) \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,s_i}}). \tag{A.9}$$

Using the likelihood in (A.9), the posterior inference via a Gibbs sampler is straightforward. At each iteration, it is necessary to update π,σ and η_σ, along with the latent variables $\mathbf{s}$ and $\mathbf{u}$. The weights π_j will be computed with the stick-breaking construction, and then the updates will be performed for the sticks V_j.

Following the recommendations in Section 3 of the chapter (Remark 1), we suggest implementing the algorithm using as a baseline a multivariate heavy-tailed distribution with Pareto Type II margins; see Equation (15). Some details on posterior inference for the extreme value index are in order. Assuming the Jeffrey's prior $\mathsf{p}(\alpha_k) = 1/\alpha_k$, it follows that the posterior distribution is

$$\mathsf{p}(\alpha_k \mid \text{else}) \propto f(\sigma \mid \alpha_k)\mathsf{p}(\alpha_k) = \frac{1}{\alpha_k} \prod_{h=1}^{N} \beta^{\alpha_k} \alpha_k (\beta + \sigma_h)^{-\alpha_k - 1} \propto \text{Gamma}\left\{N, \sum_{h=1}^{N} \log\left(\frac{\beta + \sigma_h}{\beta}\right)\right\}.$$

Algorithm 2 shows how to learn from data about the stable process scale mixture models in Section 3.1 of the chapter. Step 6 of Algorithm 2 is not conjugate, and hence a Metropolis–Hastings step is required, which is implemented with a random walk strategy proposing candidates D^* from a Beta distribution. Algorithm 3 shows how to learn about the predictor dependent model (16) in Section 3.2 of the chapter, from a random sample $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{n}$, where the co-

variate $\mathbf{x}_i \in \mathbb{R}^p$ for all i. The algorithm is tantamount to Algorithm 2, except for an additional step to update parameters $\boldsymbol{\beta}_j$ and minor changes in the updating steps of the sticks and latent variables. To update $\boldsymbol{\beta}_j$ a Metropolis–Hastings step is required, which was implemented with a random walk proposing candidates $\boldsymbol{\beta}_j^*$ from multivariate Normal distribution in $\mathbb{R}^p$. For Step 3, a Metropolis–Hastings step might also be required—depending on the kernel to be used. The Erlang kernel is particularly convenient as Steps 3 and 8 can be performed using a Gibbs sampler. For example, the version of the model in Remark 1 of the chapter can be fitted using Algorithms 1 and 2 with the following details for Steps 3 and 8:

3 Sample $\boldsymbol{\sigma}_j$ from $\mathtt{p}(\boldsymbol{\sigma}_j \mid \text{else}) \propto G_0(\boldsymbol{\sigma}_j) \prod_{\{i:s_i=j\}} \prod_{k=1}^{d} \text{Er}(y_{i,k}; \lceil \sigma_{k,s_i} \rceil, \sigma_{k,s_i}/\lambda)$.

8 Sample λ from $\text{Gamma}\big(a_\lambda + \sum_{i=1}^{n}\sum_{k=1}^{d} \lceil \sigma_{k,s_i} \rceil\,,\; b_\lambda + \sum_{i=1}^{n}\sum_{k=1}^{d} y_{i,k}/\sigma_{k,s_i}\big)$.

Here, $\text{Er}(y; a, b)$ is the density of the Erlang distribution with shape $a \in \mathbb{N}$ and scale $b > 0$, and (8) is implemented with a random walk strategy proposing candidates $\boldsymbol{\sigma}_j^*$ from a multivariate truncated Normal distribution in $\mathbb{R}_+^d$. Some final comments on notation are in order. In Algorithms 2 and 3, $\text{B}(\cdot \mid a, b)$ is the density of a $\text{Beta}(a, b)$ distribution, and $D_{i,j} \equiv \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j)/\{1 + \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j)\}$.

Algorithm 2: Slice Sampler for Stable Process Scale Mixtures Eq. (12)

1 Initialize $N, \boldsymbol{\sigma}_1^{(0)}, \ldots, \boldsymbol{\sigma}_N^{(0)}, \eta_\sigma^{(0)}, V_1^{(0)}, \ldots, V_N^{(0)}, \mathbf{u}^{(0)}$ and $\mathbf{s}^{(0)}$.

2 Sample α_k for each coordinate k, from $\text{Gamma}(N, \sum_{h=1}^{N} \log\{(1 + \sigma_{k,h})\})$.

3 Sample $\boldsymbol{\sigma}_j$ from $\mathtt{p}(\boldsymbol{\sigma}_j \mid \text{else}) \propto G_0(\boldsymbol{\sigma}_j) \prod_{\{i:s_i=j\}} \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,s_i}})$.

4 Sample V_j from $\text{Beta}\big(1 - D + \sum_{i=1}^{n} I(s_i = j)\,, jD + \sum_{i=1}^{n} I(s_i > j)\big)$, set $\pi_j = V_j \prod_{\ell<j}(1 - V_\ell)$.

5 Sample u_i from $(u_i \mid \text{else}) \sim \text{Unif}(0, \pi_{s_i})$, for $i = 1, \ldots, n$. Then, set N as the smallest integer for which $\sum_{j=1}^{N} \pi_j > 1 - u^*$, where $u^* = \min_i\{u_i\}$.

6 Sample D from $\mathtt{p}(D \mid \text{else}) \propto \prod_{d=i}^{N} \text{B}(v_i \mid 1 - D, iD)\, \text{B}(D \mid a_D, b_D)$.

7 With probability $\mathtt{p}(s_i = j \mid \text{else}) \propto I(j : \pi_j > u_i) \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,s_i}})$, set $s_i = j$ for $i = 1, \ldots, n$.

8 Update any remainder parameters in η_σ via Metropolis–Hastings.

9 Repeat Steps 2 through 7 until reaching stationarity.

A.3 The `pityoR` Package

In this section, we sketch some details on the R package `pityoR`. The scale mixture models from Sections 2 and 3 in the chapter can be fitted using the following functions:

- `SPmix`: Fits the univariate scale stable mixture model with an Erlang kernel and centered on a Pareto Type II distribution; see Example 4 in the chapter.

Algorithm 3: Slice Sampler for Conditional Stable Process Scale Mixtures Eq. (14)

1 Initialize $N, \boldsymbol{\sigma}_1^{(0)}, \ldots, \boldsymbol{\sigma}_N^{(0)}, \eta_\sigma^{(0)}, V_1^{(0)}, \ldots, V_N^{(0)}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N, \mathbf{u}^{(0)}$ and $\mathbf{s}^{(0)}$.

2 Sample α_k for each coordinate k, from $\text{Gamma}(N, \sum_{h=1}^{N} \log\{(1+\sigma_{k,h})\})$.

3 Sample $\boldsymbol{\sigma}_j$ from $\mathtt{p}(\boldsymbol{\sigma}_j \mid \text{else}) \propto G_0(\boldsymbol{\sigma}_j) \prod_{\{i:s_i=j\}} \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,s_i}})$.

4 Sample $V_j(\mathbf{x}_i)$ from $\text{Beta}\big(1 - D_{i,j} + I(s_i = j)\,, jD_{i,j} + I(s_i > j)\big)$ and set $\pi_j(\mathbf{x}_i) = V_j(\mathbf{x}_i) \prod_{\ell<j}\{1 - V_\ell(\mathbf{x}_i)\}$, for $i = 1, \ldots, n$ and $j = 1, \cdots, N$.

5 Sample $\boldsymbol{\beta}_j$ from $\mathtt{p}(\boldsymbol{\beta}_j \mid \text{else}) \propto \exp\left(\boldsymbol{\beta}_j^{\mathrm{T}}[s^2\mathbf{I}_p]^{-1}\boldsymbol{\beta}_j\right) \prod_{\{i:s_i=j\}} \text{B}(V_j(\mathbf{x}_i) \mid 1 - D_{i,j}, jD_{i,j})$.

6 Sample u_i from $(u_i \mid \text{else}) \sim \text{Unif}(0, \pi_{s_i}(\mathbf{x}_i))$, $i = 1, \ldots, n$. Then set N as the smallest integer for which $\sum_{j=1}^{N} \pi_j(\mathbf{x}_i) > 1 - u^*$, where $u^* = \min_i\{u_i\}$.

7 With probability $\mathtt{p}(s_i = j \mid \text{else}) \propto I(j : \pi_j(\mathbf{x}_i) > u_i) \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,s_i}})$, set $s_i = j$; $i = 1, \ldots, n$.

8 Update any remainder parameters in η_σ via Metropolis–Hastings.

9 Repeat steps 2 through 7 until reaching stationarity.

- `SPmix_multi`: Function to compute the multivariate scale stable mixture model of Erlang kernel, centered on a multivariate heavy-tailed distribution with Pareto Type II margins; see Section 3.1 in the chapter.

- `SPmix_cond`: Fits the conditional multivariate scale stable conditional mixture model, centered on a multivariate heavy-tailed distribution with Pareto Type II margins; see Section 3.2 in the chapter.

As noted in p. 125 of this supplement the Erlang kernel presents some computational advantages; it also follows from p. 125 that the codes made available in the package can be readily adapted to other kernels.

The package also contains commands for fitting the shape mixture model presented in Section 5 of the chapter, and a command to simulate trajectories from the Pitman–Yor process along with the asymptotic envelopes that stem from Theorem 2; namely:

- `DPmix`: Fits the univariate shape Dirichlet mixture model, centered on a Pareto Type II distribution; see Example 5 on the chapter.

- `rPYP`: Simulates trajectories from the Pitman–Yor process along with the corresponding envelopes given in Theorem 2 of the chapter. Our computational experience suggests that ε has to be set rather small, and clearly as a function of D (with larger values of D implying a considerably smaller ε).

Below we will illustrate the function `SPmix_multi` so to illustrate how to fit one of the simulation scenarios presented in Section 5 of the chapter (Bivariate Scenario 2); although below we focus on the bivariate case, the same command can be used to fit multivariate data. Before showing how to use `SPmix_multi`, we first load the required packages; as it will be illustrated below, any copula from the `copula` package can be used as a baseline by using the command `mvdc`.

```r
## Load required packages
packages <- c("copula", "pityoR")
sapply(packages, require, character.only = TRUE)
```

Before fitting the model, we set up the MCMC and specify the baseline as well as the prior information.

```r
## MCMC settings
mcmc <- list(nburn = 200, nsave = 200)
## Prior information
centering <-  mvdc(copula = archmCopula(family = "frank", param = 1),
                   margins = c("pareto", "pareto"),
                   paramMargins = list(list(shape = .5, scale = 1),
                                       list(shape = .5, scale = 1)))
prior <- list(a1 = .1, b1 = .1, aD = .5, bD = .5, centering = centering)
```

Next, we simulate data from the bivariate distribution from Bivariate Scenario 2 using the command `rvdc` from the package `copula`. For reproducibility reasons, we fix the seed using `set.seed`.

```r
set.seed(8452)
myMvd <- mvdc(copula = archmCopula(family = "frank", param = 1),
              margins = c("lgamma", "lgamma"),
              paramMargins = list(list(shapelog = 5, ratelog = 5),
                                  list(shapelog = 5, ratelog = 5)))
data <- rMvdc(1000, myMvd)
Y <- cbind(data[,1]-1, data[,2]-1)
```

Next, we learn about the proposed model from data using the function `SPmix_multi` from the package `pityoR`.

```r
fit <- SPmix_multi(Y = Y, prior, mcmc)
```

To plot the contour of the posterior distribution one can use `plot` for the object class `SPmix_multi`, that is:

```
plot(fit, which = "contour", xlim = c(0.1, 10), ylim = c(0.1, 10), data = TRUE)
```

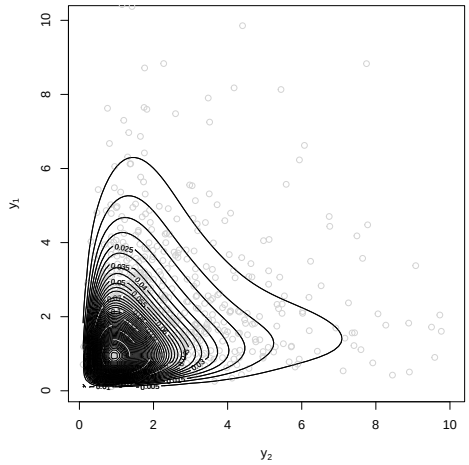


Figure A.1: Contours of joint density estimate obtained with the proposed stable process scale mixture model.

To compare the obtained fit against the true joint density, use the following code:

```
gridy1 <- gridy2 <- seq(0.1, 10, length.out = 100)
gridtrue <- expand.grid(gridy1, gridy2)
gg <- as.matrix(gridtrue)
true <- dMvdc(gg + 1, myMvd)
z <- matrix(true, ncol = 100, nrow = 100)
lvls <- pretty(range(z), 50)
par(pty = "s")
contour(gridy1, gridy2, z, levels = lvls, ylab = expression(y[2]),
        xlab = expression(y[1]), main = "")
```

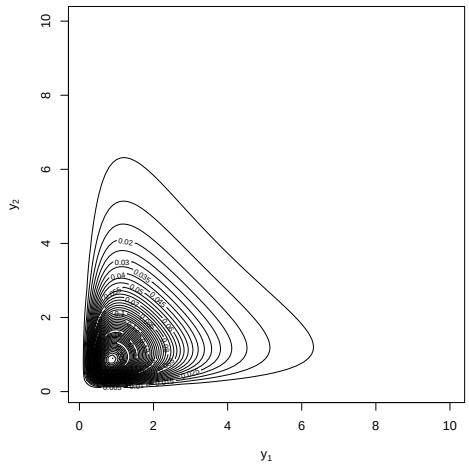


Figure A.2: True joint density.

The EVI estimated using the proposed model along with 90% credible intervals can be readily obtained as follows:

```
fit$evi.hat

## [1] 0.2039750 0.2891748

fit$evi.CI

##            95%        5%
## [1,] 0.1564305 0.2594535
## [2,] 0.2442174 0.3406467
```

For comparison, the true extreme value index of each $\mathrm{LG}(5,5)$ marginal is $1/5 = 0.20$.

Some final comments on other visualizations available from the `pityoR` package are in order. For the marginal distributions, the function `plot`, from the object class `SPmix_multi`, contains the following options:

- `which = "density"` plots the fitted marginal density for a specific component.
- `which = "logsurvival"` plots the log-survival for a specific the marginal component.
- `which = "qqplot"` shows the q-q plot of the marginal estimates for a specific component.
- `which = "qqboxplot"` shows the q-q boxplot (Rodu and Kafadar, 2022) of the randomized quantile residuals for the marginal estimates for a specific component.

Hence, for example:

```
plot(fit, which = "logsurvival", marginal = 1, bands = TRUE, xlim = c(0, 10))
```

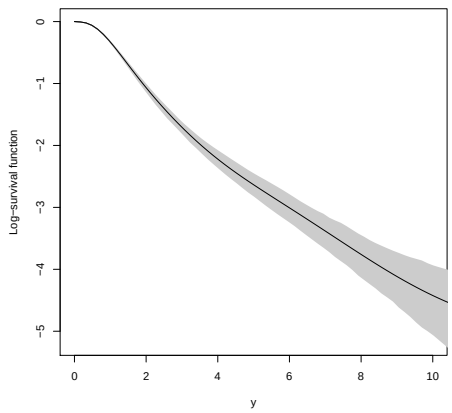


Figure A.3: Log-survival estimate for the marginal Y_1 obtained with the proposed stable process scale mixture.

As another example:

```
plot(fit, which = "qqboxplot", marginal = 1, bands = TRUE)
```

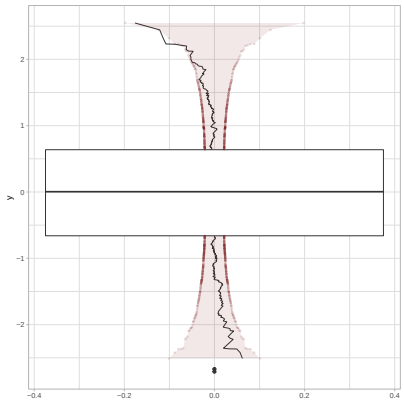


Figure A.4: q-q boxplot of randomized quantile residuals based on marginal density estimate of Y_1.

Appendix B

Supporting Materials for Chapter 5

B.1 Supplementary Simulation Studies

B.1.1 Selected Experiments with Kernel Density Estimator

The Monte Carlo experiments to be reported next illustrate: i) the boundary-bias issue of the structure scedasis kernel density estimator; ii) the better performance of the proposed method in comparison to the kernel approach. The structure scedasis kernel density estimator is defined as

$$\widehat{c}(s) = \frac{1}{k} \sum_{\{j:Z_{(t)}>Z_{(T-k)}\}} K_h\left(s - t/T\right), \tag{B.1}$$

where is the number of exceedances of $Z_t = \min(X_t, Y_t)$, $h > 0$ is the bandwidth, and $K_h(\cdot) = K(\cdot/h)/h$, with K denoting a kernel. For the experiments reported in Figure B.1 we use $h = 0.1$; a similar poor fit at the boundary would be visible using other bandwidth selection methods such as, for example, Silverman's rule of thumb (Silverman, 1986). Figure B.1 emphasizes the

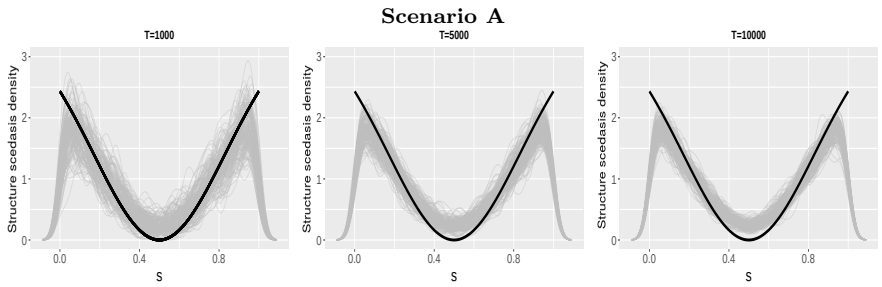


Figure B.1: Boundary-bias issue of the structure scedasis kernel density estimator; here the structure process is the time-varying minimum.

Table B.1: MISE for Scenarios A–C under the kernel density estimator and the Polya tree estimator.

Scenario A	$T = 500$	$T = 1\,000$	$T = 5\,000$	$T = 10\,000$
MISE kernel	0.2575	0.1390	0.0928	0.0853
MISE Polya	0.0207	0.0182	0.0120	0.0103
Scenario B	$T = 500$	$T = 1\,000$	$T = 5\,000$	$T = 10\,000$
MISE kernel	0.0261	0.0229	0.0078	0.0067
MISE Polya	0.0152	0.0106	0.0070	0.0057
Scenario C	$T = 500$	$T = 1\,000$	$T = 5\,000$	$T = 10\,000$
MISE kernel	0.0272	0.0186	0.0090	0.0086
MISE Polya	0.0265	0.0126	0.0089	0.0080

boundary-bias issue to which the structure scedasis plain vanilla kernel density estimator from (1) is subject to.

Next, to compare the performance of the estimation using Polya trees with the kernel estimation method, we compute the MISE (Mean integrated squared error) for Scenarios A–C under the different sample sizes, Table B.1 reports the results where we can see that the Polya tree method provides a better fit in general.

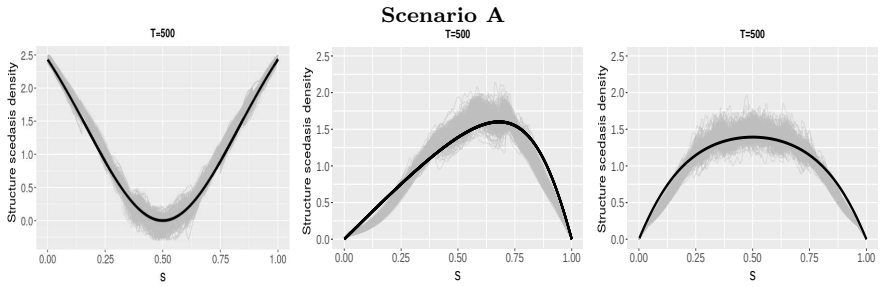


Figure B.2: Fitted scedasis density estimation under Scenarios A–C for $T = 500$, obtained using a mixture of finite Polya trees; here the structure process is the time-varying minimum.

B.1.2 Structure Scedasis of Radius

We now present the main findings from a Monte Carlo simulation study for the aggregation rule $\phi_R(x, y) = x + y$. We simulated 1 000 time series of length $T = 500, 1\,000, 5\,000$, and 10 000 from Scenarios A–C introduced in Section 5.1. A twin analysis to the one of the minimum is presented next, that is, a one shot experiment is presented in Figure B.3 and Figure B.4 shows the structure scedasis density estimates obtained with the proposed methods for Scenarios A–C over this Monte Carlo simulation study; we used the same prior specification as in Section 5.1. We can see that the estimation is overall good for all scenarios.

For Scenarios A–C, the true structure scedasis for the ratio is calculated using a result in (Falk, 2007, Theorem 3.3), which states that for marginal Fréchet distributions with parameter

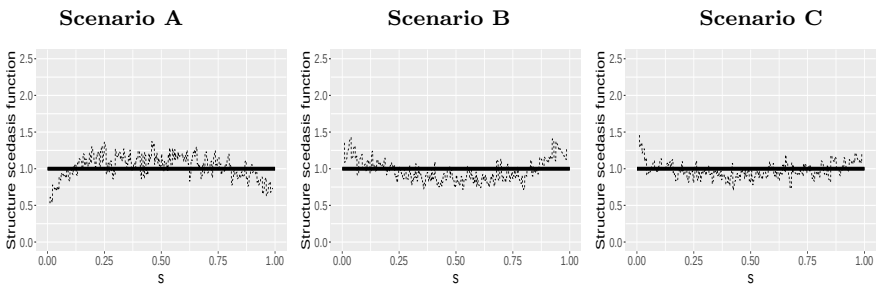


Figure B.3: Structure scedasis density estimate obtained via the posterior mean of a mixture of finite Polya trees over a single-run experiment (dashed) plotted against true (solid); here the structure process is the time-varying radius.

α, the ratio is regularly varying with index α and slowly varying function,

$$\mathcal{L}_R(z) = \int_0^1 \{1 - D(z) - D'(z)(1-z)\} z^{1/\alpha - 1} \frac{\{z^{1/\alpha} + (1-z)^{1/\alpha}\}^{\alpha-1}}{(1-z)^2} \, \mathrm{d}z + \{2 - M(1^-)\}, \tag{B.2}$$

with $D(z)$ the Pickands dependence function and $M(1^-) = \lim_{\varepsilon \to 0} M(1-\varepsilon)$ the limit from the left of M at 1 and $M(z) = D'(z) - 1$. In our case we have unit Fréchet distributions and time-varying dependence functions, then (B.2) becomes:

$$\mathcal{L}_R(z) = \int_0^1 \frac{\{1 - D_t(z) - D_t'(z)(1-z)\}}{(1-z)^2} \mathrm{d}z + \{2 - M_t(1^-)\}. \tag{B.3}$$

Then, under this setup if the baseline distribution is also unit Fréchet, the true structure scedasis is,

$$c_R^*\left(\frac{t}{T}\right) = \lim_{z \to \infty} \mathcal{L}_R(z).$$

We use numerical approximation to calculate (B.3) for the different dependence function in Scenarios A–C, which is very closed to the constant 1 in all scenarios.

B.1.3 Sensitivity Analysis

In this section we conduct a sensitivity analysis; the main results are tantamount to the ones obtained in Section 5 of the chapter. Figure B.5 shows the one shot experiment for Scenarios A–C with $T = 1000$, where we set the baseline distribution to be $C_{0,\theta} \sim \text{Beta}(a, b)$, and set $a \sim \text{Log-normal}(m_0, s_0)$ and $b \sim \text{Log-normal}(\tau_1, \tau_2)$; finally, we set $\alpha \sim \text{Gamma}(a_0, b_0)$. In terms of hyperparameters, we consider:

- Prior 1: $a_0 = .1$, $b_0 = .1$, $M = 8$, $m_0 = .5$, $s_0 = 1$, $\tau_1 = .01$, and $\tau_2 = .01$.
- Prior 2: $a_0 = 1$, $b_0 = 1$, $M = 8$, $m_0 = 1$, $s_0 = 1$, $\tau_1 = .01$, and $\tau_2 = .01$.

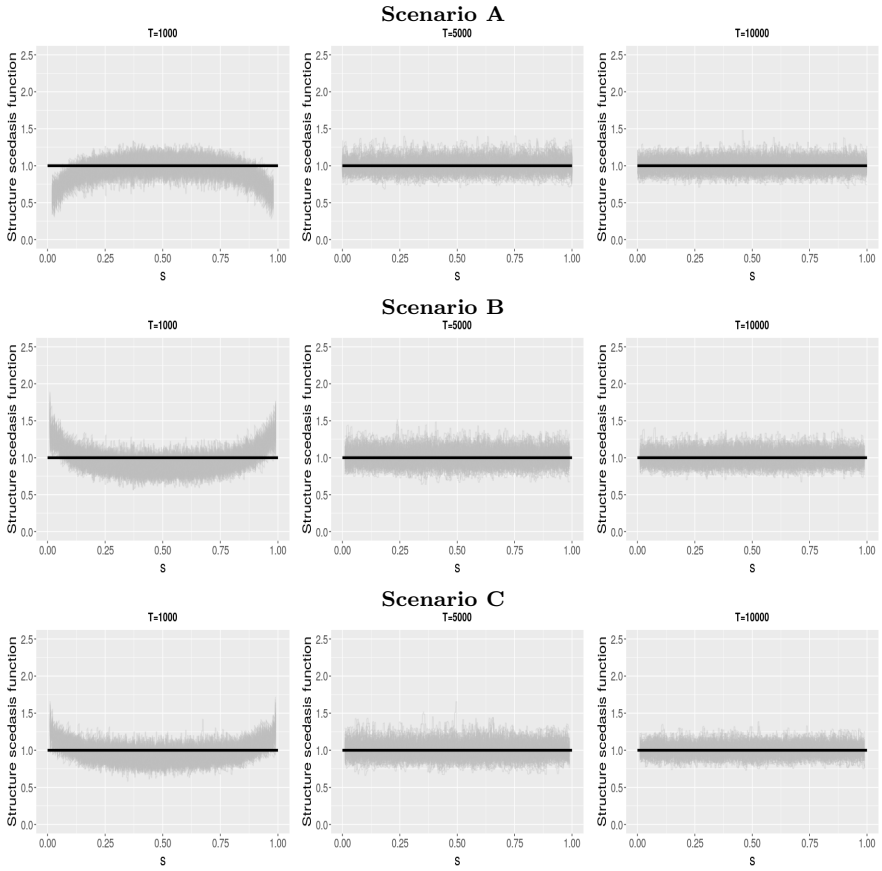


Figure B.4: First 150 trajectories of structure scedasis density estimates obtained via the posterior mean of a mixture of finite Polya trees over the Monte Carlo simulation study (gray) plotted against true (black); here the structure process is the time-varying radius.

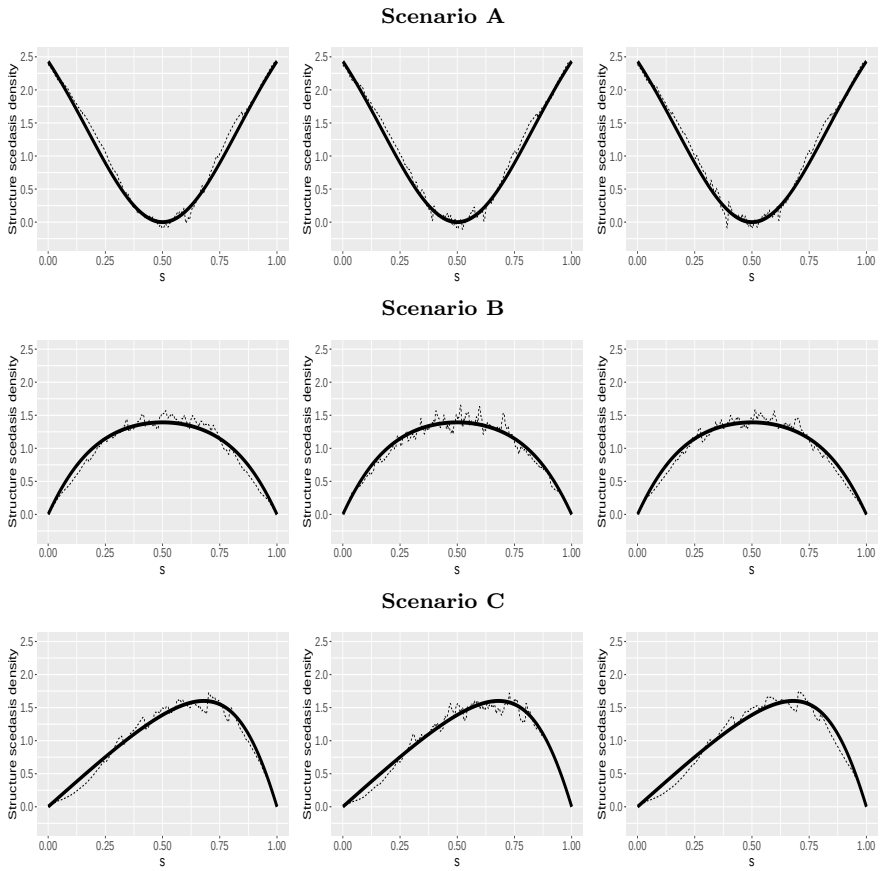


Figure B.5: Structure scedasis density estimate obtained via the posterior mean of a mixture of finite Polya trees over a single-run experiment (dashed) plotted against true (solid). For different prior specifications. Prior 1, 2 and 3 (left to right). Here, the structure process is the time-varying minimum.

- Prior 3: $a_0 = .1$, $b_0 = .1$, $M = 8$, $m_0 = \hat{\mu}$, $s_0 = \hat{\sigma}$, $\tau_1 = .01$, and $\tau_2 = .01$,

where$(\hat{\mu}, \hat{\sigma})$ is the maximum likelihood estimator of $\alpha \sim \text{Gamma}(a_0, b_0)$. The data are threshold at their 0.95% quantile, and we run a burn-in period of 5000 iterates after which we saved 5000 posterior iterates. Figure B.6 repost the Monte Carlo study. We simulate 1000 time series of length $T = 1000$ under the three scenarios and fit the proposed methods using three different prior configurations (Priors 1–3).

B.2 Additional Empirical Results

B.2.1 Time-varying Radius

We now examine the fits of the structure scedasis function for the FAANG stocks when the structure process is the time-varying radius, $Z_t = X_t + Y_t$. Figure B.7 reports the fits of the scedasis function for pairs of stocks, and Figure B.8 depicts the estimate for the sum of the five stocks. Table B.2 reports the posterior mean and credible intervals for the extreme value index

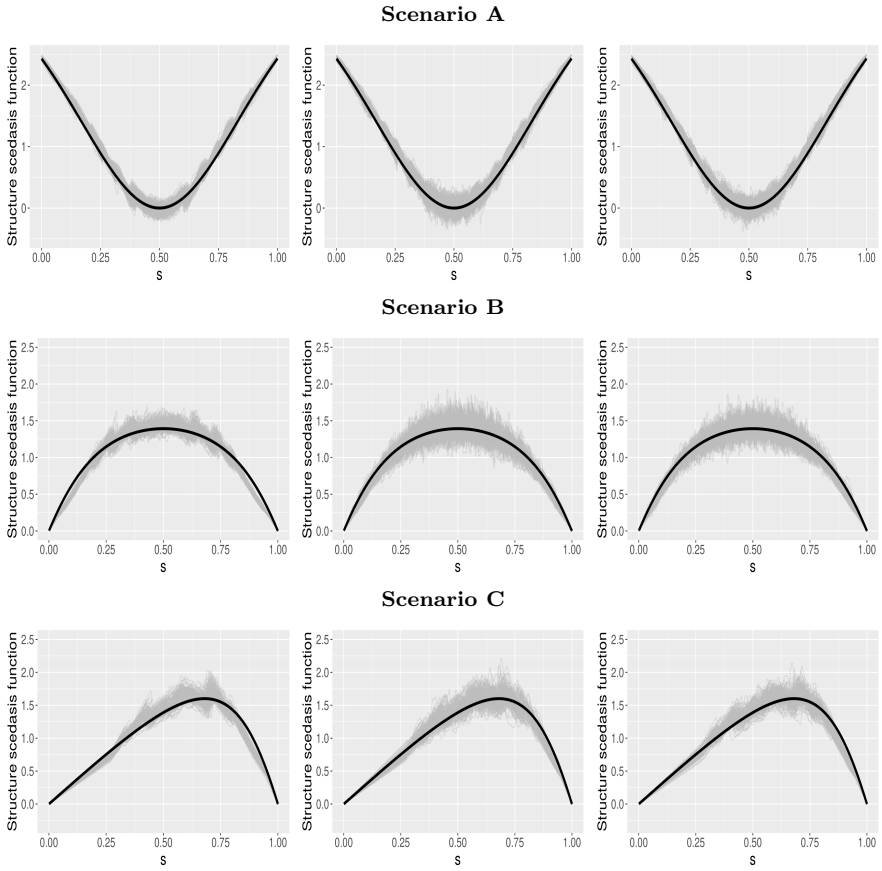


Figure B.6: First 150 trajectories of structure scedasis density estimates obtained via the posterior mean of a mixture of finite Polya trees over the Monte Carlo simulation study (gray) plotted against true (black). For different prior specifications. Prior 1, 2 and 3 (left to right). Here, the structure process is the time-varying minimum.

of the structure process; the posterior mean extreme value index for this multivariate analysis is 0.97 (CI = $(0.76, 1.29)$).

Table B.2: Posterior median and 95% credible intervals of the extreme value index of the structure process of the time-varying radius.

Pair of FAANG stocks	Lower limit	Posterior median	Upper limit
Facebook–Amazon	0.980	1.000	1.000
Facebook–Apple	0.991	1.000	1.000
Facebook–Netflix	1.000	1.000	1.000
Facebook–Google	1.000	1.000	1.000
Amazon–Apple	0.991	1.000	1.000
Amazon–Netflix	0.989	1.000	1.000
Amazon–Google	1.000	1.000	1.000
Apple–Netflix	1.000	1.000	1.000
Apple–Google	1.000	1.000	1.000
Netflix–Google	0.976	1.000	1.000

B.3 The `extremis` Package

In this section we present some notes on how to use the R (R Development Core Team, 2016) package *extremis* so to reproduce the experiments in the chapter. Before running the code chunks below, we start by cleaning workspace and installing the *extremis* packages (if not installed).

```
rm(list = ls())
if (!require("devtools")) install.packages("devtools")
require(devtools)
if (!require("extremis")) install_github("mmbbcarvalho/extremis")
require(extremis)
```

For reproducibility reasons, we fix setseed and list below the information about R, the OS, and loaded packages:

```
set.seed(12333)
sessionInfo()

## R version 4.0.3 (2022-05-10)
## Platform: x86_64-apple-darwin19.6.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /usr/local/Cellar/r/4.0.3/lib/R/lib/libRblas.dylib
## LAPACK: /usr/local/Cellar/r/4.0.3/lib/R/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/C/en_US.UTF-8/C/C/en_US.UTF-8
```

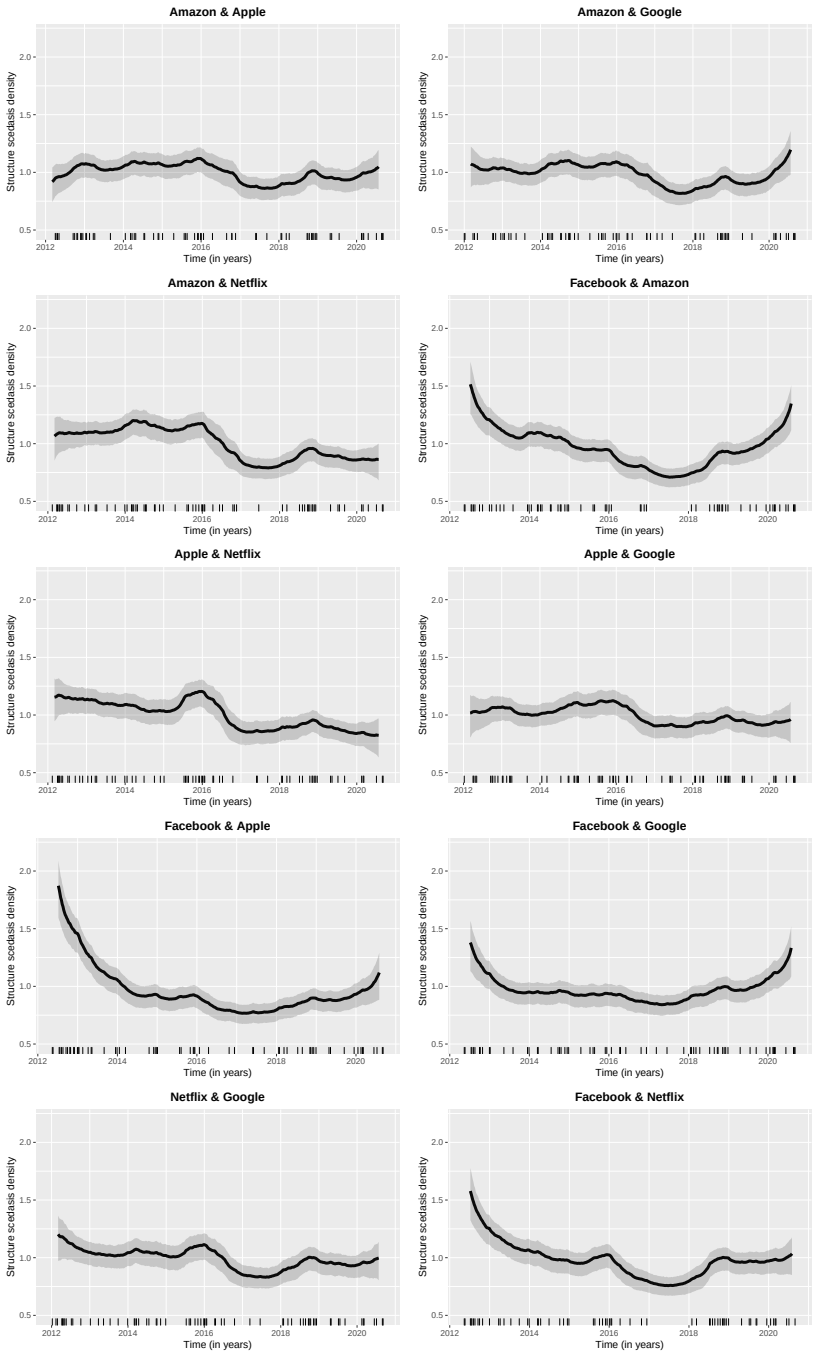


Figure B.7: Structure scedasis (for the pairwise setting) posterior mean along with pointwise credible bands using finite mixtures of Polya trees for FAANG stocks over 2012–2020; here, the structure process is the time-varying radius.

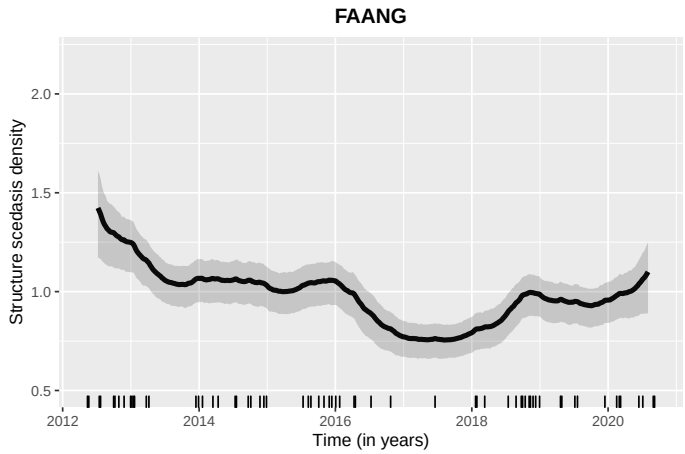


Figure B.8: Structure scedasis (for the multivariate setting) posterior mean along with pointwise credible bands using finite mixtures of Polya trees for FAANG stocks over 2012–2020; here, the structure process the time-varying radius.

```
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] extremis_1.1   devtools_2.3.2 usethis_1.6.3  knitr_1.30
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.5          compiler_4.0.3      highr_0.8           prettyunits_1.1.1
##  [5] remotes_2.2.0       tools_4.0.3         evd_2.3-3           testthat_3.0.0
##  [9] digest_0.6.27       pkgbuild_1.1.0      pkgload_1.1.0       evaluate_0.14
## [13] memoise_1.1.0       lattice_0.20-41     rlang_0.4.8         Matrix_1.2-18
## [17] cli_2.2.0           SparseM_1.78        xfun_0.19           withr_2.3.0
## [21] stringr_1.4.0       desc_1.2.0          fs_1.5.0            MatrixModels_0.4
## [25] rprojroot_2.0.2     grid_4.0.3          glue_1.4.2          R6_2.5.0
## [29] processx_3.4.4      fansi_0.4.1         emplik_1.1-1        sessioninfo_1.1.1
## [33] conquer_1.0.2       callr_3.5.1         magrittr_2.0.1      MASS_7.3-53
## [37] matrixStats_0.57.0 ps_1.4.0            ellipsis_0.3.1      assertthat_0.2.1
## [41] quantreg_5.75       stringi_1.5.3       crayon_1.3.4
```

Prior to running the code we start by specifying the prior and to setup the MCMC:

```
## Initial state
state <- NULL
## MCMC parameters
nburn <- 1000
nsave <- 1000
nskip <- 0
ndisplay <- 500
mcmc <- list(nburn = nburn, nsave = nsave, nskip = nskip, ndisplay = ndisplay,
```

```
            tune1 = 1.1, tune2 = 1.1, tune3 = 1.1)
## Prior information
prior <- list(a0 = .1, b0 = .1, M = 8, m0 = .01, S0 = .01, tau1 = .01, tau2 = .01)
```

We now simulate data from a bivariate extreme value distribution (Asymmetric Logistic, Scenario C in the chapter) using the command `rbevd` from the package evd (Stephenson, 2002).

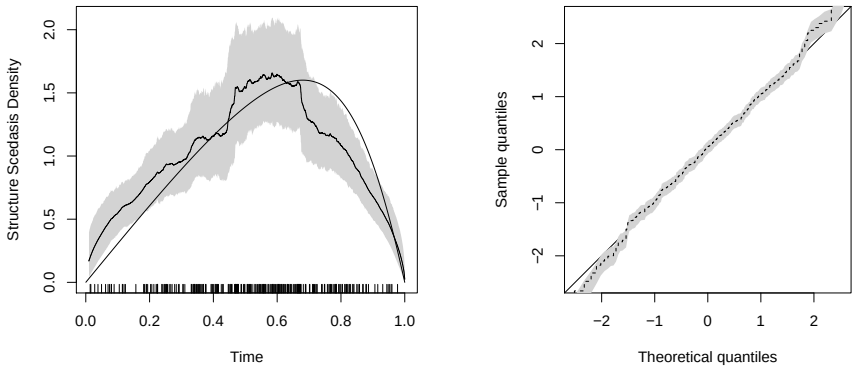
```
require(evd)
T <- 5000
time <- seq(1/T, 1, by = 1/T)
aux <- matrix(0, T, 2)
for (i in 1:T) {
    aux[i, ] <- rbvevd(1, dep = .5, asy = c(sin(pi * time[i]), time[i]),
                       model = "alog", mar1 = c(1, 1, 1), mar2 = c(1, 1, 1))
}
XY <- cbind(time, aux)
```

Next, we fit the structure scedasis function with a finite mixture of Polya trees by using the command `cPTdensity`:

```
fit <- cPTdensity(XY, prior = prior, mcmc = mcmc, state = state, status = TRUE)

##
## MCMC scan 500 of 1000 (CPU time: 30.071 s)
## MCMC scan 1000 of 1000 (CPU time: 54.247 s)

plot(fit, CI = TRUE)
## The code below plots the true structure scedasis function
aux1 <- numeric()
for (i in 1:T) {
  alpha <- .5
  t2 <- time[i]
  t1 <- sin(pi*time[i])
  aux1[i] <- (t1 + t2) - ((t1^(1/alpha)) + (t2^(1/alpha)))^alpha
}
func <- function(x) {
  alpha <- .5
  t1 <- sin(pi*x)
  t2 <- x
  (t1 + t2) - ((t1^(1/alpha)) + (t2^(1/alpha)))^alpha
};
integral <- integrate(func, 0, 1)$value;s<-aux1/integral;
lines(time, s)
## The code below plots the randomized quantiles
plot(fit, rquantiles = TRUE)
```



The package *extremis* also includes the command `cdensity` for fitting the structure scedasis density via kernel methods. Any kernel method available from the R command `density` can be used as well as any standard bandwidth selection method.

As a final note, if a single variable is entered into the commands above, then the package takes the aggregation rule to be the identity function, and estimates a scedasis functions for the setup in Einmahl et al. (2016).

Bibliography

Albrecher, H., Asmussen, S., and Kortschak, D. (2006), "Tail asymptotics for the sum of two heavy-tailed dependent risks," *Extremes*, 9, 107–130.

Albrecher, H., Beirlant, J., and Teugels, J. L. (2017), *Reinsurance: actuarial and statistical aspects*, John Wiley & Sons.

Aloui, R., Hammoudeh, S., and Nguyen, D. K. (2013), "A time-varying copula approach to oil and stock market dependence: The case of transition economies," *Energy Economics*, 39, 208–221.

Applebaum, D. (2009), *Lévy Processes and Stochastic Calculus*, Cambridge: CUP.

Arbel, J., De Blasi, P., and Prünster, I. (2019), "Stochastic Approximations to the Pitman–Yor Process," *Bayesian Analysis*, 14, 1201–1219.

Bae, T. and Ko, B. (2017), "A note on weighted infinite sums of dependent regularly varying tailed random variables," *Journal of the Korean Statistical Society*, 46, 321–327.

Barrientos, A. F., Jara, A., and Quintana, F. A. (2012), "On the Support of MacEachern's Dependent Dirichlet Processes and Extensions," *Bayesian Analysis*, 7, 277–310.

Barrios, E., Lijoi, A., Nieto-Barajas, L. E., and Prünster, I. (2013), "Modeling with Normalized Random Measure Mixture Models," *Statistical Science*, 28, 313–334.

Barron, A., Schervish, M. J., and Wasserman, L. (1999), "The consistency of posterior distributions in nonparametric problems," *The Annals of Statistics*, 27, 536–561.

Bassetti, F., Casarin, R., and Leisen, F. (2014), "Beta-product Dependent Pitman–Yor Processes for Bayesian Inference," *Journal of Econometrics*, 180, 49–72.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004), *Statistics of Extremes: Theory and Applications*, Hoboken, NJ: Wiley.

Bertoin, J. (1999), "Subordinators: Examples and Applications," in *Lectures on Probability Theory and Statistics*, New York: Springer.

Bingham, N. H., Goldie, C. M., Teugels, J. L., and Teugels, J. (1989), *Regular Variation*, Cambridge: CUP.

Bladt, M. and Rojas-Nandayapa, L. (2018), "Fitting Phase–type Scale Mixtures to Heavy–tailed Data and Distributions," *Extremes*, 21, 285–313.

Boettcher, W. S., Hahn, S. S., and Shaw, G. L. (1994), "Mathematics and Music: A Search for Insight into Higher Brain Function," *Leonardo Music Journal*, 53–58.

Boldi, M.-O. and Davison, A. C. (2007), "A Mixture Model for Multivariate Extremes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 217–229.

Bruun, J. T. and Tawn, J. A. (1998), "Comparison of approaches for estimating the probability of coastal flo oding," *J Royal Statist. Soc., Ser. C*, 47, 405–423.

Buzsaki, G. (2006), *Rhythms of the Brain*, Oxford: Oxford University Press.

Canale, A., Lijoi, A., Nipoti, B., and Prünster, I. (2017), "On the Pitman–Yor Process with Spike and Slab Base Measure," *Biometrika*, 104, 681–697.

Castillo, I. (2017), "Pólya tree posterior distributions on densities," in *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, Institut Henri Poincaré, vol. 53, pp. 2074–2102.

Castro, D. (2015), "Bivariate Extremes: Modeling, Smoothing, and Regression," Ph.D. thesis, Pontificia Universidad Católica de Chile, Santiago, Chile.

Castro, D. and de Carvalho, M. (2017), "Spectral Density Regression for Bivariate Extremes," *Stochastic Environmental Research and Risk Assessment*, 31, 1603–1613.

Castro, D., de Carvalho, M., and Wadsworth, J. L. (2018), "Time-Varying Extreme Value Dependence with Application to Leading European Stock Markets," *Annals of Applied Statistics*, 12, 283–309.

Chavez-Demoulin, V. and Davison, A. C. (2005), "Generalized Additive Modelling of Sample Extremes," *Journal of the Royal Statistical Society, Ser. C*, 54, 207–222.

Chavez-Demoulin, V., Embrechts, P., and Hofert, M. (2016), "An extreme value approach for modeling operational risk losses depending on covariates," *J. Risk Insur.*, 83, 735–776.

Christensen, R., Hanson, T., and Jara, A. (2008), "Parametric nonparametric ntatistics: An introduction to mixtures of finite Polya trees," *The American Statistician*, 62, 296–306.

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

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

Corradin, R., Canale, A., and Nipoti, B. (2021), "BNPmix: An R Package for Bayesian Nonparametric Modeling via Pitman-Yor Mixtures," *Journal of Statistical Software*, 100, 1–33.

Davis, J. R. and Uryasev, S. (2016), "Analysis of tropical storm damage using buffered probability of exceedance," *Natural Hazards*, 83, 465–483.

Davison, A. C. and Huser, R. (2015), "Statistics of Extremes," *Annual Review of Statistics and its Application*, 2, 203–235.

Davison, A. C. and Smith, R. L. (1990), "Models for Exceedances over High Thresholds," *Journal of the Royal Statistical Society, Ser. B*, 393–442.

de Carvalho, M. (2016a), "Mean, What Do You Mean?" *The American Statistician*, 70, 270–274.

— (2016b), "Statistics of extremes: Challenges and opportunities," in *Extreme Events in Finance: A Handbook of Extreme Value Theory and its Applications*, ed. Longin, F., Hoboken: Wiley.

de Carvalho, M., Barney, B. J., and Page, G. L. (2020), "Affinity-based measures of biomarker performance evaluation," *Statistical Methods in Medical Research*, 20, 837–853.

de Carvalho, M., Pereira, S., Pereira, P., and de Zea Bermudez, P. (2022), "An Extreme Value Bayesian Lasso for the Conditional Left and Right Tails," *Journal of Agricultural, Biological and Environmental Statistics*, 27, 222–239.

de Haan, L. (2015), "Convergence of Heteroscedastic Extremes," *Statistics & Probability Letters*, 101, 38–39.

de Haan, L. and Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, New York: Springer.

Denisov, D. and Zwart, B. (2007), "On a Theorem of Breiman and a Class of Random Difference Equations," *Journal of Applied Probability*, 44, 1031–1046.

Dombry, C., Engelke, S., and Oesting, M. (2017), "Bayesian inference for multivariate extreme value distributions," *Electronic Journal of Statistics*, 11, 4813–4844.

Doocy, S., Daniels, A., Murray, S., and Kirsch, T. D. (2013), "The human impact of floods: a historical review of events 1980-2009 and systematic literature review," *PLoS currents*, 5.

Doss, H. and Sellke, T. (1982), "The Tails of Probabilities Chosen From A Dirichlet Prior," *The Annals of Statistics*, 10, 1302–1305.

Dunn, P. K. and Smyth, G. K. (1996), "Randomized Quantile Residuals," *Journal of Computational and Graphical Statistics*, 5, 236–244.

Eastoe, E. F. and Tawn, J. A. (2009), "Modelling Non-Stationary Extremes with Application to Surface Level Ozone," *Journal of the Royal Statistical Society, Ser. C*, 58, 25–45.

Einmahl, J. H., De Haan, L., and Piterbarg, V. I. (2001), "Nonparametric Estimation of the Spectral Measure of an Extreme Value Distribution," *Annals of Statistics*, 29, 1401–1423.

Einmahl, J. H., Ferreira, A., de Haan, L., Neves, C., and Zhou, C. (2021), "Spatial Dependence and Space-Time Trends in Extreme Events," *Ann. Statist. (to appear)*.

Einmahl, J. H. J., de Haan, L., and Zhou, C. (2016), "Statistics of Heteroscedastic Extremes," *Journal of the Royal Statistical Society: Series B*, 78, 31–51.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, New York: Springer.

Embrechts, P., Lindskog, F., and McNeil, A. (2001), "Modeling dependence with copulas," *Technical Report, Department of Mathematics, F 'e d é ral Institute of Technology Zurich, Zurich*, 14.

Escobar-Bach, M., Goegebeur, Y., and Guillou, A. (2018), "Local Robust Estimation of the Pickands Function," *The Annals of Statistics*, 46, 2806–2843.

Falk, M. (2007), "On the excess distribution of sums of random variables in bivariate EV models," *REVSTAT—Statistical Journal*, 5, 137–161.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230.

— (1974), "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 2, 615–629.

Fermanian, J.-D. and Wegkamp, M. H. (2012), "Time-Dependent Copulas," *Journal of Multivariate Analysis*, 110, 19–29.

Fisher, R. A. and Tippett, L. H. C. (1928), "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," in *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge Univ Press, vol. 24, pp. 180–190.

Frank, M. G. (2009), *Brain Rhythms*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 482–483.

Franssen, S. and van der Vaart, A. (2021), "The Bernstein-von Mises theorem for the Pitman-Yor process of nonnegative type," *arXiv preprint arXiv:2102.06059*.

Fuentes-Garcia, R., Mena, R. H., and Walker, S. G. (2010), "A new Bayesian nonparametric mixture model," *Communications in StatisticsSimulation and Computation®*, 39, 669–682.

Gelman, A. (2014), *Bayesian Data Analysis*, Boca Raton: CRC Press, third edition. ed.

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. (1999), "Consistent semiparametric Bayesian inference about a location parameter," *Journal of statistical planning and inference*, 77, 181–193.

Ghosal, S. and Van der Vaart, A. W. (2015), *Fundamentals of Nonparametric Bayesian Inference*, CUP, Cambridge.

Glen, A. G., Leemis, L. M., and Drew, J. H. (2004), "Computing the Distribution of the Product of Two Continuous Random Variables," *Computational Statistics & Data Analysis*, 44, 451–464.

Gnedenko, B. (1943), "Sur la distribution limite du terme maximum d'une serie aleatoire," *Annals of Mathematics*, 423–453.

Gong, Y. and Huser, R. (2019), "Asymmetric tail dependence modeling, with application to cryptocurrency market data," *arXiv preprint arXiv:1905.05056*.

Gudendorf, G. and Segers, J. (2010), "Extreme-Value Copulas," in *Copula Theory and Its Applications*, Springer, Berlin, Heidelberg, pp. 127–145.

Guillotte, S., Perron, F., and Segers, J. (2011), "Non-parametric Bayesian inference on bivariate extremes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 377–406.

Hanson, T. and Johnson, W. O. (2002), "Modeling regression error with a mixture of Polya trees," *Journal of the American Statistical Association*, 97, 1020–1033.

Hanson, T. E. (2006), "Inference for mixtures of finite Polya tree models," *Journal of the American Statistical Association*, 101, 1548–1565.

Hanson, T. E., de Carvalho, M., and Chen, Y. (2017), "Bernstein polynomial angular densities of multivariate extreme value distributions," *Statistics & Probability Letters*, 128, 60–66.

Harvey, A. and Oryshchenko, V. (2012), "Kernel density estimation for time series data," *International Journal*

of Forecasting, 28, 3–14.

Heffernan, J. E. and Tawn, J. A. (2004), "A Conditional Approach for Multivariate Extreme Values (with Discussion)," *Journal of the Royal Statistical Society, Ser. B*, 66, 497–546.

Ishwaran, H. and James, L. F. (2001), "Gibbs Sampling Methods for Stick-Breaking Priors," *Journal of the American Statistical Association*, 161–173.

Jackman, S. (2017), "Impeachment 101: The history, process and prospect of a Trump impeachment," *The United States Studies Centre at the University of Sydney*.

Jain, S. and Neal, R. M. (2004), "A Split-Merge Markov Chain Monte Carlo Procedure for the Dirichlet Process Mixture Model," *Journal of Computational and Graphical Statistics*, 13, 158–182.

James, L. F., Lijoi, A., and Prünster, I. (2009), "Posterior analysis for normalized random measures with independent increments," *Scandinavian Journal of Statistics*, 36, 76–97.

Jondeau, E. and Rockinger, M. (2006), "The copula-garch model of conditional dependencies: An international stock market application," *Journal of international money and finance*, 25, 827–853.

Jones, M. C. (1993), "Simple boundary correction for kernel density estimation," *Statistics and Computing*, 3, 135–146.

Kalli, M., Griffin, J. E., and Walker, S. G. (2011), "Slice Sampling Mixture Models," *Statistics and Computing*, 21, 93–105.

Kiriliouk, A., Segers, J., and Warchoł, M. (2015), "Nonparametric Estimation of Extremal Dependence," in *Extreme Value Modelling and Risk Analysis: Methods and Applications*, Boca Raton, FL: Chapman and Hall/CRC.

Klimesch, W. (2012), "Alpha-band Oscillations, Attention, and Controlled Access to Stored Information," *Trends in Cognitive Sciences*, 16, 606–617.

Kolmogorov, A. (1930), "Sur la notion de la moyenne," *Atti Accad. Naz. Lincei*, 12, 388391.

Koopman, S. J. and Harvey, A. (2003), "Computing observation weights for signal extraction and filtering," *Journal of Economic Dynamics and Control*, 27, 1317–1333.

Kraft, C. H. (1964), "A class of distribution function processes which have derivatives," *Journal of Applied Probability*, 1, 385–388.

Lavine, M. et al. (1992), "Some aspects of Polya tree distributions for statistical modelling," *Ann. Statist.*, 20, 1222–1235.

— (1994), "More aspects of Polya tree distributions for statistical modelling," *Ann. Statist.*, 22, 1161–1176.

Leadbetter, M. R. (1974), "On extreme values in stationary sequences," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 28, 289–303.

Ledford, A. W. and Tawn, J. A. (1996), "Statistics for near independence in multivariate extreme values," *Biometrika*, 83, 169–187.

Li, C., He, C., and Lin, C. (2018), "Economic impacts of the possible China–US trade war," *Emerging Markets Finance and Trade*, 54, 1557–1577.

Li, C., Lin, L., and Dunson, D. B. (2019), "On Posterior Consistency of Tail Index for Bayesian Kernel Mixture Models," *Bernoulli*, 25, 1999–2028.

Lijoi, A., Mena, R. H., and Prünster, I. (2007), "Controlling the Reinforcement in Bayesian Non-Parametric Mixture Models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 715–740.

Lijoi, A., Prünster, I., and Rigon, T. (2020), "The Pitman–Yor Multinomial Process for Mixture Modelling," *Biometrika*, 107, 891–906.

Lin, C.-H. and Kao, T.-C. (2008), "Multiple structural changes in the tail behavior: evidence from stock index futures returns," *Nonlinear Analysis: Real World Applications*, 9, 1702–1713.

Lin, N., Emanuel, K. A., Smith, J. A., and Vanmarcke, E. (2010), "Risk assessment of hurricane storm surge for New York City," *Journal of Geophysical Research: Atmospheres*, 115.

Liu, T. and Woo, W. T. (2018), "Understanding the US-China trade war," *China Econ. J.*, 11, 319–340.

Longin, F. (ed.) (2017), *Extreme Events in Finance: A Handbook of Extreme Value Theory and Its Applications*, New York: Wiley.

Longin, F. and Solnik, B. (2001), "Extreme Correlation of International Equity Markets," *The journal of finance*, 56, 649–676.

MacEachern, S. N. (1994), "Estimating Normal Means with a Conjugate Style Dirichlet Process Prior," *Communications in Statistics—Simulation and Computation*, 23, 727–741.

Marcon, G., Padoan, S. A., and Antoniano-Villalobos, I. (2016), "Bayesian inference for the extremal dependence," *Electronic Journal of Statistics*, 10, 3310–3337.

McKibbin, W. J. and Fernando, R. (2020), "The global macroeconomic impacts of COVID-19: Seven scenarios," .

Mefleh, A., Biard, R., Dombry, C., and Khraibani, Z. (2020), "Trend detection for heteroscedastic extremes," *Extremes*, 23, 85–115.

Mena, R. H., Ruggiero, M., and Walker, S. G. (2011), "Geometric stick-breaking processes for continuous-time Bayesian nonparametric modeling," *Journal of Statistical Planning and Inference*, 141, 3217–3230.

Mendes, J. M., Bermudez, P. C. d. Z., Pereira, J., Turkman, K. F., and Vasconcelos, M. J. P. (2010), "Spatial Extremes of Wildfire Sizes: Bayesian Hierarchical Models for Extremes," *Environmental and Ecological Statistics*, 17, 1–28.

Mhalla, L., Chavez-Demoulin, V., and Naveau, P. (2017), "Non-linear Models for Extremal Dependence," *Journal of Multivariate Analysis*, 159, 49–66.

Mhalla, L., de Carvalho, M., and Chavez-Demoulin, V. (2019), "Regression Type Models for Extremal Dependence," *Scandinavian Journal of Statistics*, 49, 1141–1167.

Miller, J. W. and Harrison, M. T. (2014), "Inconsistency of Pitman–Yor Process Mixtures for the Number of Components," *The Journal of Machine Learning Research*, 15, 3333–3370.

Müller, P. and Mitra, R. (2013), "Bayesian Nonparametric Inference—Why and How," *Bayesian Analysis*.

Müller, P., Quintana, F. A., Jara, A., and Hanson, T. (2015a), *Bayesian Nonparametric Data Analysis*, New York: Springer.

— (2015b), *Bayesian Nonparametric Data Analysis*, New York: Springer.

Naveau, P., Huser, R., Ribereau, P., and Hannart, A. (2016), "Modeling Jointly Low, Moderate, and Heavy Rainfall Intensities without a Threshold Selection," *Water Resources Research*.

Nelsen, R. B. (2006), *An Introduction to Copulas*, Springer series in statistics, New York: Springer, 2nd ed.

Ombao, H., Lindquist, M., Thompson, W., and Aston, J. (2016), *Handbook of Neuroimaging Data Analysis*, Boca Raton, FL: Chapman and Hall/CRC.

Patton, A. J. (2006), "Modelling Asymmetric Exchange Rate Dependence," *International economic review*, 47, 527–556.

Pickands, J. (1975), "Statistical inference using extreme order statistics," *the Annals of Statistics*, 119–131.

— (1981), "Multivariate Extreme Value Distributions," in *Proc. 43th Sess. Int. Statist. Inst*, pp. 859–878.

Pitman, J. and Yor, M. (1997), "The Two-Parameter Poisson-Dirichlet Distribution Derived from a Stable Subordinator," *The Annals of Probability*, 855–900.

Poon, S.-H., Rockinger, M., and Tawn, J. (2003), "Modelling Extreme-Value Dependence in International Stock Markets," *Statistica Sinica*, 929–953.

— (2004), "Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications," *Review of Financial Studies*, 17, 581–610.

Quintana, F. A. and Müller, P. (2004), "Nonparametric Bayesian Data Analysis," *Statistical Science*, 19, 95–110.

Quintana, F. A., Müller, P., Jara, A., and MacEachern, S. N. (2022), "The Dependent Dirichlet Process and Related Models," *Statistical Science*, 37, 24–41.

Quintos, C., Fan, Z., and Phillips, P. C. (2001), "Structural Change Tests in Tail Behaviour and the Asian Crisis," *The Review of Economic Studies*, 68, 633–663.

R Development Core Team (2016), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing.

Ray, K. and van der Vaart, A. (2021), "On the Bernstein-von Mises theorem for the Dirichlet process," *Electronic Journal of Statistics*, 15, 2224–2246.

Resnick, S. I. (2007), *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, New York: Springer.

Roberts, J. A., Boonstra, T. W., and Breakspear, M. (2015), "The Heavy Tail of the Human Brain," *Current Opinion in Neurobiology*, 31, 164–172.

Rodríguez, A. and Dunson, D. B. (2011), "Nonparametric Bayesian Models through Probit Stick-Breaking Processes," *Bayesian Analysis*, 6, 145–177.

Rodu, J. and Kafadar, K. (2022), "The q–q Boxplot," *Journal of Computational and Graphical Statistics*, 31, 26–39.

Sabourin, A. and Naveau, P. (2014), "Bayesian Dirichlet mixture model for multivariate extremes: a re-parametrization," *Computational Statistics & Data Analysis*, 71, 542–567.

Sarabia Alegría, J. M., Gómez Déniz, E., et al. (2008), "Construction of Multivariate Distributions: A Review of Some Recent Results," *SORT*, 32, 3–36.

Sato, K.-I. (1999), *Lévy Processes and Infinitely Divisible Distributions*, Cambridge: CUP.

Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 4, 639–650.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London, England.

Sklar, M. (1959), "Distribution functions in dimensions and their margins," *Publ. inst. statist. univ. Paris*, 8, 229–231.

Stephenson, A. G. (2002), "`evd`: Extreme value distributions," *R News*, 2, 31–32.

Teh, Y. W. (2006), "A Hierarchical Bayesian Language Model Based on Pitman–Yor Processes," in *Proceedings of the 21st International Conference on Computational Linguistics*, pp. 985–992.

Tressou, J. (2008), "Bayesian Nonparametrics for Heavy Tailed Distribution. Application to Food Risk Assessment," *Bayesian Analysis*, 3, 367–391.

Tsay, R. S. (2002), *Analysis of Financial Time Series*, New York: Wiley.

Van den Goorbergh, R. W., Genest, C., and Werker, B. J. (2005), "Bivariate option pricing using dynamic copula models," *Insurance: Mathematics and Economics*, 37, 101–114.

Walker, S. G. (2007), "Sampling the Dirichlet Mixture Model with Slices," *Communications in Statistics—Simulation and Computation*, 36, 45–54.

Wand, M. P. and Jones, M. C. (1995), *Kernel Smoothing*, London: Chapman & Hall.

Wang, H. and Tsai, C.-L. (2009), "Tail Index Regression," *Journal of the American Statistical Association*, 104, 1233–1240.

Wigglesworth, R. (2020), "How Big Tech got even bigger in the Covid-19 era," *Financial Times*.

Zellner, A. (1996), "Introduction to Bayesian inference in econometrics," .

Zhang, D., Hu, M., and Ji, Q. (2020), "Financial markets under the global pandemic of COVID-19," *Financial Research Letters*, 101528.