technische universität
dortmund

fakultät
statistik

Modelling interexceedance times of temporally clustered extreme events

by

Christina Henrika Mathieu

Dissertation

in partial fulfilment of the requirements for the degree of
Doktor der Naturwissenschaften
presented to the

Department of Statistics, TU Dortmund University

Dortmund, 2025

Referees:
Prof. Dr. R. Fried
Prof. Dr. K. Schorning

Date of oral examination:
06 November 2025

Abstract

This dissertation presents a unified framework for modelling interexceedance times of temporally clustered extreme events. Building on two established models, we introduce a generalised model, which combines underlying stationary event magnitudes with random, heavy-tailed waiting times. When a sufficiently high threshold is used to define exceedances, the resulting interexceedance times can be approximated by a mixture of a Dirac measure at zero and a Mittag-Leffler distribution.

The generalised mixed distribution is characterised by three parameters: the tail parameter, the extremal index, and a scale parameter. To estimate them from data, we propose two minimum distance estimators, CMmod1 and CMmod2, based on modified Cramér-von Mises distances. Their weak consistency is shown theoretically, and their finite sample performance is assessed through extensive simulations. Results highlight CMmod2's robustness to threshold selection and various parameter constellations. The estimators are also evaluated against existing methods in the special cases of the compound Poisson process and the fractional Poisson process. Despite requiring the estimation of an additional parameter, CMmod1 and CMmod2 show competitive accuracy, suggesting that explicit model selection may be avoidable.

A case study on North Atlantic cyclone data illustrates the practical value of the FCPP, revealing spatial differences in clustering behaviour and confirming the real-world applicability of the proposed methodology.

Contents

List of symbols

The following tables list the key notations used throughout this thesis. The list is not intended to be exhaustive and is organised thematically rather than alphabetically.

General notations

symbol	**meaning**
$\xrightarrow{d}$	convergence in distribution
$\overset{d}{=}$	equality in distribution
$\lfloor \cdot \rfloor$	floor function: $\lfloor x \rfloor = \max\{n \in \mathbb{N} \mid n \leqslant x\}$
$\lceil \cdot \rceil$	ceiling function: $\lceil x \rceil = \min\{n \in \mathbb{N} \mid n \geqslant x\}$
$\mathbb{1}\{\cdot\}$	indicator function
$f(x) \sim g(x)$	two functions $f, g : (0, \infty) \to (0, \infty)$ such that $\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1$
$\Gamma(\cdot)$	Gamma function

Model notations

symbol	**meaning**
$(X_n, T_n)_{n\in\mathbb{N}}$	marked point process with
X_n	magnitude: value of the n-th event
T_n	arrival time: time of the n-th event
W_n	waiting time: time between the $n-1$ and n-th event $W_n = T_n - T_{n-1}$.
$u \in \mathbb{R}$	threshold
$p(u)$	excess probability $p(u) = \mathbb{P}(X_1 > u)$
$\eta(u)$	stopping time: position of the first exceedance $\eta(u) = \min\{k \geqslant 1 \mid X_k > u\}$
$T(u)$	time of the first exceedance $T(u) = \sum_{i=1}^{\eta(u)} W_i$ (given $X_0 > u$)
$(X_k(u), T_k(u))_{k\in\mathbb{N}}$	continuous time random exceedances (CTRE) with
$X_k(u)$	exceedance: value of the k-th exceedance
$T_k(u)$	interexceedance time (IET): time between the $k-1$ and k-th exceedances; distributed as $T(u)$
$b(\cdot)$	regularly varying funcion with index $1/\beta$, $\beta \in (0,1]$, $b(n) = n^{1/\beta}L(n)$ with $L(\cdot)$ slowly varying
σ_u	$\sigma_u = b(1/p(u)) = p(u)^{-1/\beta}L(1/p(u))$
l_u	$l_u = p(u)^{1/\beta}\sigma_u$
$\mathrm{Exp}(\lambda)$	exponential distribution with rate $\lambda > 0$
δ_0	Dirac measure at point zero
$\mathrm{ML}(\beta, \varsigma)$	Mittag-Leffler distribution with tail parameter $\beta \in (0,1]$ and scale parameter $\varsigma > 0$
$\mathbb{P}_{(\beta,\theta,\sigma)}$	mixture distribution: $\mathbb{P}_{(\beta,\theta,\sigma)} = (1-\theta)\,\delta_0 + \theta\,\mathrm{ML}(\beta, \theta^{-1/\beta}\sigma)$
$F^{(\beta,\theta,\sigma)}$	cumulative distribution function of $\mathbb{P}_{(\beta,\theta,\sigma)}$, $F^{(\beta,\theta,\sigma)}(t) = (1-\theta) + \theta\, F_*^{(\beta,\theta,\sigma)}(\mathrm{t})$ for $t > 0$
$F_*^{(\beta,\theta,\sigma)}$, $f_*^{(\beta,\theta,\sigma)}$	distribution function and density of $\mathrm{ML}(\beta, \theta^{-1/\beta}\sigma)$
$T_{(\beta,\theta,\sigma)}$	random variable following the mixture distribution $\mathbb{P}_{(\beta,\theta,\sigma)}$

Notations of Section 5.2 and Section 5.4

symbol	**meaning**
$\phi \in \Phi$	parameter vector ϕ and compact set $\Phi \subset \mathbb{R}^m$
μ_ϕ	positive valued continuous distribution with parameter vector ϕ; continuous regarding ϕ
$F_*^{(\phi)}$, $f_*^{(\phi)}$	cumulative distribution function and density of μ_ϕ
$\mathbb{P}_{(\phi,\theta)}$	mixture distribution: $\mathbb{P}_{\phi,\theta} = (1-\theta)\,\delta_0 + \theta\,\mu_\phi$ with $\theta \in [a,1]$, $a > 0$
$F^{(\phi,\theta)}$	cumulative distribution function of $\mathbb{P}_{(\phi,\theta)}$, $F^{(\phi,\theta)}(t) = (1-\theta) + \theta\,F_*^{(\phi)}(t)$ for $t > 0$
$T_{(\phi,\theta)}$	random variable following the mixture distibution $\mathbb{P}_{(\phi,\theta)}$
$T^{(n)}$	random variable on the positive real numbers with $\frac{T^{(n)}}{\sigma_n} \xrightarrow{d} T_{(\phi,\theta)}$, $\sigma_n > 0$
$F^{(n)}$	cumulative distribution function of $T^{(n)}$
$T_{1,n}, \ldots, T_{k_n,n}$	k_n i.i.d. copies of $T^{(n)}$
F_{k_n}	empirical cumulative distribution function of $T_{1,n}, \ldots, T_{k_n,n}$

1 Introduction

Extreme weather events, such as heatwaves, heavy rainfall, or storms, pose significant challenges for risk assessment and decision-making in fields such as climate science, meteorology, finance, and insurance. Accurate modelling of such phenomena is crucial yet inherently difficult, as, by definition, extreme events are rarely observed.

In extreme value theory, two classical approaches are used to model such events: annual maxima are modelled using the generalised extreme value distribution (block maxima approach), while exceedances over a high threshold are modelled using the generalised Pareto distribution (peak-over-threshold approach). These models can be used to estimate key risk measures, such as return periods or return levels. However, these metrics do not capture any temporal structure, such as the clustering of extremes in time. For instance, if extreme events always occurred in immediate pairs every two years, the return period would be approximately one year, which would not reflect the actual recurrence behaviour due to its high variance. Empirical observations show that extreme events often do not occur isolated but tend to appear in temporal clusters. Accounting for such clustering in modelling can be of crucial importance, as extreme events, which occur close together, often cause more severe damage than temporally isolated ones.

This dissertation addresses this phenomenon by focusing on the interarrival times of extremes, where an extreme event is identified by threshold exceedance. The time between two such exceedances is referred to as the interexceedance time. In the literature, two primary models are used to explain clustering in the peak-over-threshold framework. These models are based on fundamentally different mechanisms: Dependency between magnitudes leads to interexceedance times that follow a mixture distribution with a point mass at zero and an exponential component, while heavy-tailed waiting times between independent magnitudes yield interexceedance times that follow a Mittag-Leffler distribution in the limit.

The goal of this work is to introduce a generalised model that captures both mechanisms simultaneously, thus offering greater modelling flexibility while encompassing both established models as special cases. This unified framework may eliminate the need for a prior model choice. To estimate the parameters of the proposed model, novel minimum distance estimators are developed and evaluated. These estimators are based on the Cramér–von Mises distance.

The dissertation is structured as follows: Since the models and methods developed build on concepts from extreme value theory, point processes, and heavy-tailed distributions, the necessary theoretical background is presented in Chapter 2. Chapter 3 is dedicated to models for interexceedance times. This includes a discussion of existing models, where the asymptotic distributions of the IETs are introduced, along with the corresponding point process representations. The new generalised model is then presented and discussed in detail. Chapter 4 focuses on the development of an appropriate estimation method. Existing estimation methods for the submodels are reviewed, and the challenges in fitting the generalised model to data are highlighted before introducing the newly developed estimators. Chapters 5 and 6 evaluate the performance of the two estimators: First, weak consistency under certain conditions is established, and then their behaviour in finite samples is investigated through a simulation study. A further chapter demonstrates the practical relevance of the work by applying it to a dataset of mid-latitude cyclones. With the exception of the background chapter, each chapter concludes with a summary and a critical discussion of the findings. The final chapter summarises the results of the dissertation as a whole and provides an outlook on potential future research directions arising from this work. Additional results, derivations, and figures, excluded from the main text for the sake of clarity and to avoid repetition, can be found in the Appendix.

Parts of this thesis, specifically the development of the generalised model, the estimation method based on a modification of the Cramér–von Mises distance, as well as parts of the simulation study and the data analysis, have been published as a working paper entitled “Models for temporal clustering of extreme events with applications to mid-latitude winter cyclones” on arXiv (Mathieu et al., 2025).

2 Basics

In this chapter, we present the theoretical foundations necessary for understanding the subsequent chapters. We begin with the mathematical concept of *heavy-tailed distributions*. In this context, we discuss *regularly varying functions*, which are a key tool for describing the tail behaviour of heavy-tailed distributions, and we introduce the *class of stable distributions* as well as the *Mittag-Leffler distribution*, both of which exhibit heavy-tailed behaviour. In the second section, we provide a brief introduction to the fundamentals of extreme value theory. We first present the main theorems and concepts in the context of independent sequences and then extend the discussion to stationary time series. The third section gives a concise overview of renewal processes, which form the theoretical basis for the study of interexceedance times (IET), the main subjects of interest in this dissertation. The results presented in this chapter are well-established in the literature. For each section, relevant references are provided to guide the reader to more detailed treatments of the respective topics.

2.1 Heavy-tailed distributions on the positive real numbers

The term *heavy-tailed* is used variably across different applications. Basically, it refers to probability distributions in which a large amount of probability mass lies in one or both tails, particularly when compared to the normal or exponential distribution.

Foss et al. (2013) define a heavy-tailed distribution as one whose exponential moments do not exist, meaning

$$\int_{-\infty}^{\infty} \exp(\lambda x)\, \mathrm{d}F(x) = \infty \quad \text{for all } \lambda > 0. \tag{2.1}$$

This property is always true for distributions whose k-th moment does not exist for some $k \in \mathbb{N}$, e.g., distributions without variance or without expected value. However, there are also distributions that fulfill (2.1) and yet all moments exist. A well-known example would be the Weibull distribution with a shape parameter less than one.

In this thesis, unless stated otherwise, we use the term *heavy-tailed* to refer specifically to distributions without a finite expected value.

In Chapter 3, we will present several convergence theorems concerning the limiting distribution of IETs. Especially in the case of heavy-tailed waiting times, the concepts of *regularly varying functions*, the class of *stable distributions* and the *Mittag-Leffler distribution* play a central role. We therefore provide a concise overview of these topics and highlight the properties most relevant to our context in the following subsections.

For all three concepts, there exists a wide range of introductory textbooks. We recommend Bingham et al. (1987), Resnick (1987) and Seneta (1976) for regular variation, Samorodnitsky and Taqqu (1994) for stable distributions, and Haubold et al. (2011) and Pillai (1990) for the Mittag-Leffler distribution.

2.1.1 Regularly varying functions

For two functions $f, g : (0, \infty) \to (0, \infty)$, we write $f(x) \sim g(x)$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

A function $h : (0, \infty) \to (0, \infty)$ is *regularly varying* at ∞ with index $\alpha \in \mathbb{R}$ if, for all $\lambda > 0$,

$$\lim_{x\to\infty} \frac{h(\lambda x)}{h(x)} = \lambda^\alpha. \tag{2.2}$$

For $\alpha = 0$, the function h is referred to as *slowly varying*.

Every regularly varying function h with index α admits a representation of the form $h(x) = x^\alpha L(x)$, where L is a slowly varying function. Conversely, any function of the form $h(x) = x^\alpha L(x)$, with L slowly varying, is regularly varying with index α.

For indices $\alpha < 0$, it follows that $\lim_{x\to\infty} h(x) = 0$. Therefore, regularly varying functions can be used to describe the tail behaviour of probability distributions. A random variable X with tail function

$$\bar{F}(x) = 1 - F(x) \sim x^{-\alpha} L(x), \quad \text{as } x \to \infty, \tag{2.3}$$

for some $\alpha > 0$ and slowly varying function L is said to follow a *power-law* distribution. In this case, the moments of X satisfy $\mathbb{E}(|X|^p) < \infty$ for $p < \alpha$ and $\mathbb{E}(|X|^p) = \infty$ for $p > \alpha$.

2.1.2 Stable distributions and the generalised central limit theorem

Let $X, X_1, \ldots, X_n$ be a sequence of stochastically independent and identically distributed (i.i.d.) normal random variables with $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$ and

$S_n = X_1 + \cdots + X_n$. Then, for each $n \in \mathbb{N}$,

$$\frac{S_n - (n - \sqrt{n})\mu}{\sqrt{n}} \stackrel{d}{=} X, \tag{2.4}$$

where $\stackrel{d}{=}$ denotes equality in distribution. This shows that the family of normal distributions is closed under convolution (up to affine transformations), and therefore belongs to the class of *stable distributions*.

A non-degenerate random variable X, or its distribution, is said to be *stable* if there exist sequences $(a_n)_{n\in\mathbb{N}}$, with $a_n > 0$, and $(b_n)_{n\in\mathbb{N}}$ such that

$$\frac{S_n - b_n}{a_n} \stackrel{d}{=} X, \tag{2.5}$$

where $X_1, \ldots, X_n$ are i.i.d. copies of X. If $b_n = 0$ for all n, the distribution is called *strictly stable*.

A stable distributed random variable X is fully characterised by its characteristic function, which takes the form

$$\mathbb{E}\left(\exp(-i\,s\,X)\right) = \exp\left(i\mu s + \varsigma^{\alpha}(-|s|^{\alpha} + i\,s\,\omega(s,\alpha,\delta))\right), \tag{2.6}$$

where

$$\omega(s,\alpha,\delta) = \begin{cases} \delta|s|^{\alpha-1}\tan(\frac{\pi\alpha}{2}), & \alpha \neq 1, \\ -\delta\frac{2}{\pi}\ln|s|, & \alpha = 1, \end{cases} \tag{2.7}$$

and the four distribution parameters are defined as follows:

- stability parameter (or index) $\alpha \in (0, 2]$,
- scale parameter $\varsigma > 0$,
- skewness parameter $\delta \in [-1, 1]$,
- location parameter $\mu \in \mathbb{R}$.

Note that there are various parametrisations in the literature. We refer to the parametrisation of Samorodnitsky and Taqqu (1994).

We write $X \sim S_\alpha(\varsigma, \delta, \mu)$, and say that X is α-*stable*. The index α determines the tail behaviour of the distribution and is of particular importance. It holds that

$$\mathbb{E}(|X|^p) < \infty \;\text{ if } p < \alpha \;\text{ and }\; \mathbb{E}(|X|^p) = \infty \;\text{ if } p > \alpha \;\text{ for } \alpha < 2.$$

For $\alpha = 2$, the characteristic function reduces to that of a normal distribution with mean μ and variance 2ς, and the skewness parameter vanishes since $\omega(s, 2, \delta) = 0$ for all s and δ. Therefore, the normal distribution is the only α-stable distribution with finite variance.

When $\alpha < 1$ and $\delta = 1$ (respectively, $\delta = -1$), the distribution is supported on the positive (respectively, negative) real numbers, and we refer to X as being *one-sided stable*. For $\alpha \neq 1$, X is strictly stable if and only if $\mu = 0$. A strictly one-sided stable variable $X \sim S_\alpha(\varsigma, 1, 0)$ has the Laplace transform

$$\mathbb{E}(\exp(-tX)) = \exp\left(-\frac{\varsigma^\alpha}{\cos(\pi\alpha/2)}t^\alpha\right), \quad \text{for } t > 0, \tag{2.8}$$

see Proposition 1.2.12 in Samorodnitsky and Taqqu (1994). Figure 2.1 characterises density and cumulative distribution function (c.d.f.) of the one-sided stable distribution with Laplace transform $\exp(-t^\alpha)$ and different stability parameters.

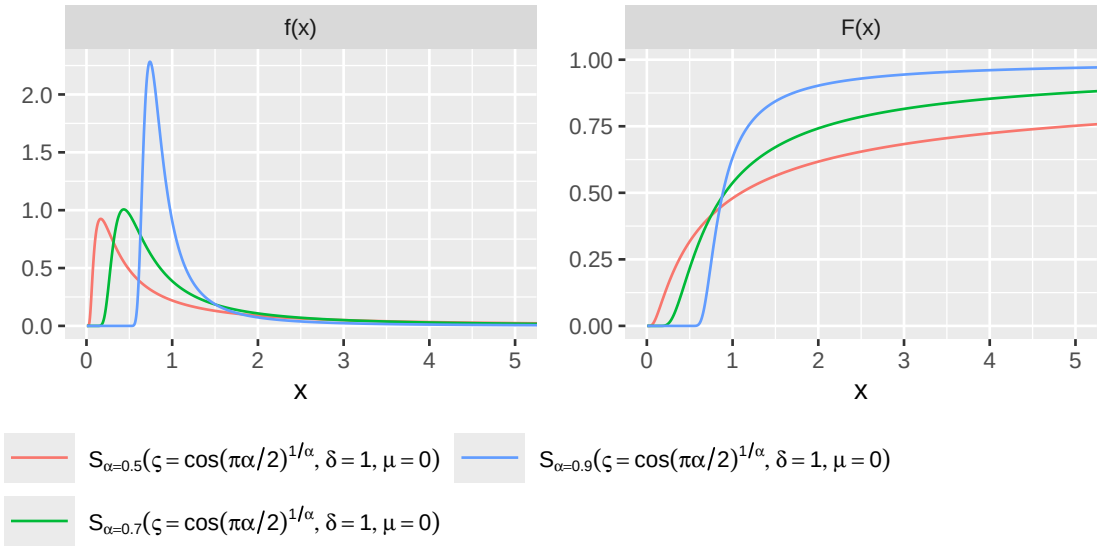

Figure 2.1: Density (left) and cumulative distribution function (right) of a one-sided stable distribution with Laplace transform $\exp(-t^\alpha)$.

Now let $Y, Y_1, Y_2, \ldots$ be an abitrary sequence of independent and identically distributed random variables, and define $S_n = Y_1 + \cdots + Y_n$. Suppose there exist sequences $(a_n)_{n\in\mathbb{N}}$, with $a_n > 0$, and $(b_n)_{n\in\mathbb{N}}$ such that

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} X, \tag{2.9}$$

for some non-degenerate random variable X, where $\xrightarrow{d}$ denotes convergence in distribution. Then X is stable and we say that the distribution of Y belongs to the *domain of attraction* of X. In this sense, the classical Central Limit Theorem is a special case of this generalised convergence result, corresponding to the case $\alpha = 2$.

2.1.3 Mittag-Leffler distribution

At the beginning of the 20th century, the mathematician M. G. Mittag-Leffler introduced the function now known as the *Mittag-Leffler function* (see Mittag-Leffler, 1903a, Mittag-Leffler, 1903b). It is defined by

$$E_\beta(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\beta k + 1)}, \quad \text{for } \beta, z \in \mathbb{C} \text{ and } Re(\beta) > 0, \tag{2.10}$$

where Γ denotes the Gamma function. For $\beta \in (0, 1]$ and real value z, the Mittag-Leffler function is continuous, strictly increasing, satisfying $\lim_{x \to -\infty} E_\beta(x) = 0$ and $E_\beta(0) = 1$. This allows the definition of a probability distribution via the c.d.f.

$$F_\beta(x) = \begin{cases} 1 - E_\beta(-x^\beta), & x > 0, \\ 0, & x \leqslant 0, \end{cases} \tag{2.11}$$

known as the *Mittag-Leffler distribution*, characterised by the tail parameter $\beta \in (0, 1]$.

The Laplace transform of a Mittag-Leffler distributed random variable X is given by

$$\mathbb{E}(\exp(-sX)) = \frac{1}{1 + s^\beta}, \ s > 0. \tag{2.12}$$

For a scale parameter $\varsigma > 0$, we write $\varsigma X \sim \mathrm{ML}(\beta, \varsigma)$. The distribution is continuous regarding its parameters, heavy-tailed for $\beta < 1$ in the sense that $\mathbb{E}(|X|) = \infty$, and in the special case $\beta = 1$, it reduces to the exponential distribution with rate $\lambda = 1/\varsigma$. Figure 2.2 shows the density and c.d.f. for different tail parameters.

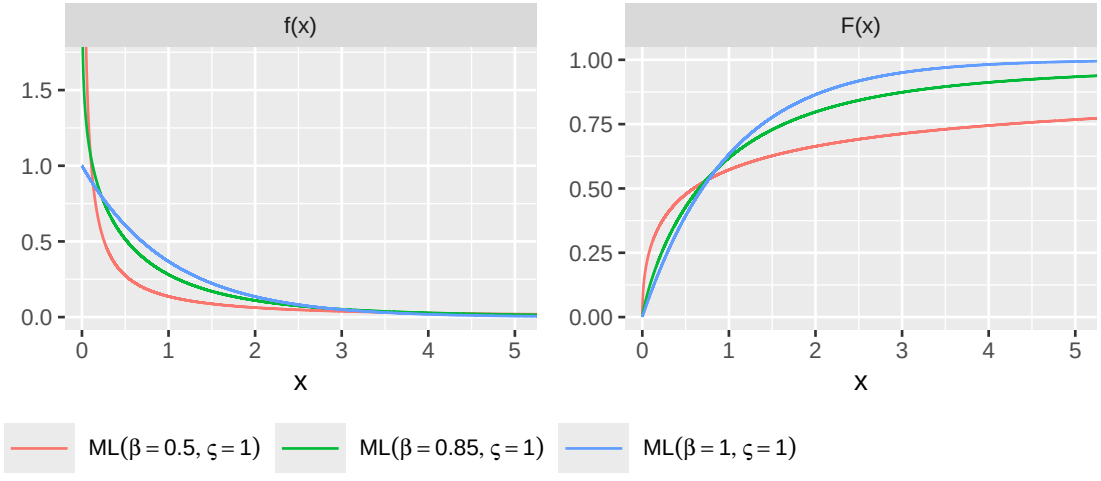

Figure 2.2: Density (left) and cumulative distribution function (right) of the Mittag-Leffler distribution.

Using the Laplace transform, we can show that the Mittag-Leffler distribution lies in the domain of attraction of a strictly stable distribution $S_\beta(\cos(\pi\beta/2)^{1/\beta}, 1, 0)$, whose Laplace transform is given by $\exp(-s^\beta)$:

Let $X_1, \ldots, X_n$ be a sequence of independent and identically $\mathrm{ML}(\beta, \varsigma)$-distributed random variables, and let $D_\beta \sim S_\beta(\cos(\pi\beta/2)^{1/\beta}, 1, 0)$. We consider the Laplace transform of the normalised sum $\frac{\sum_{i=1}^n X_i}{\varsigma n^{1/\beta}}$:

$$\begin{aligned}
\mathbb{E}\left(\exp\left(-s\frac{\sum_{i=1}^n X_i}{\varsigma n^{1/\beta}}\right)\right) &= \mathbb{E}\left(\prod_{i=1}^n \exp\left(-sn^{-1/\beta}\frac{X_i}{\varsigma}\right)\right) \\
&\overset{\text{i.i.d.}}{=} \mathbb{E}\left(\exp\left(-sn^{-1/\beta}\frac{X_1}{\varsigma}\right)\right)^n \\
&= \left(\frac{1}{1+(sn^{-1/\beta})^\beta}\right)^n \\
&= \left(1+\frac{s^\beta}{n}\right)^{-n} \xrightarrow{n\to\infty} \exp\left(-s^\beta\right).
\end{aligned} \tag{2.13}$$

This implies that the normalised sum converges in distribution,

$$\sum_{i=1}^n \frac{X_i}{(\varsigma n^{1/\beta})} \xrightarrow{d} D_\beta, \tag{2.14}$$

and the claim is thus proven (Pillai, 1990, Theorem 4.1).

Haubold et al. (2011, Theorem 19.1.1) state that for two stochastically independent random variables E, exponential distributed with rate one, and D_β, β-stable with Laplace transform as in (2.12), it holds

$$E^{1/\beta} D_\beta \overset{d}{=} \mathrm{ML}(\beta, 1). \tag{2.15}$$

As the Mittag-Leffler distribution, unlike the exponential distribution, is not memoryless, the conditional probability of a value x depends on the value x_0 that has already been exceeded, more precisely:

$$p(x \mid x_0) = f(x + x_0 \mid X > x_0) = \frac{f(x+x_0)}{\mathbb{P}(X > x_0)} \neq f(x), \tag{2.16}$$

where f is the density of the Mittag-Leffler distribution. Figure 2.3 illustrates the conditional distribution for three values of x_0. The higher the value of x_0 is, the greater is the probability of observing larger values of x.

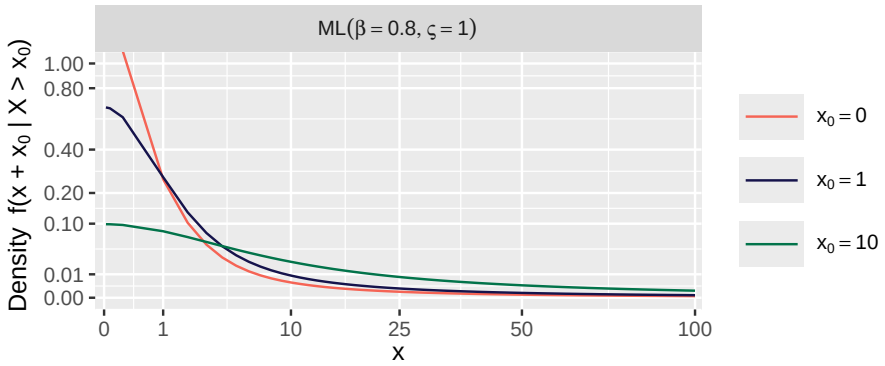

Figure 2.3: Conditional distribution that X exceeds $x + x_0$ given that it already exceeded x_0, for X Mittag-Leffler distributed with tail parameter $\beta = 0.9$ and scale parameter $\varsigma = 1$.

2.2 Statistical modelling of extreme events

Extreme value theory (EVT) deals with the analysis of the behaviour and frequency of rare, extreme events, i.e. events that occur with very low probability. In the case of a unimodal distribution, we are particularly interested in describing the behaviour of the distribution's tails, where such rare events lie.

The following subsections introduce the two fundamental results of EVT, the *Fisher-Tippett-Gnedenko* and *Pickands-Balkema-de Haan theorems*, which form the mathematical basis for the most widely used methods to identify extreme events in univariate samples. These theorems assume that the sample consists of i.i.d. observations. In the final subsection, we relax this assumption by considering strictly stationary sequences. We then extend the discussed approaches by introducing the *extremal index* $\theta \in (0, 1]$, which plays a central role in modelling the temporal clustering of interexceedance times.

This section only gives a concise overview of the topic. For more details, we refer to Beirlant et al. (2004), Coles (2001), de Haan and Ferreira (2006) or Embrechts et al. (1997).

2.2.1 Fisher-Tippett-Gnedenko theorem and block maximum

Let $X, X_1, \ldots, X_n$ be i.i.d. random variables with c.d.f. F. We begin by considering the maximum $M_n = \max\{X_1, \ldots, X_n\}$. It is well known that

$$\mathbb{P}(M_n \leqslant x) = \mathbb{P}(X_1 \leqslant x, \ldots, X_n \leqslant x) \overset{i.i.d.}{=} \mathbb{P}(X_1 \leqslant x)^n = F^n(x).$$

Thus, in principle, we could use F to compute the probability that none of n consecutive events exceed a high value x, or conversely, the probability that at

least one of them does. However, since the true distribution function F is typically unknown, we must estimate the distribution of M_n directly. The empirical distribution function is often not suitable in this context, because only very few observations may exceed x, making tail estimation highly unreliable. Instead, we analyse the limiting behaviour of M_n. For fixed x, since $F(x) \in [0,1]$, $F^n(x)$ converges either to zero or one as $n \to \infty$. To obtain a non-degenerate limit, we normalise M_n appropriately. According to the theorem of *Fisher-Tippett-Gnedenko* (Fisher and Tippett, 1928; Gnedenko, 1943), if there are real valued sequences $(a_n)_{n\in\mathbb{N}}$, with $a_n > 0$, and $(b_n)_{n\in\mathbb{N}}$ such that

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G, \tag{2.17}$$

with G being a non-degenerate distribution, then G belongs to the family of generalised extreme value distributions (GEV). In this case, we say that X, or equivalently its distribution F, *belongs to the max-domain of attraction* of G (Leadbetter et al., 1983, Theorem 1.4.2). The GEV family is defined by its c.d.f.

$$G_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-1/\xi}\right), & \text{für} \quad 1+\xi x > 0, \quad \xi \neq 0, \\ \exp\left(-\exp(-x)\right), & \text{für} \quad x \in \mathbb{R}, \qquad\quad \xi = 0, \end{cases} \tag{2.18}$$

where $\xi \in \mathbb{R}$ is the shape parameter. By incorporating a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\sigma > 0$, we obtain the generalised form

$$G_{\xi,\mu,\sigma}(x) = G_\xi\left(\frac{x-\mu}{\sigma}\right), \tag{2.19}$$

with the domain adjusted accordingly. The shape parameter $\xi \in \mathbb{R}$ determines the distribution tails and divides the GEV into three subfamilies:

- Weibull distribution for $\xi < 0$,
- Gumbel distribution for $\xi = 0$,
- Fréchet distribution for $\xi > 0$.

While the Gumbel distribution has unbounded support in both directions, the Weibull distribution has a bounded upper endpoint, and the Fréchet distribution has a heavy upper tail and finite support from below. Figure 2.4 shows the density and distribution function for illustrations.

As a consequence of the theorem, for sufficiently large $n \in \mathbb{N}$,

$$\mathbb{P}(M_n \leqslant x) \approx G_{\xi,b_n,a_n}(x). \tag{2.20}$$

To apply the GEV model to data, we must estimate the three parameters ξ, μ, and σ, e.g., using the maximum likelihood method. This requires a sample of i.i.d. maxima.

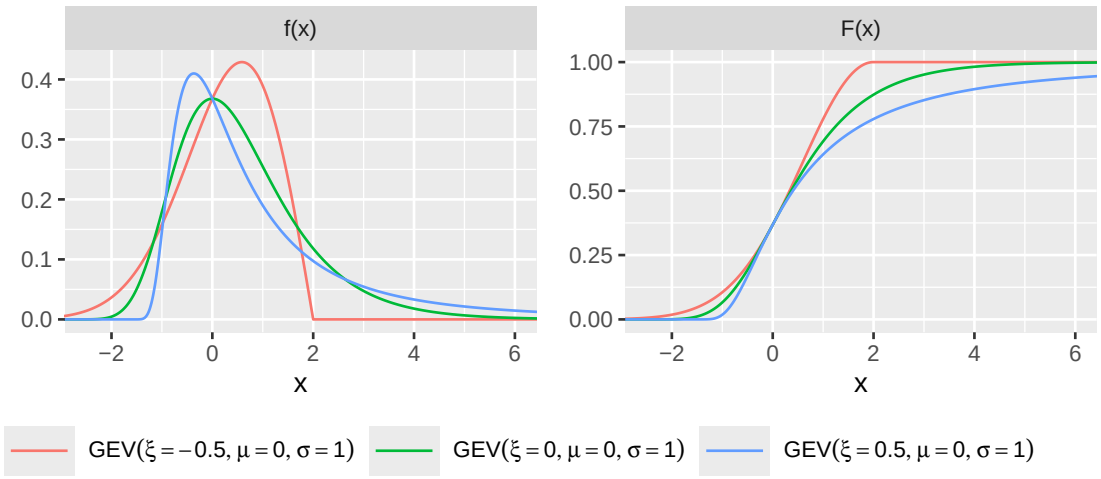

Figure 2.4: Density (left) and cumulative distribution function (right) of the GEV distribution for the three subclasses: Weibull ($\xi < 0$), Gumbel ($\xi = 0$) and Fréchet ($\xi > 0$).

Block maximum approach

The block maximum method partitions the original sample $X_1, \ldots, X_n$ in k non-overlapping blocks of size r, such that $n \geqslant k \cdot r$. Assuming independency of the X_i, the maximum within each block, denoted by $M(J_1), \ldots, M(J_k)$, follows the same distribution as $M_r = \max\{X_1, \ldots, X_r\}$. If we assume $n = k \cdot r$ for simplicity, then

$$\mathbb{P}(M_n \leqslant x) = \mathbb{P}(M(J_1) \leqslant x, \ldots, M(J_k) \leqslant x) = \mathbb{P}(M_r \leqslant x)^k = \mathbb{P}(M_r \leqslant x)^{\frac{n}{r}}.$$

Thus, once we fit the GEV distribution to the sample of block maxima, we can approximate $\mathbb{P}(M_n \leqslant x)$ by $G_{\hat{\xi},\hat{\mu},\hat{\sigma}}(x)^{n/r}$ for all $n \geqslant r$. The choice of block size r and the resulting block number k is critical. On one hand, r must be large enough for the asymptotic approximation in the Fisher-Tippett-Gnedenko theorem to hold. On the other hand, k must be large enough to ensure a sufficient number of maxima for reliable parameter estimation. Figure 2.6 (left) shows a scheme of the block maxima approach.

2.2.2 Pickands-Balkema-de Haan theorem and peak-over-threshold

An alternative approach for modelling extreme events is the peak-over-threshold (POT) method, which focuses on the behaviour of values exceeding a certain threshold u. Unlike the block maxima approach, which considers the maximum

within fixed blocks of data, the POT method directly models the exceedances of the threshold, potentially using all available data.

Let $X, X_1, \ldots, X_n$ be a sequence of i.i.d. random variables with c.d.f. F. Given a threshold u, we define the excess as the random variables

$$Y_i = \begin{cases} X_i - u, & \text{if } X_i > u, \\ 0, & \text{if } X_i \leqslant u. \end{cases} \tag{2.21}$$

The excess distribution function of X is then

$$F_u(x) = \mathbb{P}(X - u \leqslant x \mid X > u) = \mathbb{P}(X \leqslant x + u \mid X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}.$$

The main idea behind the POT method is that, for sufficiently high thresholds, the distribution of excesses, after proper rescaling, can be approximated by a generalised Pareto distribution (GPD) with c.d.f.:

$$H_{\xi,\sigma}(x) = \begin{cases} 1 - \left(1 + \xi\left(\frac{x}{\sigma}\right)\right)^{-1/\xi}, & \text{for } \xi \neq 0, \\ 1 - \exp\left(-\frac{x}{\sigma}\right), & \text{for } \xi = 0, \end{cases} \tag{2.22}$$

with $x > 0$ for $\xi \geqslant 0$ and $0 < x \leqslant -\sigma/\xi$ for $\xi < 0$, where $\xi \in \mathbb{R}$ is the shape parameter and $\sigma > 0$ is the scale parameter. Figure 2.5 illustrates the density and c.d.f. for different shape parameters.

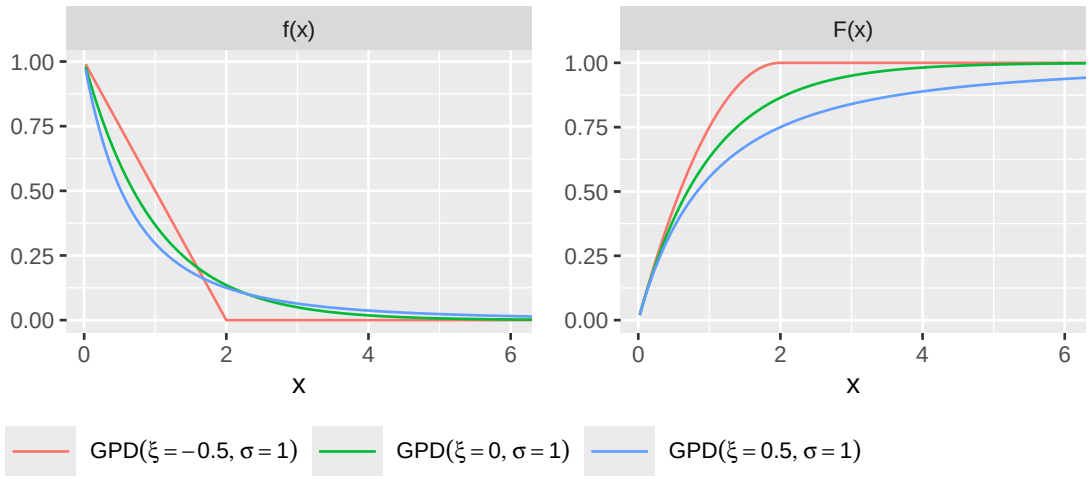

Figure 2.5: Density (left) and cumulative distribution function (right) of the GPD.

According to the Pickands-Balkema-de Haan theorem (Balkema and De Haan, 1974; Pickands, 1975), if X is in the max-domain of attraction of a GEV $G_{\xi,\mu,\sigma}$, then for a sufficiently high threshold u, the distribution of the excesses (scaled appropriately)

converges in distribution to a GPD:

$$F_u(x) \to H_{\xi,\sigma_u}(x), \quad \text{as} \quad u \to x_R, \tag{2.23}$$

where $x_R = \sup\{x \mid F(x) < 1\}$ is the right endpoint of F. In other words, as the threshold u increases, the distribution of exceedances above u converges to the GPD with parameters ξ and σ_u. The GPD is closely related to the GEV distribution and shares the same shape parameter ξ and it holds $\sigma_u = \sigma + \xi(u - \mu)$.

Thus, we can approximate the excess distribution for a sufficiently large threshold u by the GPD:

$$H_{\xi,\sigma_u}(x) \approx F_u(x) = \frac{F(x+u) - F(u)}{1 - F(u)}, \quad x > 0 \tag{2.24}$$

and therefore we can approximate the distribution tail of X by

$$F(x+u) \approx (1 - F(u)) \cdot H_{\xi,\sigma_u}(x) + F(u), \quad x > 0. \tag{2.25}$$

$F(u)$ may be approximated by the empirical distribution function and the parameters ξ and σ_u can be estimated using the exceedances of a sample, typically by maximum likelihood estimation.

Peak-over-threshold approach

The POT method uses the events that exceed a threshold u of the original i.i.d. sample $X_1, \ldots, X_n$ for a sample of excesses. The GPD can then be fitted to that sample. The number of exceedances $K = \sum_{i=1}^{n} \mathbb{1}\{X_i > u\}$ is a random number and follows a binomial distribution with parameters n and $p(u) = \mathbb{P}(X > u) = 1 - F(u)$. Thus, the expected number of exceedances is

$$\nu_n(u) = n\, p(u) \tag{2.26}$$

and it is monotone decreasing as $u \to x_R$. If there is a sequence of thresholds $(u_n)_{n\in\mathbb{N}}$ such that

$$\nu_n(u_n) \to \nu \in (0, \infty), \tag{2.27}$$

K is asymptotically Poisson distributed with parameter ν and it holds

$$\mathbb{P}(M_n \leqslant u_n) = F(u_n)^n = \left(1 - \frac{n\, p(u)}{n}\right)^n \to \exp(-\nu), \quad \text{as } n \to \infty. \tag{2.28}$$

The choice of the threshold u is crucial for estimation. On one hand, u should be

high enough for the asymptotic approximation in the Pickands-Balkema-de Haan theorem to hold and on the other hand, u should not be too high, so that there are enough exceedances to analyse. In practice, instead of choosing a threshold u, the sample size k is specified to ensure enough exceedances in the sample. Figure 2.6 (right) shows a scheme of the POT approach.

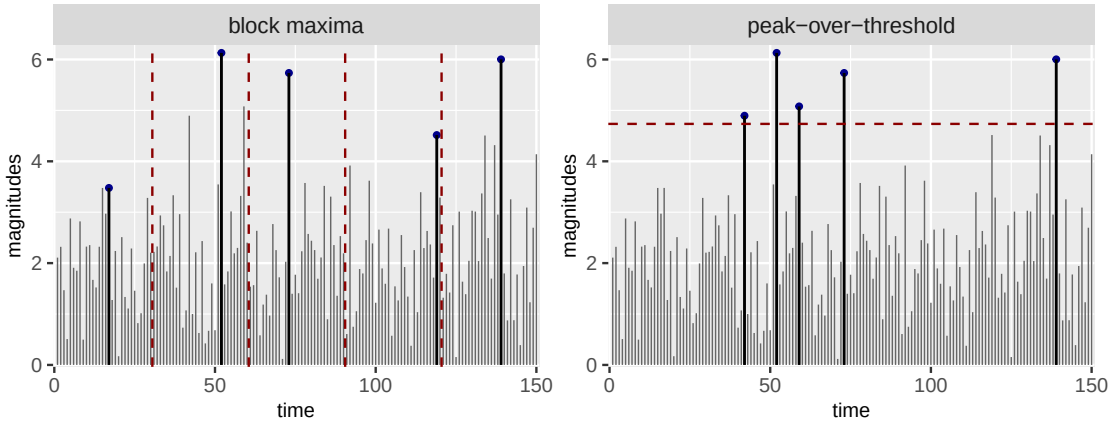

Figure 2.6: Comparison of the block maxima and POT approach on simulated data using the Gumbel distribution.

2.2.3 Extension for stationary data

In the previous two subsections, we assumed i.i.d. random variables $X, X_1, \ldots, X_n$. But since in many applications the events are somehow dependent, we relax the requirement to a strictly stationary sequence $(X_n)_{n \in N}$, i.e.,

$$\forall h \in \mathbb{N},\ \forall k \in \mathbb{N}: \quad (X_{h+1}, \ldots, X_{h+k}) \stackrel{d}{=} (X_1, \ldots, X_k). \tag{2.29}$$

In addition we consider a mixing condition that limits the long-range dependence. Therefore we define $M(I) = \max\{X_i \mid i \in I \subset \mathbb{N}\}$.

Condition 1. *Let be* $\mathcal{I}_{j,k}(u_n) = \{\{M(I) \leqslant u_n\} \mid I \subseteq \{j, \ldots, k\}\}$. *For all* $A_1 \in \mathcal{I}_{1,l}(u_n), A_2 \in \mathcal{I}_{l+s,n}(u_n)$ *and* $1 \leqslant l \leqslant n - s$,

$$|P(A_1 \cap A_2) - P(A_1)P(A_2)| \leqslant \alpha(n, s)$$

and $\alpha(n, s_n) \to 0$ *as* $n \to \infty$ *for some positive integer sequence* s_n *such that* $s_n = o(n)$, *i.e.,* $\frac{s_n}{n} \to 0$. *This mixing condition is called* $D(u_n)$ *condition.*

The $D(u_n)$ condition says two events $\{M(I_1) \leqslant u_n\}$ and $\{M(I_2) \leqslant u_n\}$ are asymptotically independent, when the index sets are separated by a relatively short distance s_n.

We can now generalise the Fisher-Tippett-Gnedenko theorem for strictly stationary sequences $(X_n)_{n\in\mathbb{N}}$.

According to Leadbetter et al. (1983), if there are sequences $(a_n)_{n\in\mathbb{N}}$, with $a_n > 0$, and $(b_n)_{n\in\mathbb{N}}$ such that X is in the max-domain of attraction of a GEV with c.d.f. $G_{\xi,\mu,\sigma}$, if the $D(u_n)$ condition holds with $u_n = a_n x + b_n$ for each x such that $G_{\xi,\mu,\sigma}(x) > 0$, and if $\mathbb{P}((M_n - a_n)/b_n \leqslant x)$ converges, then

$$\mathbb{P}\left(\frac{M_n - a_n}{b_n} \leqslant x\right) \xrightarrow{d} G^{\theta}_{\xi,\mu,\sigma}(x), \tag{2.30}$$

as $n \to \infty$, for some constant $\theta \in [0,1]$. θ is called the *extremal index*. For $\theta > 0$, $G^{\theta}_{\xi,\mu,\sigma}$ is a GEV with parameters $\tilde{\xi} = \xi$, $\tilde{\mu} = \mu - \sigma\frac{1-\theta^{\xi}}{\xi}$ and $\tilde{\sigma} = \sigma\theta^{\xi}$, or for $\xi = 0$, $\tilde{\mu} = \mu + \sigma\log(\theta)$ and $\tilde{\sigma} = \sigma$, respectively. For $\theta = 0$, $G^{\theta}_{\xi,\mu,\sigma}$ is a degenerative distribution and $\theta = 1$ holds for stochastically independent sequences $(X_n)_{n\in\mathbb{N}}$.

In the following, we always assume $\theta \in (0,1]$. Then, for every value $\nu \in (0,\infty)$, there is a sequence of thresholds $(u_n)_{n\in\mathbb{N}}$ such that

$$\begin{aligned} np(u_n) &\to \nu \qquad \text{and} \\ \mathbb{P}(M_n \leqslant u_n) &\to \exp\left(-\nu\theta\right). \end{aligned} \tag{2.31}$$

Max-autoregressive process

One example for a stationary process with extremal index $\theta \in (0,1]$ is the *max-autoregressive process* (MAR) (see, e.g., Ferro and Segers, 2003):

Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. Fréchet random variables with c.d.f.

$$\phi_1(x) = G_{1,1,1}(x) = \exp(-x^{-1}), \quad x > 0, \tag{2.32}$$

with $G_{\xi,\mu,\sigma}$ the c.d.f. of the GEV distribution introduced in (2.18) and (2.19).

We define the max-autoregressive process $(X_n)_{n\in\mathbb{N}}$ inductively by

$$\begin{aligned} X_1 &= Y_1 \\ X_n &= \max\{(1-\theta)\,X_{n-1}, \theta\,Y_n\}. \end{aligned} \tag{2.33}$$

It holds

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - n}{n} \leqslant x\right) = G^{\theta}_1(x) = G_{1,-(1-\theta),\theta}(x). \tag{2.34}$$

2.3 Renewal processes for modelling interexceedance times

In the previous Section 2.2, we introduced the foundation of extreme value theory which deals with the precise description of distribution tails. In this work, however, we focus on the recurrence time of extreme events, i.e. the times that elapse between two successive extreme events. Since we identify extremes with the POT method, we call this time *interexceedance time* (IET).

We introduce IETs in the next chapter in depth using the framework of *renewal processes*. To this end, we provide a concise overview tailored to the context of our study and introduce four specific types of renewal processes that we utilise. Further details can be found in numerous introductory textbooks on the subject, e.g., Borovkov (1987) and Reiss (2012). For the fractional Poisson and the fractional compound Poisson process we recommend, e.g., Laskin (2003).

2.3.1 Renewal processes

Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of positive i.i.d. random variables. We can interpret those as waiting times between two successive events. Then, the sum of the first n waiting times

$$T_n = \sum_{i=1}^{n} W_i \tag{2.35}$$

denotes the occurrence time of the n-th event and the process $(N_t)_{t>0}$ with

$$N_t = \sum_{n=1}^{\infty} \mathbb{1}\{T_n \leqslant t\}, \tag{2.36}$$

which counts the number of events until time $t > 0$, is called *renewal process*. Since $\{N_t \geqslant n\} = \{T_n \leqslant t\}$, also the waiting times $(W_n)_{n\in\mathbb{N}}$ identify the renewal process uniquely. If the increments $N_{t+s} - N_t$, $t, s > 0$, are stationary, meaning that their distribution depends only on the length s and not on the time point t, the process is called *homogeneous*, otherwise, it is referred to as *inhomogeneous*. However, since this work focuses solely on homogeneous renewal processes, we will disregard the inhomogeneous case for the remainder of this dissertation. Thus, any mention of a renewal process henceforth will implicitly refer to a homogeneous renewal process. The most famous renewal process is the *Poisson process*.

Poisson process

A renewal process $(N_t)_{t>0}$ is called *Poisson process* (PP) with intensity $\lambda > 0$ if and only if the waiting times W_n are exponential distributed with rate λ for all $n \geqslant 1$. For the counting process $(N_t)_{t>0}$ it follows that

- N_t is Poisson distributed with rate λt for all $t > 0$,
- the increments $N_{t+s} - N_t$ are Poisson distributed with rate λs for all $s, t > 0$,
- for any time points $0 < t_1 < t_2 < \cdots < t_n < \infty$ the increments $N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}$ are stochastically independent.

This means that for any time interval of length $s > 0$ the number of events within this interval is identically distributed regardless of the considered time point and the number of past events has no stochastical influence on the number of future events.

Fractional Poisson process

A *fractional Poisson process* (FPP) is a renewal process with Mittag-Leffler distributed waiting times $(W_n)_{n\in\mathbb{N}}$ with tail parameter $\beta \in (0, 1]$ and scale parameter $\varsigma > 0$, i.e., $W_1 \sim ML(\beta, \varsigma)$. For $\beta = 1$ we have exponential distributed waiting times with rate $\lambda = 1/\varsigma$ and thus a standard Poisson process. The distribution of the counting process $(N_t)_{t>0}$ of a FPP is given by

$$\mathbb{P}(N_t = k) = \frac{\left(\frac{t}{\varsigma}\right)^{\beta k}}{k!} \sum_{j=0}^{\infty} \frac{(j+k)!}{j!} \frac{\left(-\left(\frac{t}{\varsigma}\right)^{\beta}\right)^{j}}{\Gamma\left(\beta(j+k)+1\right)}, \quad \text{for all } k \in \mathbb{N}_0. \tag{2.37}$$

The mean and variance of the counting process N_t are:

$$\begin{aligned} \mathbb{E}(N_t) &= \frac{\left(\frac{t}{\varsigma}\right)^{\beta}}{\Gamma(1+\beta)}, \\ \mathrm{Var}(N_t) &= \left(\frac{\left(\frac{t}{\varsigma}\right)^{\beta}}{\Gamma(1+\beta)}\right)^2 \left(\frac{\beta\Gamma(\beta)^2}{\Gamma(2\beta)} - 1\right) + \frac{\left(\frac{t}{\varsigma}\right)^{\beta}}{\Gamma(1+\beta)}. \end{aligned} \tag{2.38}$$

Since the Mittag-Leffler distribution is not memoryless, see (2.16), disjoint increments of the FPP are no longer stochastically independent.

2.3.2 Compound renewal processes

By adding so-called *marks* Y_n to each occurrence time T_n of a classical renewal process, we receive a *marked renewal process*. In case of i.i.d. marks, we also call it a *compound renewal process*. More precisely, let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables with distribution π. Then the process $(Z_t)_{t>0}$ with

$$Z_t = \sum_{n=1}^{\infty} Y_n \cdot \mathbb{1}\{T_n \leqslant t\} = \sum_{n=1}^{N_t} Y_n \tag{2.39}$$

is called *compound renewal process*, i.e. Z_t sums until time t occurred marks. If the processes $(Y_n)_{n\in\mathbb{N}}$ and $(W_n)_{n\in N}$ are also stochastically independent, the compound renewal process is called *uncoupled*.

In the uncoupled case, the mean and variance can be calculated using Wald's theorem and the law of total variance:

$$\begin{aligned} \mathbb{E}(Z_t) &= \mathbb{E}(N_t)\,\mathbb{E}(Y_1) \\ \operatorname{Var}(Z_t) &= \mathbb{E}(N_t)\operatorname{Var}(Y_1) + \operatorname{Var}(N_t)\,\mathbb{E}(Y_1)^2 \end{aligned} \tag{2.40}$$

In this work, we will use compound renewal processes to describe temporal clustered extreme events. Then W_n describes the time between the $n-1$ and n-th cluster and Y_n is the number of extremes within the n-th cluster. In this case Z_t counts the number of extremes until time t. The process $(Y_n, W_n)_{n\in\mathbb{N}}$ is the governing sequence of the compound renewal process.

Two examples of compound renewal processes are the *compound Poisson process* (CPP) and the *fractional compound Poisson process* (FCPP), where the underlying waiting times $(W_n)_{n\in\mathbb{N}}$ correspond to a PP or FPP, respectively. If $(Z_t)_{t>0}$ is a CPP, we can write shortly $Z_t \sim CPP(\lambda, \pi)$, where $\lambda > 0$ is the intensity of the corresponding PP. Analogously, we write $Z_t \sim FCPP((\beta, \varsigma), \pi)$ if $(Z_t)_{t>0}$ is a FCPP.

2.4 Software and computations

All implementations, simulations, and most figures in this dissertation were produced using the statistical software R (R Core Team, 2024). To model and analyse heavy-tailed phenomena, several specialised R packages were employed. We used the package `stabledist` (Wuertz et al., 2016) for working with stable distributions. Functions for the density, cumulative distribution function, quantiles and random number generation of the Mittag-Leffler distribution as well as several estimation procedures are provided in `MittagLeffleR` (Gill and Straka, 2017). Additionally, to work with regularly varying distributions such as the Pareto distribution, we utilised

the package `ReIns` (Reynkens and Verbelen, 2020), a toolbox tailored to extreme value theory and the analysis of heavy tails. The package `exdex` (Northrop and Christodoulides, 2023) provides several estimation methods for the extremal index.

For graphical visualisations, we relied on `ggplot2` (Wickham, 2016).

Source code written specifically for the analyses in this dissertation, such as the estimation procedures introduced in Chapter 4, is implemented in the package `FCPP` (Mathieu, 2025), which is publicly available on the open-source platform GitHub.

3 Asymptotic models for interexceedance times

In Section 2.2, we introduced the foundation of extreme value theory, which deals with the precise description of distribution tails. However, in this work, we focus on the return time of extreme events. Our goal is thereby to describe temporally clustered extremes.

This chapter is structured as follows: After first introducing the framework conditions of our models, we describe the concepts of interexceedance times (IETs) and temporal clustering in the first section. Thereby we introduce the dispersion statistic, which is an indicator for clustering behaviour. In the second section, we introduce three asymptotic models for the IETs. While the first of them corresponds to a Poisson process and cannot describe clustering, the other two are based on different mechanisms that lead to a compound Poisson process and a fractional compound process, respectively. Both can describe temporal clustering. For each model we describe the clustering behaviour and calculate the dispersion statistic. The main section is the third section. Here we present our contribution to the topic: We provide a generalised model that captures the previous three models and is more flexible. We show that this model corresponds to a fractional compound Poisson process. We conclude the chapter with a summary and discussion.

3.1 Introduction of interexceedance times and temporal clustering

From now on, we identify extreme events with the POT method, i.e. extremes are events that exceed a chosen threshold. Therefore, we also call them *exceedances*. The waiting time between two successive exceedances is then called *interexceedance time*. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of identically distributed random variables belonging to the max-domain of attraction of a GEV. We call them *magnitudes* or *event magnitudes* since they indicate certain values of occurring events. In addition, we introduce a temporal component by joining each magnitude X_n with a (random) occurrence time $T_n > 0$, such that $T_{n-1} < T_n$ for all n. The time between the $n-1$ and n-th event is then the waiting time $W_n = T_n - T_{n-1}$. We require the sequence $(W_n)_{n\in\mathbb{N}}$ to be stochastically independent and also independent of $(X_n)_{n\in\mathbb{N}}$. Thus,

$(X_n, W_n)_{n\in\mathbb{N}}$ is an uncoupled marked renewal process, with possibly stochastically dependent marks $(X_n)_{n\in\mathbb{N}}$.

3.1.1 Interexceedance times

Without loss of generality, we set $X_0 > u$ and $T_0 = 0$ for a chosen threshold u, otherwise we skip the first elements of the sequence until the first excess.

Then the interexceedance time is defined as

$$T(u) = \sum_{i=1}^{\eta(u)} W_i, \tag{3.1}$$

with stopping time $\eta(u) = \min\{k \geqslant 1 \mid X_k > u\}$, see Figure 3.1 for illustration. By using the POT method and restarting the process after every excess, we receive the process $(X_k(u), T_k(u))_{k\in\mathbb{N}}$, called *continuous time random exceedances* (CTRE), where $X_k(u)$ is the k-th exceedance and $T_k(u)$ is the IET between the $k-1$ and k-th exceedances. For all $k \in \mathbb{N}$, $T_k(u) \stackrel{d}{=} T(u)$ applies.

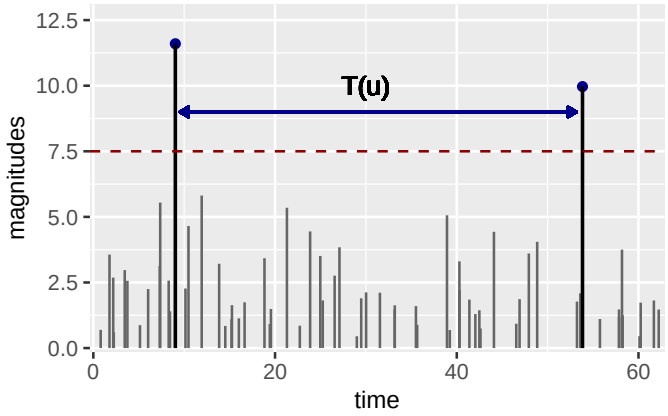

Figure 3.1: Illustration of the interexceedance time between two extreme events.

3.1.2 Temporal clustering

Temporally clustered extreme events refer to the occurrence of multiple extremes within a relatively short time interval. Those extremes form one *cluster*, see Figure 3.2 for illustration. In mathematical sense, we talk about clustering behaviour if it cannot be explained by the random behaviour of a PP. It is well known and we will roughly explain it again in the next section that in the simplest case, when (X_n) is a sequence of i.i.d. random variables separated equidistantly, i.e. $W_n \equiv 1$ for all

$n \in \mathbb{N}$, the IETs asymptotically form a PP. That means that the IETs are exponential distributed with rate $\lambda > 0$ and, for a fixed time interval s, $N_s(u)$, the number of exceedances within this interval, is Poisson distributed with rate $s\lambda$.

A deviation from that behaviour indicates that the number of exceedances within time intervals of that length is either overdispersive (if $\mathrm{Var}(N_s(u)) > \mathbb{E}(N_s(u))$) or underdispersive (if $\mathrm{Var}(N_s(u)) < \mathbb{E}(N_s(u))$). In case of overdispersion, we observe a stronger clustering behaviour than for PPs and in case of underdispersion the extremes occur more regularly than expected for a PP.

The *dispersion statistic*

$$\Psi_s = \frac{\mathrm{Var}(N_s(u))}{\mathbb{E}(N_s(u))} - 1 \tag{3.2}$$

equals zero in case of a PP and can be used for measuring the dispersion and therefore the clustering behaviour (see Dacre and Pinto, 2020; Mailier et al., 2006).

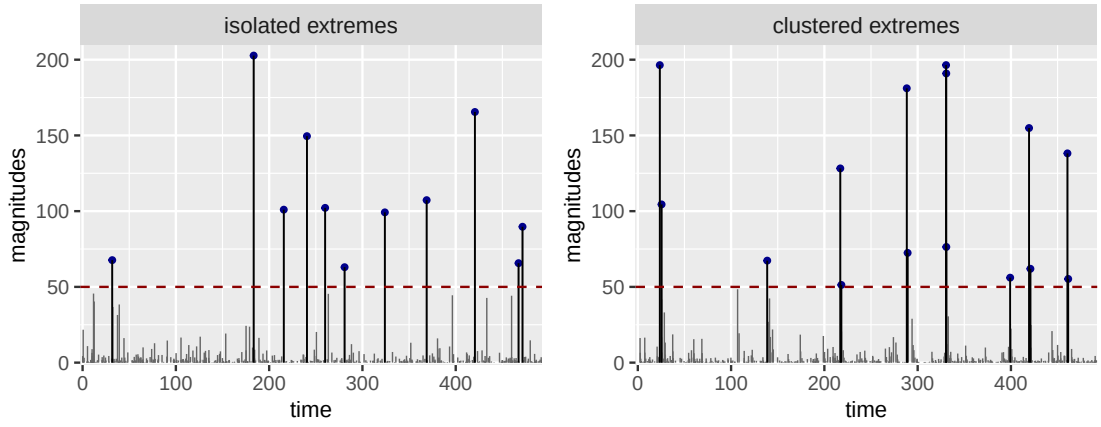

Figure 3.2: Comparison of simulated time series with isolated exceedances (left) and clustered exceedances (right).

3.2 Established models for IETs

In the following three subsections, we introduce three models for describing interexceedance times (IETs). We establish a connection to renewal process theory, which we use to explain clustering behaviour.

In classical EVT, time series are typically considered in which events occur at deterministic, equidistant time points. In the first model, we additionally assume i.i.d. event magnitudes. This leads, in the limit, to exponentially distributed IETs, and the model corresponds to a Poisson process. As such, it does not capture

clustering and serves primarily as a benchmark, as many practitioners adopt this naive approach for modelling waiting times, and hence IETs. The second model assumes that the event magnitudes form a stationary sequence. This setting is well known in classical EVT and leads to dependent extremes, which may exhibit clustering. In the third model, the i.i.d. event magnitudes are combined with random and heavy-tailed waiting times. This model goes beyond the classical framework and is described in detail by Hees et al. (2021). It allows for a different type of clustering behaviour. Both of these models describe time series with clustered extreme events.

3.2.1 Independent event magnitudes

Firstly, we consider the simple case of a sequence $(X_n)_{n\in\mathbb{N}}$ of i.i.d. event magnitudes with c.d.f. F and equidistant waiting times $W_n \equiv 1$. Let $(u_n)_{n\in\mathbb{N}}$ be a sequence of thresholds such that $u_n \to x_R$ and $n\,p(u_n) \to \nu \in (0,\infty)$, with $x_R = \sup\{x \mid F(x) < \infty\}$ being the right endpoint. It holds that $F(u_n)^n \to \exp(-\nu)$ (see Equation (2.28)). For the stopping time $\eta(u_n)$, we then receive for all $t > 0$

$$\begin{aligned}\mathbb{P}(p(u_n)\eta(u_n) > t) &= \mathbb{P}\left(\eta(u_n) > \tfrac{t}{p(u_n)}\right) = \mathbb{P}(M_{\lfloor \frac{t}{p(u_n)} \rfloor} \leqslant u_n) \\ &= F(u_n)^{\lfloor \frac{t}{p(u_n)} \rfloor} \sim F(u_n)^{\frac{t}{p(u_n)}} = (F(u_n)^n)^{\frac{t}{n\,p(u_n)}} \\ &\to \exp(-\nu)^{\frac{t}{\nu}} = \exp(-t) \quad \text{as} \quad n \to \infty.\end{aligned} \tag{3.3}$$

Note, that the convergence is independent of ν and thus of the chosen threshold sequence. In total, we get

$$p(u)\eta(u) \xrightarrow{d} T_{(1)} \quad \text{as} \quad u \to x_R, \tag{3.4}$$

where $T_{(\lambda)}$ is exponential distributed with rate λ.

Since $T(u) = \sum_{i=1}^{\eta(u)} W_i = \sum_{i=1}^{\eta(u)} 1 = \eta(u)$, we receive

$$p(u)T(u) \xrightarrow{d} T_{(1)} \quad \text{as} \quad u \to x_R. \tag{3.5}$$

It follows that the CTRE $(X_k(u), T_k(u))_{k\in\mathbb{N}}$ leads asymptotically to a Poisson process.

3.2.2 Stationary event magnitudes

Let $(X_n)_{n\in\mathbb{N}}$ be a strictly stationary sequence with extremal index $\theta \in (0,1]$, i.e., there are sequences $(u_n)_{n\in\mathbb{N}}$ that the $D(u_n)$ condition holds, and waiting times $W_n \equiv 1$ for

all $n \in \mathbb{N}$. Ferro and Segers (2003) state that then, for all $t > 0$,

$$p(u)T(u) \xrightarrow{d} T_{(\theta,1)} \quad \text{as} \quad u \to x_R, \tag{3.6}$$

where $T_{(\theta,\lambda)}$ is a random variable with mixture distribution $(1-\theta)\,\delta_0 + \theta\,\mathrm{Exp}(\lambda\theta)$, δ_0 the Dirac measure at zero and $\mathrm{Exp}(\lambda\theta)$ the exponential distribution with rate $\lambda\theta$. The c.d.f. is then

$$F^{(\theta,\lambda)}(t) = 1 - \theta\exp(-\lambda\theta t) \quad \text{for} \quad t \geqslant 0. \tag{3.7}$$

A proof using a stricter mixing condition is given in Ferro and Segers (2003), the result remains for the $D(u_n)$ condition (see Beirlant et al., 2004; Segers, 2002). The author of this dissertation has presented a more detailed version of the proof in her Master's thesis to make it accessible to undergraduate students (Meschede, 2018). This proof has been translated into English, adapted to the notation of this work, and is included in the Appendix.

Hsing et al. (1988) stated that under a slightly stricter mixing condition the CTRE $(X_k(u), T_k(u))_{k\in\mathbb{N}}$ leads to a compound Poisson process with cluster sizes as marks. Since it illustrates that in this model the exceedances occur in clusters and since with help of the CPP we can calculate the dispersion statistic, we will give a brief derivation based on Beirlant et al. (2004, Chapter 10).

Derivation of the compound Poisson process

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of stationary random variables with extremal index $\theta \in (0,1]$ and waiting times $W_n \equiv 1$ for all $n \in \mathbb{N}$. Let $(u_n)_{n\in\mathbb{N}}$ be a sequence of thresholds and $(s_n)_{n\in\mathbb{N}}$ a sequence of integers such that the following mixing condition $\Delta(u_n)$ holds:

Condition 2. *Let $\mathcal{F}_{j,k}(u_n) = \sigma(\{X_i > u_n\} \mid j \leqslant i \leqslant k)$ be a σ-algebra generated by $\{X_i > u_n\}$, $i = j, \ldots, k$, with integers $j < k$. For all $A_1 \in \mathcal{F}_{1,l}(u_n)$, $A_2 \in \mathcal{F}_{l+s,n}(u_n)$ and $1 \leqslant l \leqslant n-s$,*

$$|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| \leqslant \alpha(n,s),$$

and $\alpha(n, s_n) \to 0$ as $n \to \infty$ for some positive integer s_n with $s_n = o(n)$. This mixing condition is called $\Delta(u_n)$.

Further, we choose the sequence of thresholds $(u_n)_{n\in\mathbb{N}}$ such that $np(u_n) \to \nu \in (0,\infty)$ and $\mathbb{P}(M_n \leqslant u_n) \to \exp(-\nu\theta)$ as $n \to \infty$, and another sequence of integers $(r_n)_{n\in\mathbb{N}}$ such that $s_n = o(r_n)$ and $r_n = o(n)$, e.g., $r_n = \lfloor \max\{\sqrt{n\,s_n}, n\sqrt{\alpha(n,s_n)}\} \rfloor$.

We part the sequence in consecutive blocks of length r_n and consider the first $k_n = \left\lfloor \frac{n}{r_n} \right\rfloor$ blocks, $J_1, \ldots, J_{k_n}$. Each block maximum $M(J_1), \ldots, M(J_{k_n})$ is distributed as $M_{r_n} = \max\{X_1, \ldots, X_{r_n}\}$.

From this assumption it follows:

1. The probability that none of the next r_n events after one magnitude exceeded the threshold converges towards θ (O'Brien, 1987), i.e.,

$$\mathbb{P}(M_{r_n} \leqslant u_n \mid X_0 > u_n) \to \theta. \tag{3.8}$$

2. The expected number of exceedances of a block with length r_n with at least one excess converges towards $1/\theta$ (Hsing et al., 1988), i.e.,

$$\mathbb{E}\left(\sum_{i=1}^{r_n} \mathbb{1}\{X_i > u_n\} \mid M_{r_n} > u_n\right) = \frac{r_n\, p(u_n)}{\mathbb{P}(M_{r_n} > u_n)} \to \frac{1}{\theta}. \tag{3.9}$$

3. $\mathbb{1}\{M(J_1) \leqslant u_n\}, \ldots, \mathbb{1}\{M(J_{k_n}) \leqslant u_n\}$ are asymptotially independent.

4. $\mathbb{P}(\bigcup_{i=1}^{k_n}\{M(J_{i,s_n}) \geqslant u_n\}) \to 0$ as $n \to \infty$, where $M(J_{i,s_n})$ is the maximum of the last s_n events of the i-th block, i.e. $M(J_{i,s_n}) = \max\{X_{i\cdot r_n - s_n + 1}, \ldots, X_{i\cdot r_n}\}$. It means that, in the limit, the exceedances of different blocks are almost surely seperated by at least s_n events that do not exceed the threshold.

For $n < \infty$, we now identify a cluster as the exceedances of one block of length r_n. Point 1 states that in case of $\theta = 1$, all exceedances are almost surely separated by at least r_n events which do not exceed the threshold, i.e., in one block of size r_n an excess is almost surely isolated, while for $\theta < 1$ the probability that there are more exceedances within the next r_n events is $1 - \theta$. Point 4 says that the clusters are at least separated by length s_n.

Due to point 3 and the Poisson approximation the number of clusters

$$\sum_{i=1}^{k_n} \mathbb{1}\{M(J_i) > u_n\} \tag{3.10}$$

is asymptotically Poisson distributed with rate $\nu\theta = \lim_{n\to\infty} k_n\, \mathbb{P}(M_{r_n} > u_n)$. Note, that the expected number of cluster within n event magnitudes therefore converges towards $\nu\theta$.

By scaling the time by $p(u_n)$, we obtain $T_i = i\, p(u_n)$, such that the first n events occur asymptotically in the interval $[0, \nu]$, since $\lim_{n\to\infty} np(u_n) = \nu$. The scaling has the advantage that the time interval does not increase to infinity for raising n. Therefore, the number of clusters until time ν is Poisson distributed with rate $\theta\nu$ and in general

the number of clusters until time t is Poisson distributed with rate θt. Since the scaled block size $r_n p(u_n)$ converges towards zero, in limit every point on the interval $[0, \infty)$ is the relative position of one block and thus the time between two clusters, i.e., the time between two blocks which both exceed the threshold, is asymptotically exponential distributed with rate θ.

We can define the distribution of the cluster size as

$$\pi_n(j) = \mathbb{P}\left(\sum_{i=1}^{r_n} \mathbb{1}\{X_i > u_n\} = j \mid M_{r_n} > u_n\right), \quad j \in \mathbb{N}. \tag{3.11}$$

Point 2 says that the expected cluster size is asymptotically $1/\theta$, more precisely $\lim\limits_{n\to\infty} \sum\limits_{j=1}^{\infty} j\, \pi_n(j) = \frac{1}{\theta}$. Note that even if $\pi_n \to \pi$, in general $\sum\limits_{j=1}^{\infty} \pi(j) \neq \frac{1}{\theta}$, but Hsing et al. (1988) gave some extra assumption under which $\sum\limits_{j=1}^{\infty} \pi(j) = \frac{1}{\theta}$ is true.

We consider for all $n \in \mathbb{N}$ the process $(Z_{n,t})_{t>0}$ with

$$Z_{n,t} = \sum_{i=1}^{\infty} \mathbb{1}\left\{X_i > u_n\right\} \mathbb{1}\left\{ip(u_n) \leqslant t\right\}. \tag{3.12}$$

$Z_{n,t}$ counts the exceedances until time t.

Let $(Z_t)_{t>0}$ be the limiting process of $(Z_{n,t})_{t>0}$, such that $Z_{n,t} \to Z_t$ for every $t > 0$. Thus, if there is a distribution π, such that $\pi_n(j) \to \pi(j)$ for every $j \in \mathbb{N}$, the process $(Z_t)_{t>0}$ is a compound Poisson process as defined in Section 2.3.2 with the cluster sizes $(Y_n)_{n\in\mathbb{N}}$ as marks, i.e.,

$$Z_t \sim CPP(\theta, \pi). \tag{3.13}$$

Cluster interpretation

With this mathematical results, we can interpret that for $\theta = 1$ the exceedances occur isolated, while for $\theta < 1$ the exceedances appear in clusters. The mean cluster size is asymptotically $1/\theta$, i.e., the smaller θ, the larger the average cluster. The time between two consecutive clusters, also called *intercluster time*, is asymptotically exponential distributed, while the time between two consecutive exceedances within the same cluster, named *intracluster time*, is asymptotically zero. This fits to our asymptotically IET distribution introduced in (3.7).

Lastly, we have a look at the dispersion statistic introduced in (3.2). Applied on the

$CPP(\theta, \pi)$ we are interested in

$$\Psi_s = \frac{\mathrm{Var}(Z_s)}{\mathbb{E}(Z_s)} - 1, \tag{3.14}$$

with $Z_s = \sum_{n=1}^{N_s} Y_n$, N_s Poisson distributed with rate $s\nu\theta$ and Y_n has distribution π for all $i \in \mathbb{N}$. Using (2.40), we receive

$$\begin{aligned} \mathbb{E}(Z_s) &= \mathbb{E}(N_s)\,\mathbb{E}(Y_1) = s\,\theta\,\mathbb{E}(Y_1), \\ \mathrm{Var}(Z_s) &= \mathbb{E}(N_s)\,\mathrm{Var}(Y_1) + \mathrm{Var}(N_s)\;\mathbb{E}(Y_1)^2 = s\,\theta\,\mathbb{E}(Y_1^2) \end{aligned} \tag{3.15}$$

and thus

$$\Psi_s = \frac{\mathbb{E}(Y_1^2)}{\mathbb{E}(Y_1)} - 1. \tag{3.16}$$

The dispersion statistic only depends on the cluster size distribution π. Since each cluster includes at least one excess, it holds $\mathbb{E}(Y_1) \geqslant 1$. Therefore $\Psi_s = 0$ if and only if $\mathbb{E}(Y_1) = 1$ and $\mathrm{Var}(Y_1) = 0$, i.e., when the exceedances are almost surely isolated ($\theta = 1$). Otherwise $\Psi_s > 0$ since $\mathbb{E}(Y_1) > 1$ for $\theta < 1$ and therefore $\mathbb{E}(Y_1^2) \geqslant \mathbb{E}(Y_1)^2 > \mathbb{E}(Y_1)$. Thus, Z_s is overdispersed.

3.2.3 Heavy-tailed waiting times

Let the event magnitudes $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables as introduced in Subsection 3.2.1.

Another mechanism that leads to temporal clustering behaviour is heavy-tailed distributed waiting times $(W_n)_{n\in\mathbb{N}}$, such that the distribution tails are regular varying with index $-\alpha$, $0 < \alpha < 1$. More precisely,

$$\mathbb{P}(W > t) = t^{-\alpha}\,L(t), \tag{3.17}$$

where L is slowly varying, and it follows that $\mathbb{E}(W_1) = \infty$. A prominent example of a heavy-tailed distribution is the Pareto distribution and another example is the Mittag-Leffler distribution, we introduced in Subsection 2.1.3.

While the result in Equation (3.4), that

$$p(u)\eta(u) \to T_{(1)}, \tag{3.18}$$

still applies, Hees et al. (2021) showed that

$$\frac{T(u)}{b(1/p(u))} \xrightarrow{d} T_{(\beta,1)} \quad \text{as} \quad u \to x_R, \tag{3.19}$$

with $\beta = \alpha$ and b regularly varying with index $1/\alpha$, where $T_{(\beta,\varsigma)} \sim \mathrm{ML}(\beta, \varsigma)$. It follows that the CTRE $(X_k(u), T_k(u))_{k\in\mathbb{N}}$ leads to a fractional Poisson process.

Cluster interpretation

The Mittag-Leffler distribution possesses heavy tails on both ends, i.e. very low and high values occur more frequently than the exponential distribution would suggest. This leads to the clustering of extremes when they are seperated by Mittag-leffler distributed IETs. The lower the tail parameter is, the stronger is the clustering. Since the mechanism differs from the one in the previous Subsection 3.2.2, Hees et al. (2021) say that the extreme events occur in *bursts*.

Again we have a look at the dispersion statistic:

$$\begin{aligned} \Psi_s &= \frac{\mathrm{Var}(N_s)}{\mathbb{E}(N_s)} - 1 \overset{(2.38)}{=} \mathbb{E}(N_s)\left(\frac{\beta\,\Gamma(\beta)^2}{\Gamma(2\beta)} - 1\right) \\ &= \frac{s^\beta}{\Gamma(1+\beta)}\left(\frac{\beta\,\Gamma(\beta)^2}{\Gamma(2\beta)} - 1\right) \end{aligned} \tag{3.20}$$

For $\beta = 1$ we are in the special case of exponential distributed IETs and $\Psi_s = 0$. For $\beta < 1$, $\Psi_s > 0$. In contrary to the CPP dispersion, Ψ_s actually depends on the time interval s. Figure 3.3 illustrates the behaviour of the dispersion statistic dependent on s. It is noteworthy that the dispersion statistic decreases rather than increases for smaller tail parameters above a certain s.

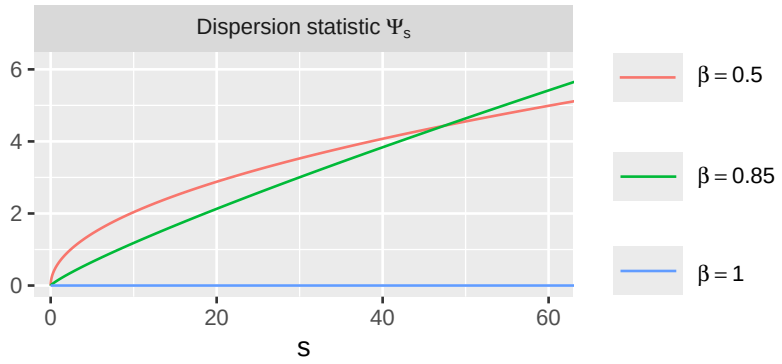

Figure 3.3: Dispersion statistic Ψ_s of an FPP for various tail parameters β.

3.3 A generalised asymptotic model

After introducing existing models for temporal clustered extreme events in the previous section, in this section we present our contribution to the topic. We provide a generalised model for IETs of temporal clustered extreme events that takes both mechanisms, stationary event magnitudes and heavy-tailed waiting times, simultaniously into account and includes the three models of the previous section. Additionally, we show that the result of the Subsection 3.2.2 remains true for random waiting times with finite mean. We show that then the CTRE leads to a fractional compound Poisson process.

3.3.1 Theorem and proof

Theorem 1. *Let $T_{(\beta,\theta,\sigma)}$ be a random variable with mixture distribution*

$$\mathbb{P}_{(\beta,\theta,\sigma)} = (1-\theta)\,\delta_0 + \theta\, ML(\beta, \theta^{-1/\beta}\,\sigma), \tag{3.21}$$

where δ_0 is the Dirac measure at point zero, $ML(\beta,\varsigma)$ is the Mittag-Leffler distribution with tail parameter β and scale parameter ς, and $\beta \in (0,1]$, $\theta \in (0,1]$ and $\sigma > 0$.

Let $(X_n)_{n\in\mathbb{N}}$ be a strictly stationary process of magnitudes in the max-domain of attraction of a GEV distribution with extremal index $\theta \in (0,1]$, x_R is the right endpoint of the distribution of X_1, and $p(u) = \mathbb{P}(X_1 > u)$ is the excess probability.

Let $(T_n)_{n\in\mathbb{N}}$ be corresponding arrival times such that the sequence of waiting times $(W_n)_{n\in\mathbb{N}}$, with $W_n = T_n - T_{n-1}$, are i.i.d. and stochastically independent of the magnitudes $(X_n)_{n\in\mathbb{N}}$.

Without loss of generality, we set $X_0 > u$ and $T_0 = 0$. The interexceedance time regarding the process $(X_n, W_n)_{n\in\mathbb{N}}$ and a threshold u is defined as

$$T(u) = \sum_{i=1}^{\eta(u)} W_i, \quad \text{with stopping time } \eta(u) = \min\{k \geqslant 1 \mid X_k > u\}.$$

(i) If $\mathbb{E}(W_n) = \frac{1}{\lambda} < \infty$ for all $n \in \mathbb{N}$,

$$\frac{T(u)}{b(1/p(u))} \xrightarrow{d} T_{(\beta,\theta,1)} \quad as \quad u \to x_R, \tag{3.22}$$

with tail parameter $\beta = 1$ and $b(n) = n\,\frac{1}{\lambda}$ regularly varying with index 1.

(ii) If $\mathbb{E}(W_n) = \infty$*, for all* $n \in \mathbb{N}$*, and* $\mathbb{P}(W_n > t)$ *regularly varying with index* $-\alpha$*,* $0 < \alpha < 1$*,*

$$\frac{T(u)}{b(1/p(u))} \xrightarrow{d} T_{(\beta,\theta,1)} \quad as \quad u \to x_R, \tag{3.23}$$

with tail parameter $\beta = \alpha$ *and* b *regularly varying with index* $1/\alpha$*.*

Proof.

(i) The proof is analogous to that of Theorem 1 in Gut and Hüsler (1999), who consider the case of $(X_n)_{n\in\mathbb{N}}$ being stochastically independent. Using (3.4) we have

$$\begin{aligned} p(u)T(u) &= p(u)\eta(u)\,\frac{1}{\eta(u)}\sum_{i=1}^{\eta(u)} W_i \\ &\xrightarrow{d} T_{(1,\theta,1)} \cdot \tfrac{1}{\lambda} = T_{(1,\theta,1/\lambda)} \quad \text{as} \quad u \to x_R, \end{aligned} \tag{3.24}$$

since we can deduce from Theorem 2 of Richter (1965) together with Kolmogorov's SLLN that $\dfrac{1}{\eta(u)}\sum_{i=1}^{\eta(u)} W_i \to \mathbb{E}(W_1) = \dfrac{1}{\lambda}$ in probability if $\eta(u) \to \infty$ in probability as $u \to x_R$. The stopping time fulfills $\eta(u) \to \infty$ since for all $m \in \mathbb{N}$

$$\begin{aligned} \mathbb{P}(\eta(u) > m) &= \mathbb{P}(X_1 \leqslant u, \ldots, X_m \leqslant u) = 1 - \mathbb{P}\left(\textstyle\bigcup_{i=1}^{m}\{X_i > u\}\right) \\ &\geqslant 1 - \sum_{i=1}^{m} \mathbb{P}\left(X_i > u\right) = 1 - m \cdot \mathbb{P}(X_1 > u) \to 1. \end{aligned} \tag{3.25}$$

(ii) Consider an abitrary sequence of thresholds $(u_n)_{n\in\mathbb{N}}$ such that the $D(u_n)$ condition is fulfilled and $n\,p(u_n) \to \nu$ as $n \to \infty$ with $0 < \nu < \infty$. Obviously, we have $u_n \to x_R$ as $n \to \infty$. It follows from (3.4) that

$$\frac{\eta(u_n)}{n} = \frac{p(u_n)\eta(u_n)}{n \cdot p(u_n)} \xrightarrow{d} E, \tag{3.26}$$

where $E \overset{d}{=} T_{(1,\theta,1/\nu)} \sim (1-\theta)\cdot\delta_0 + \theta\cdot\mathrm{Exp}(\nu\theta)$ with $\mathrm{Exp}(\nu\theta)$ being the exponential distribution with rate $\nu\theta$. For $0 < \alpha < 1$ the waiting time distribution is in the strict domain of attraction of a positively skewed sum-stable distribution with stability parameter α and Laplace transform

$$\mathcal{L}_{D_\alpha}(s) := \mathbb{E}(\exp(-sD_\alpha)) = \exp(-s^\alpha). \tag{3.27}$$

More precisely, there is a regularly varying function b with index $1/\alpha$ such that

$$\frac{W_1 + \cdots + W_n}{b(n)} \xrightarrow{d} D_\alpha, \quad n \to \infty, \tag{3.28}$$

(see Meerschaert and Sikorskii, 2012, Theorem 4.5 and Propositions 4.15 & 4.16). Using (3.27) and Gnedenko's transfer theorem (see Gnedenko, 1983), it follows

$$\sum_{i=1}^{\eta(u_n)} \frac{W_i}{b(n)} \xrightarrow{d} Z, \quad \text{as} \quad n \to \infty, \tag{3.29}$$

where the distribution of Z has the Laplace transform

$$\begin{aligned} \mathcal{L}_Z(s) &= \textstyle\int_0^\infty (\mathcal{L}_{D_\alpha}(s))^y \, \mathrm{d}\mathbb{P}^E(y) \\ &= (1-\theta) \textstyle\int_0^\infty (\mathcal{L}_{D_\alpha}(s))^y \, \mathrm{d}\delta_0(y) + \theta \int_0^\infty (\mathcal{L}_{D_\alpha}(s))^y \, \mathrm{d}\mu_{\nu\theta}(y) \\ &= (1-\theta) + \theta \tfrac{1}{1-\log(\mathcal{L}_{D_\alpha}(s))/(\nu\theta)} = (1-\theta) + \theta \tfrac{1}{1+(s(\nu\theta)^{-1/\alpha})^\alpha)} \end{aligned} \tag{3.30}$$

which is the Laplace transform of the mixture distribution

$$(1-\theta) \cdot \delta_0 + \theta \cdot \mathrm{ML}(\alpha, (\nu\theta)^{-1/\alpha}). \tag{3.31}$$

The Laplace transform $\mathcal{L}_{D_\alpha}$ of D_α is the same as in (3.27) and $\mu_{\nu\theta}$ denotes the exponential measure with rate $\nu\theta$.

Since b is a regularly varying function, we have

$$\frac{b(n)}{b(1/p(u_n))} = \frac{b(n)}{b(n/(n \cdot p(u_n)))} \sim (n \cdot p(u_n))^{1/\alpha} \to \nu^{1/\alpha} \quad \text{as} \quad n \to \infty, \tag{3.32}$$

where $f(x) \sim g(x)$ for real valued functions f and g means $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$. We finally get

$$\frac{1}{b(1/p(u_n))} \sum_{i=1}^{\eta(u_n)} W_i = \frac{b(n)}{b(1/p(u_n))} \, \frac{1}{b(n)} \sum_{i=1}^{\eta(u_n)} W_i \xrightarrow{d} \nu^{1/\alpha} Z \tag{3.33}$$

with $\nu^{1/\alpha} Z \sim (1-\theta) \cdot \delta_0 + \theta \cdot \mathrm{ML}(\alpha, \theta^{-1/\alpha})$. Since $u_n \to x_R$ has been arbitrary, the assertion of the Theorem follows with $u := u_n$ and $p(u) := p(u_n)$.

□

3.3.2 (Fractional) Compound Poisson process

In Subsection 3.2.2, we derived that for equidistant waiting times $W_n \equiv 1$ and the fulfillment of the mixing condition $\Delta(u_n)$, the CTRE $(X_k(u), T_k(u))_{k\in\mathbb{N}}$ leads to a compound Poisson process with the cluster size as marks.

As far as we know, there is no literature that deals with the more general case of i.i.d. random waiting times $(W_n)_{n\in\mathbb{N}}$ with $\mathbb{E}(W_1) = 1/\lambda < \infty$. Therefore, we show that the result

$$Z_t \sim CPP(\theta, \pi) \tag{3.34}$$

remains true for $\beta = 1$, and in addition, we show that for $0 < \beta < 1$ the CTRE $(X_k(u), T_k(u))_{k\in\mathbb{N}}$ leads to a fractional compound Poisson process.

Derivation of the compound Poisson process for $\beta = 1$

Taking random waiting times into account, we have to distinguish between the block size r_n, i.e., the number of events within each block, and the block time length which is random and distributed as $T_{r_n} = \sum_{i=1}^{r_n} W_i$. The by $\lambda\, p(u_n)$ scaled process $(Z_{n,t})_{t>0}$ is then defined as

$$Z_{n,t} = \sum_{i=1}^{\infty} \mathbb{1}\{X_i > u_n\}\, \mathbb{1}\{\lambda\, p(u_n)\, T_i \leqslant t\}\,.$$

Due to the SLLN

$$p(u_n)\,\lambda\, T_n = n\, p(u_n)\,\lambda\, \frac{1}{n} T_n \to \nu\,\lambda\,\frac{1}{\lambda} = \nu \quad \text{almost surely,} \tag{3.35}$$

and

$$p(u_n)\,\lambda T_{r_n} = \frac{r_n}{n}\, n\, p(u_n)\,\lambda \frac{1}{r_n} T_{r_n} \sim \frac{r_n}{n}\nu\,\lambda\,\frac{1}{\lambda} = \frac{r_n}{n}\,\nu \to 0 \quad \text{almost surely.} \tag{3.36}$$

Thus, when the time is scaled by $p(u_n)\,\lambda$, the first n events occur asymptotically in the interval $[0, \nu]$ and the scaled block time length converges towards zero. Therefore, the result remains true.

Derivation of the fractional compound Poisson process for $\beta < 1$

For the second case with $0 < \beta < 1$, the results from Subsection 3.2.2 still remain since the event magnitudes and waiting times are stochastically independent. Thus,

for a sample $X_1, \ldots, X_{k_n \cdot r_n}$, with $k_n = \lfloor \frac{n}{r_n} \rfloor$, the number of clusters

$$\sum_{i=1}^{k_n} \mathbb{1}\{M(J_i) > u_n\} \quad \text{with} \quad M(J_i) = \max\{X_{(i-1)\,r_n+1}, \ldots, X_{i\,r_n}\}$$

is asymptotically Poisson distributed with rate $\nu\theta$ due to the Poisson approximation.

Since $W_1 > 0$ is in the domain of attraction of the β-stable distribution $D_\beta > 0$, we receive

$$\begin{aligned} \frac{T_n}{b(1/p(u_n))} &= \frac{b(n)}{b(1/p(u_n))} \frac{T_n}{b(n)} = \frac{b(n)}{b(n/(n\,p(u_n))} \frac{T_n}{b(n)} \\ &\sim \frac{b(n)}{b(\frac{1}{\nu}\,n)} \frac{T_n}{b(n)} \xrightarrow{d} \nu^{1/\beta} D_\beta. \end{aligned} \tag{3.37}$$

Furthermore, we receive

$$\frac{T_{r_n}}{b(1/p(u_n))} = \frac{b(r_n)}{b(n)} \frac{b(n)}{b(1/p(u_n))} \frac{T_{r_n}}{b(r_n)} \sim \left(\frac{r_n}{n}\right)^{1/\beta} \nu^{1/\beta} D_\beta \to 0 \tag{3.38}$$

weakly, using the Uniform Convergence Theorem for regularly varying functions (see Bingham et al., 1987, Theorem 1.5.2). Thus, the scaled block time length converges in probability to zero, so that in the limit, every point on the inteval $[0, \infty)$ corresponds to the position of one block.

Let $\tau(u_n) = \min\{k \geqslant 1 \mid M(J_k) > u_n\}$ be the number of blocks until the first block that exceeds the threshold. It holds that $\frac{\tau(u_n)}{k_n}$ converges towards an exponential distribution with rate $\nu\theta$. Now, the time until the first block that exceeds the threshold can be defined as

$$S(u_n) = \sum_{i=1}^{\tau(u_n)} T_{i,r_n}, \tag{3.39}$$

where T_{i,r_n} is the time length of the i-th block and thus, all T_{i,r_n}, $i \in \mathbb{N}$, are stochastically independent and distributed as T_{r_n}. Using Gnedenko's transfer theorem analogously as in the proof of Theorem 1, we receive

$$\frac{S(u_n)}{b(1/p(u_n))} \xrightarrow{d} ML(\beta, \theta^{-1/\beta}). \tag{3.40}$$

In total, with the same requirements for the event magnitudes $(X_n)_{n \in N}$ and if π_n converges towards π as in Subsection 3.2.2, the CTRE $(X_k(u), W_k(u))_{k \in \mathbb{N}}$ leads to

a fractional compound Process, $Z_t \sim FCPP((\beta, \theta^{-1/\beta}), \pi)$ with the cluster size as marks.

3.3.3 Cluster interpretation

From the perspective of the FCPP, exceedances for $\theta < 1$ occur in clusters, in the sense that the intra-cluster times are asymptotically zero, while for $\beta < 1$ the times between two consecutive clusters asymptotically follow a Mittag-Leffler distribution. Hence, while in Subsection 3.2.3 the extremes appeared in bursts, we now observe that the clusters themselves occur in bursts. Due to the heavy-tailed nature of the Mittag-Leffler distribution, both very short and very long inter-cluster times are more frequent. As a result, in finite CTREs, short inter-cluster times may be indistinguishable from intra-cluster times. In other words, due to the strong variability induced by the Mittag-Leffler distribution, multiple clusters may subjectively appear as a single large cluster. For $\beta = 1$, we have $ML(\beta, \theta) = \text{Exp}(1/\sigma)$, placing us in the classical Poissonian setting discussed in Subsection 3.2.2. Conversely, when $\theta = 1$, the Dirac component of the distribution vanishes, corresponding to the case described in Subsection 3.2.3. As illustrated in Figure 3.4, both parameters influence the distribution of the IETs $T(u)$ and, consequently, the clustering behavior. While β affects both tails of the distribution, smaller values of θ lead to stronger zero-inflation and primarily affect the lower tail.

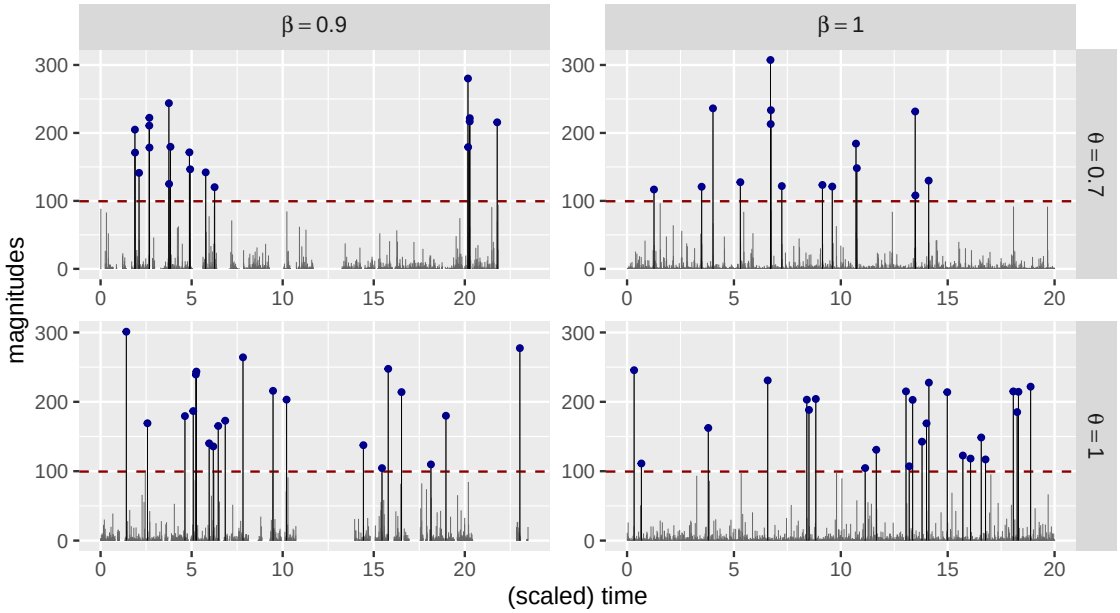

Figure 3.4: Comparison of CTREs with tail parameter and extremal index equal to one and smaller than one using simulated data.

Using the formulas of (2.38) and (2.40), the dispersion statistic is given by

$$\begin{aligned}\Psi_s &= \frac{\mathrm{Var}(Z_s)}{\mathbb{E}(Z_s)} - 1 = \frac{\mathbb{E}(N_s)\mathrm{Var}(Y_1) + \mathrm{Var}(N_s)\mathbb{E}(Y_1)^2}{\mathbb{E}(N_s)E(Y_1)} - 1 \\ &= \frac{\mathrm{Var}(Y_1)}{\mathbb{E}(Y_1)} + \frac{\mathrm{Var}(N_s)\mathbb{E}(Y_1)}{\mathbb{E}(N_s)} - 1 \\ &= \frac{\mathrm{Var}(Y_1)}{\mathbb{E}(Y_1)} + \left(\mathbb{E}(N_s)\left(\frac{\beta\Gamma(\beta)^2}{\Gamma(2\beta)} - 1\right) + 1\right)\mathbb{E}(Y_1) - 1 \\ &= \frac{\mathrm{Var}(Y_1)}{\mathbb{E}(Y_1)} + \left(\theta s^{\beta}\,\frac{\beta\Gamma(\beta)^2 - \Gamma(2\beta)}{\beta\Gamma(\beta)\Gamma(2\beta)} + 1\right)\mathbb{E}(Y_1) - 1.\end{aligned} \tag{3.41}$$

If $\beta < 1$ or $\theta < 1$, $\Psi_s > 0$ holds and therefore Z_s is overdispersed for all $s > 0$. As for the FPP, the dispersion statistic depends not only on the distribution of the cluster sizes Y_1 and the parameter β and θ, but also on the time length s (see Figure 3.5).

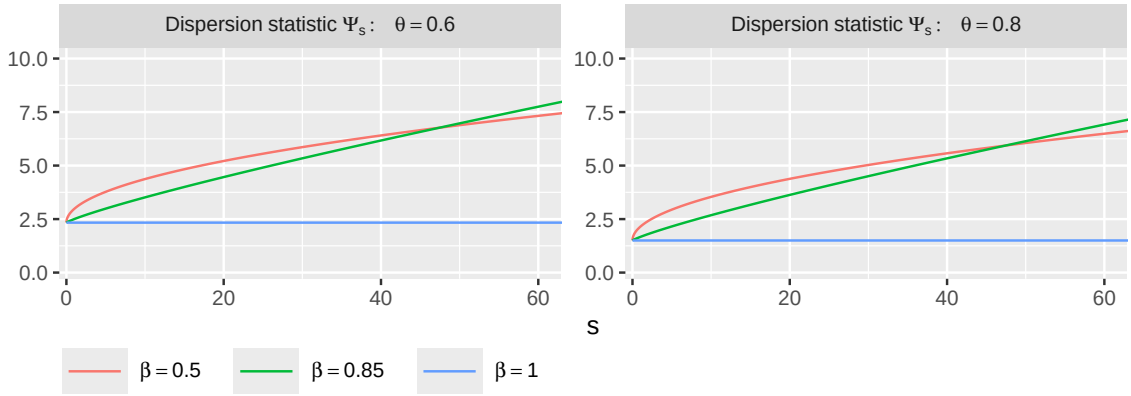

Figure 3.5: Dispersion statistic Ψ_s of an FCPP for various tail parameters β, extremal index $\theta = 0.6$ (left) and $\theta = 0.8$ (right), and $\mathbb{E}(Y_1) = 1/\theta$ and $\mathrm{Var}(Y_1) = 1/\theta^2$.

3.4 Conclusions and discussion

In this chapter, we presented models that describe clustering behaviour of extreme events. These models are based on the assumption of an underlying process $(X_n, W_n)_{n\in\mathbb{N}}$, where the magnitudes and waiting times are stochastically independent. A well-established clustering model in EVT assumes stationary magnitudes and equidistant waiting times. In this case, the IETs asymptotically follow a mixture distribution, consisting of a Dirac measure at zero and an exponential distribution. The extremal index $\theta \in (0, 1]$ determines the mixture weights and thereby governs the clustering strength. We extended this model by introducing i.i.d. random waiting

times with finite means, and showed that this generalisation leads to the same asymptotic results. Subsequently, we investigated a further extension in which the i.i.d. waiting times follow a heavy-tailed distribution with infinite mean. Under the assumption of i.i.d. magnitudes, it is known that the IETs are asymptotically Mittag-Leffler distributed. We showed that, in the stationary case, the IETs again follow a mixture distribution with the Mittag-Leffler distribution replacing the exponential component. This shift yields a FCPP in place of the CPP. In summary, we have developed a generalised model that includes the previously discussed models as special cases and offers greater flexibility in terms of clustering behaviour. With accurate parameter estimation, this model may eliminate the need for explicit model selection.

Beyond the models presented in this chapter, there are other models that describe clustering behaviour, but they do not belong to the same class of models using the POT approach and thus are not described here. These include so-called *self-exciting processes*, such as *Hawkes processes*. In a Hawkes process, each event increases the likelihood of further events occurring in the near future, an effect referred to as excitation. In other words, the occurrence of an event temporarily raises the probability of subsequent events. This mechanism is mathematically described by an intensity function that adapts the rate of future events based on past occurrences. For further details, see, e.g., Laub et al. (2021). Another relevant class of models are *log-Gaussian Cox processes* (see, e.g., Møller et al., 1998). These models describe spatially and temporally inhomogeneous and clustered events by modelling the event rate not as a constant or deterministic function, as in Poisson processes, but stochastically. More precisely, the intensity function is given by the exponential transformation of a Gaussian process. In this case, clustering arises from an underlying, time-varying latent influence.

Moreover, one may question whether the assumption of independence in the underlying process $(X_n, W_n)_{n\in\mathbb{N}}$ is too restrictive for applications to real-world data, such as the climate data discussed in Chapter 7. On the one hand, we require the independence between magnitudes and waiting times for our asymptotic arguments. On the other hand, it is arguably plausible in various settings that the length of the preceding waiting time (or even of several past waiting times) may influence the magnitude of the next event. Relaxing the independence assumption, however, is a non-trivial task and would require considerable theoretical and methodological effort in future research. Another potential relaxation for future work concerns the assumption of independence or stationarity among the magnitudes themselves. In many practical applications, e.g., climate date, it is realistic to expect the presence of trends or seasonality. Such features inherently introduce dependence structures that violate the assumption of stationary magnitudes. Incorporating non-stationarity or serial dependence in the extremes would therefore constitute a meaningful and relevant extension of the current modelling framework.

4 Construction of an estimation method

In Chapter 3, we derived that under certain assumptions the IETs, which are identically distributed as $T(u)$, of a CTRE are asymptotically mixture distributed as described in Theorem 1. This result implies that for u near to the right endpoint x_R the distribution of $T(u)$ can be approximated by

$$T(u) \overset{d}{\approx} (1-\theta)\,\delta_0 + \theta\,\mathrm{ML}(\beta, \theta^{-1/\beta}\,\sigma_u), \tag{4.1}$$

where $\sigma_u = b(1/p(u)) = p(u)^{-1/\beta}\,L(1/p(u))$ with L denoting a slowly varying function and $p(u) = \mathbb{P}(X_1 > u)$.

So, if we can estimate the parameters β, θ and σ_u, we are able to fit the model to empirical data and take advantage of its flexibility, thereby avoiding the need for prior model selection. This chapter is dedicated to this objective.

We assume we have a finite marked point process $(X_i, W_i)_{i=1,\dots,n}$ as sample with n events, and a chosen threshold u such that $X_0 > u$. By applying the POT approach, we receive the CTRE $(X_i(u), T_i(u))_{i=1,\dots,k}$ with k exceedances (not counting X_0) and k IETs. Note that $k = \sum_{i=1}^{n} \mathbb{1}\{X_i > u\}$ is actually a random variable and dependent on n and u. Since in this chapter we do not focus on asymptotic behaviour, we neglect this aspect for now.

This chapter is structured as follows: In the first section, we describe the estimation methods used to fit PP, CPP and FPP submodels. On the one hand, this is done to explain why these methods have not simply been transferred to the generalised model, the FCPP. On the other hand, some of these methods will later on be used in the simulation study for comparison purposes. Afterwards, we discuss the challenges and difficulties of finding a suitable and well performing method for the FCPP model. We then propose a minimum distance estimator with a modification of the Cramér-von Mises criterion. We conclude the chapter with a summary and discussion.

4.1 Established estimation methods for submodels

For all three submodels, various estimation methods exist. We focus on procedures based on the interexceedance times. Table 4.1 summarises the parameters of each model including the FCPP.

Table 4.1: Summary of the model parameters for all four (sub)models.

	PP	CPP	FPP	FCPP
tail parameter β	×	×	✓	✓
extremal index θ	×	✓	×	✓
scale parameter σ_u	✓	✓	✓	✓

The scale parameter $\sigma_u = p(u)^{-1/\beta} L(1/p(u))$ can be decomposed into two components: $p(u)$ and $l_u = L(1/p(u))$, which can be estimated separately. For all models, the exceedance probability $p(u)$ can be estimated by the empirical exceedance frequency

$$\hat{p}_n(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i > u\} = \frac{k}{n} \tag{4.2}$$

or by more advanced methods as discussed in Section 2.2.

4.1.1 Poisson process

If the CTRE leads to a PP, the IETs follow asymptotically an exponential distribution with rate $\sigma_u^{-1} = \dfrac{p(u)}{\mathbb{E}(W_1)}$. Thus, σ_u can be estimated by estimating $p(u)$ and $\mathbb{E}(W_1)$ separately using $\hat{p}_n(u) = \dfrac{k}{n}$ and the sample mean of the waiting times $W_1, \ldots, W_n$. Alternatively, we can estimate σ_u by the sample mean of the IETs $T_1(u), \ldots, T_k(u)$. For a suitable threshold sequence $(u_n)_{n\in\mathbb{N}}$, both approaches yield asymptotically equivalent estimators, as

$$\frac{n}{k} \frac{1}{n} \sum_{i=1}^{n} W_i = \frac{1}{k} \sum_{i=1}^{n} W_i = \frac{1}{k} \sum_{i=1}^{k} T_i(u_n) + \frac{1}{k} \sum_{i=\nu_k+1}^{n} W_i = \frac{1}{k} \sum_{i=1}^{k} T_i(u_n) + o(1), \tag{4.3}$$

where X_{ν_k} is the k-th exceedance. The main advantage of the alternative estimator is that it does not require information about the underlying process $(X_i, W_i)_{i=1,\ldots,n}$.

4.1.2 Compound Poisson process

If the CTRE leads to a CPP, the IETs follow asymptotically a mixture distribution with the Dirac measure at point zero and the exponential distribution as components. The scale parameter σ_u can be estimated as in Subsection 4.1.1.

As described in Subsection 3.2.2, in EVT typically deterministic, equidistant waiting times $W_n \equiv 1$, $n \in \mathbb{N}$, are considered. In this setting, a wide range of estimators for the extremal index θ have been proposed (see, e.g., Ferreira, 2024). Most of them need auxility parameters that must be chosen by the statistican additionally to the threshold. According to Beirlant et al. (2004, Subsection 10.3.4) and Ferreira (2018), estimators can be grouped into three classes: block estimators based on (3.9), runs estimators based on (3.8), or interexceedance times estimators based on (3.7). We focus on the third category, which are initially introduced by Ferro and Segers (2003).

Interval estimator

The interval estimator proposed by Ferro and Segers (2003) is defined as

$$\hat{\theta}_n^{(I)} = \begin{cases} \max\left\{1, \dfrac{2\left(\sum_{i=1}^{k} T_i(u)\right)^2}{k\sum_{i=1}^{k} T_i(u)^2}\right\}, & \text{if all } \ T_i(u) \leqslant 2 \\ \max\left\{1, \dfrac{2\left(\sum_{i=1}^{k} (T_i(u)-1)\right)^2}{k\sum_{i=1}^{k} (T_i(u)-1)(T_i(u)-2)}\right\}, & \text{else.} \end{cases} \tag{4.4}$$

This estimator relies on the first and second moments of $T_{(1,\theta,1)}$, since $\frac{\mathbb{E}(T^2_{(1,\theta,1)})}{\mathbb{E}(T_{(1,\theta,1)})^2} = \frac{2}{\theta}$, and includes a bias correction. A key advantage is that, apart from the choice of the threshold u, no auxiliary parameters are required. Furthermore, it does not depend on a prior or simultaneous estimation of $p(u)$.

(K-)gaps, censored and truncated estimators

More recently, several maximum likelihood based estimators have been proposed which also rely on IETs. They share the challenge of handeling the discrepancy between the positive IETs and the positive probability mass at zero of the mixture distribution. The gaps and K-gaps estimator introduced by Süveges (2007) and

Süveges and Davidson (2010), respectively, treats very short IETs as zero in order to better distinguish intra- and inter-cluster times. The censored and truncated estimators proposed by Holešovský and Fusek (2020) and Holešovský and Fusek (2022) apply censoring or truncation techniques to the IET sample to make maximum likelihood estimation feasible. All three approaches require the local dependence condition $D^{(k)}(u_n)$ to hold and involve an auxiliary parameter choice by the statistican. As stated in Holešovský and Fusek (2020), the $D^{(k)}(u_n)$ condition says, that within a cluster an exceedance is almost always followed by another exceedance within $k-1$ consecutive observations.

4.1.3 Fractional Poisson process

If the CTRE leads to a FPP, the IETs asymptotically follow a Mittag-Leffler distribution with tail parameter β and scale parameter σ_u. There are at least three methods for jointly estimating these parameters: The fractional moment estimator proposed by Kozubowski (2001), the log-moment estimator proposed by Cahoy et al. (2010) and the maximum likelihood estimator described in Gill and Straka (2017). The first two are based on the method of moments. However, since the Mittag-Leffler distribution has no finite moments, the fractional moment estimator uses moments of order $p < \beta$. Consequently, a constant p must be selected in advance and chosen to be sufficiently small. The log-moment estimator, in contrast, uses log-transformed data and does not require an auxiliary parameter. Additionally, in simulation studies Cahoy (2013) showed that the log-moment estimator outperforms the fractional moment estimator. Conversely, the maximum likelihood estimator does not rely on moment-based techniques but requires numerical optimisation, as the probability density function of the Mittag-Leffler distribution is not available in closed form. As a result, this method slightly outperforms the log-moment estimator, but is computationally more expensive (Hees et al., 2021).

4.2 Challenges

As seen in the previous section, even within the submodels, standard methods cannot be directly applied to estimate the parameters β, θ and σ_u, respectively. Although the extremal index has been investigated for many years, recently still more approaches have been proposed to improve estimation techniques.

Two properties of the distribution $\mathbb{P}_{(\beta,\theta,\sigma_u)}$, and of the IET sample $T_1(u), \ldots, T_k(u)$ make the search for suitable estimators particularly challenging:

1. The sample is only asymptotically distributed like $T_{(\beta,\theta,\sigma_u)}$. In particular, the discrepancy at zero poses difficulties since $T_i(u) > 0$ almost surely for all $i = 1, \ldots, k$, whereas $\mathbb{P}(T_{(\beta,\theta,\sigma_u)} = 0) = 1 - \theta$.

2. For $\beta < 1$, the distribution of $T_{(\beta,\theta,\sigma_u)}$ does not have finite moments.

In a first attempt to find a suitable estimation method, we adapted the interval estimator by Ferro and Segers (2003) by incorporating fractional moments. For $q < \beta$, it holds that

$$m_q = \mathbb{E}\left(\left(b\left(1/p(u)\right) T_{(\beta,\theta,1)}\right)^q\right) = \theta \frac{q\pi}{\beta} \frac{\left(\theta^{-1/\beta} b(1/p(u))\right)^q}{\Gamma(1-q)\sin\left(\frac{q\pi}{\beta}\right)} \quad \text{for } q < \beta, \tag{4.5}$$

and for $2q < \beta$ we obtain

$$\frac{m_q^2}{m_{2q}} = \theta \, \frac{q\pi}{2\beta} \frac{\Gamma(1-2q)\sin\left(\frac{2q\pi}{\beta}\right)}{\Gamma(1-q)^2 \sin\left(\frac{q\pi}{\beta}\right)^2} \longrightarrow \theta \quad \text{as } q \to 0. \tag{4.6}$$

Unfortunately, both theoretical calculations and simulations indicated that

$$\frac{\hat{m}_q^2}{\hat{m}_{2q}} \longrightarrow 1 \quad \text{as } q \to 0, \tag{4.7}$$

where $\hat{m}_q = \frac{1}{k} \sum_{i=1}^{k} T_i(u)^q$ denotes the empirical fractional moment of order q. This implies that the fractional moments of the sample do not converge to the fractional moments of the asymptotic distribution.

Conversely, if we want to apply maximum likelihood methods, similar to the gaps, K-gaps, censored and truncated estimators of the CPP, the data must first be pre-processed. This prevents all IETs, including the small ones, from being solely attributed to the Mittag-Leffler component. As a result, at least one auxiliary parameter is required, along with additional constraints like the $D^{(k)}(u_n)$ condition.

However, rather than relying on variants of the method of moments or maximum likelihood estimation, we introduce a minimum distance approach in the following section.

4.3 Minimum distance estimation with Cramér-von Mises criterion

The minimum distance approach has been introduced by Wolfowitz (1957) and explored in many further works, see, e.g., Drossos and Philippou (1980) or Parr (1981). The main idea is to measure the "similarity" of the sample data with a parametric model, minimising a distance measure between the probability density function or the cumulative distribution function of the parametric model and a

non-parametric density estimate or the empirical distribution function of the sample data. Here, we use distances based on distribution functions.

Definition 1. *Let $Z_1, \ldots, Z_n$ be random variables and Z another random variable with c.d.f. F_ϑ, $\vartheta \in \Theta \subset \mathbb{R}^p$, $p \geqslant 1$, F_n the empirical c.d.f. corresponding to $Z_1, \ldots, Z_n$, and $\Delta(\cdot, \cdot) \geqslant 0$ a function quantifying the distance between two c.d.f.s. If there is a $\hat{\vartheta} \in \Theta$ such that*

$$\Delta(F_n, F_{\hat{\vartheta}}) = \inf_{\vartheta \in \Theta} \Delta(F_n, F_\vartheta), \tag{4.8}$$

then $\hat{\vartheta}$ is called a minimum distance estimate *of ϑ. $\Delta(\cdot, \cdot)$ is called* criterion function.

Thus, out of the distribution family $\{F_\vartheta \mid \vartheta \in \Theta \subset \mathbb{R}^p\}$, the distribution $F_{\hat{\vartheta}}$ is the best distribution that fits to the empirical cumulative distribution function (e.c.d.f.) F_n regarding the criterion function Δ.

The Cramér-von-Mises distance, first proposed by Cramér (1928) and von Mises (1931), between two c.d.f.s G and H is defined as

$$\Delta^{[\mathrm{CM}]}(G, H) = \int_{-\infty}^{\infty} (G(x) - H(x))^2 \, \mathrm{d}H(x) = \mathbb{E}\left((G(X) - H(X))^2\right), \tag{4.9}$$

where X is a random variable with distribution function H. Therefore $\Delta^{[\mathrm{CM}]}(G, H) \in [0, 1]$. If H is absolutely continuous and $G = F_n$ an empirical c.d.f. of observations $z_{(1)} \leqslant \ldots \leqslant z_{(n)}$, then

$$\Delta^{[\mathrm{CM}]}(F_n, H) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{i - \frac{1}{2}}{n} - H(z_{(i)}) \right)^2 + \frac{1}{12n^2}. \tag{4.10}$$

4.3.1 Applying the original Crámer-von Mises estimator

Let $F^{(\beta,\theta,\sigma_u)}$ be the c.d.f. of the mixture distribution $(1-\theta)\,\delta_0 + \theta\,\mathrm{ML}(\beta, \theta^{-1/\beta}\sigma_u)$, $F_*^{(\beta,\theta,\sigma_u)}$ is the c.d.f. and $f_*^{(\beta,\theta,\sigma_u)}$ is the density of $\mathrm{ML}(\beta, \theta^{-1/\beta}\sigma_u)$, then

$$F^{(\beta,\theta,\sigma_u)}(x) = (1-\theta) + \theta \cdot F_*^{(\beta,\theta,\sigma_u)}(x), \quad \text{for } \; x \geqslant 0. \tag{4.11}$$

The Cramér-von Mises distance between $F^{(\beta,\theta,\sigma_u)}$ and the empirical c.d.f. F_k of IETs $T_1(u), \ldots, T_k(u)$ is

$$\begin{aligned} \Delta^{[\mathrm{CM}]}(F_k, F^{(\beta,\theta,\sigma_u)}) &= (1-\theta)^3 + \theta \int_0^\infty (F_k(x) - F^{(\beta,\theta,\sigma_u)}(x))^2 \, \mathrm{d}F_*^{(\beta,\theta,\sigma_u)}(x) \\ &= (1-\theta)^3 + \theta \int_0^\infty (F_k(x) - F^{(\beta,\theta,\sigma_u)}(x))^2 f_*^{(\beta,\theta,\sigma_u)}(x) \, \mathrm{d}x. \end{aligned} \tag{4.12}$$

The smaller the value of θ, the less influence the data have on the distance $\Delta^{[\mathrm{CM}]}$. Irrespective of the underlying true parameter values, it holds that $\Delta^{[\mathrm{CM}]}(F_k, F_{\beta,\theta,\sigma_{p(u)}}) > (1-\theta)^3$ and $\lim_{\theta\to 0} \Delta^{[\mathrm{CM}]}(F_k, F_{\beta,\theta,\sigma_{p(u)}}) = 1$. This leads to a high bias, esspecially if θ is low, which is illustrated by the heatmaps in Figure 4.1.

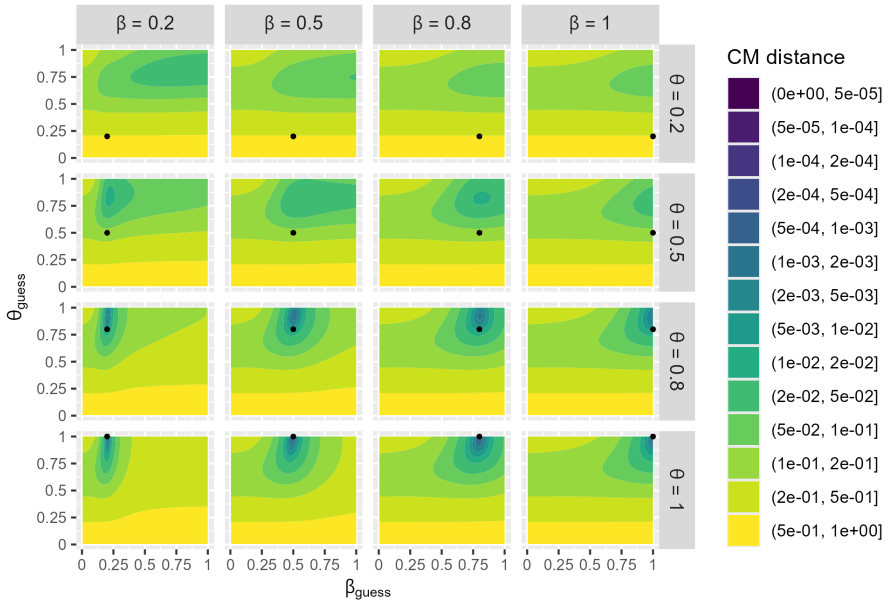

Figure 4.1: CM distances between a sample with true parameter values β, θ and $\sigma_u = p(u)^{-1/\beta}$ and $T_{(\beta_{\text{guess}},\theta_{\text{guess}},\sigma_u)}$. The black dots show where the distance should be minimal.

4.3.2 Modifications of the Crámer-von Mises criterion

We consider two modifications of the criterion function to improve the estimation procedure.

CMmod1 distance

The first modification is quite intuitive, since we simply use the continuous part of the integrator of the original distance (4.12):

$$\Delta^{[\mathrm{CMmod1}]}(F_k, F^{(\beta,\theta,\sigma_u)}) = \int_0^\infty (F_k(x) - F^{(\beta,\theta,\sigma_u)}(x))^2 f_*^{(\beta,\theta,\sigma_u)}(x)\,\mathrm{d}x. \tag{4.13}$$

With some cumbersome but straightfoward calculations (see Appendix B.1), we obtain the sum representation

$$\begin{aligned}\Delta^{[\mathrm{CMmod1}]}(F_k, F^{(\beta,\theta,\sigma_u)}) &= \frac{1}{\theta k}\sum_{i=1}^{k}\left(\frac{i-\frac{1}{2}}{k} - F^{(\beta,\theta,\sigma_u)}(T_{(i)}(u))\right)^2 \\ &\quad + \frac{1}{12\theta k^2} - \frac{1}{3}\frac{(1-\theta)^3}{\theta},\end{aligned} \tag{4.14}$$

where $T_{(1)}(u) \leqslant \ldots \leqslant T_{(k)}(u)$ are the ordered IETs. The sum representation is particularly advantageous for calculations and optimisation. It holds for all $k \in \mathbb{N}$ that

$$\lim_{\theta\to 0} \Delta^{[\mathrm{CMmod1}]}(F_k, F^{(\beta,\theta,\sigma_u)}) = 0, \tag{4.15}$$

see Appendix B.2 for calculations. Therefore, it is essential that we restrict the parameter space of θ to a compact set $[a, 1]$, $a > 0$. The heatmaps in Figure 4.2 show that the CMmod1 distance performs much better than the original CM distance but also confirms (4.15).

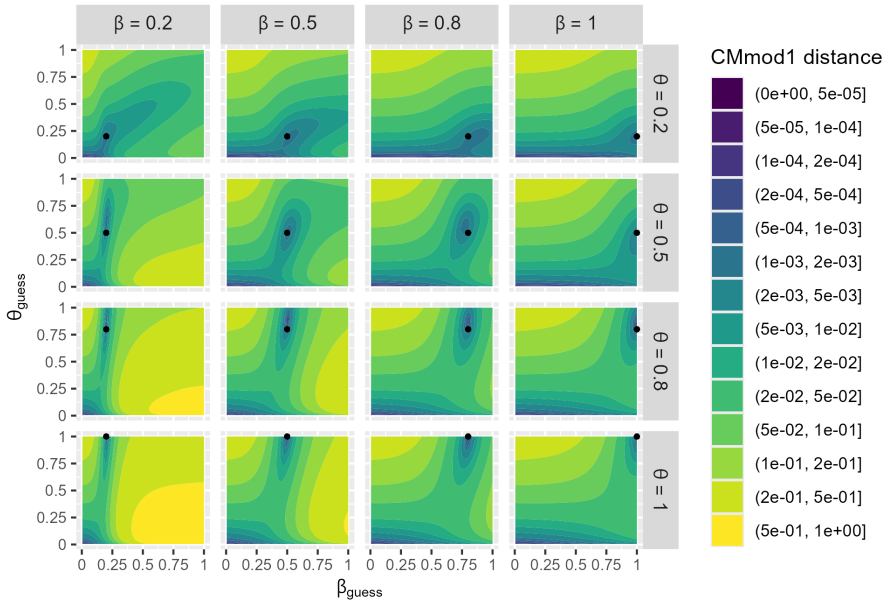

Figure 4.2: CMmod1 distances between a sample with true parameter values β, θ and $\sigma_u = p(u)^{-1/\beta}$ and $T_{(\beta_{\mathrm{guess}},\theta_{\mathrm{guess}},\sigma_u)}$. The black dots show where the distance should be minimal.

CMmod2 distance

The second modified criterion function is

$$\begin{aligned}
&\Delta^{[\text{CMmod2}]}(F_k, F^{(\beta,\theta,\sigma_u)}) \\
&\quad= \frac{1}{\theta^2}\int_0^\infty (\max\{F_k(x), 1-\theta\} - F^{(\beta,\theta,\sigma_u)}(x))^2 f_*^{(\beta,\theta,\sigma_u)}(x)\,\mathrm{d}x \\
&\quad= \int_0^\infty \left(\max\left\{\frac{F_k(x)-(1-\theta)}{\theta}, 0\right\} - F_*^{(\beta,\theta,\sigma_u)}(x)\right)^2 f_*^{(\beta,\theta,\sigma_u)}(x)\,\mathrm{d}x \\
&\quad= \Delta^{[\text{CM}]}(F_{k,\theta}^{\max}, F_*^{(\beta,\theta,\sigma_u)}),
\end{aligned} \tag{4.16}$$

and adresses the discrepancy of the two distribution functions at zero:

$$|F_k(0) - F^{(\beta,\theta,\sigma_u)}(0)| = 1-\theta. \tag{4.17}$$

Therefore, we consider $\max\{F_k(x), 1-\theta\}$ instead of $F_k(x)$, such that

- $|\max\{F_k(0), 1-\theta\} - F^{(\beta,\theta,\sigma_u)}(0)| = 0$,
- $|\max\{F_k(x), 1-\theta\} - F^{(\beta,\theta,\sigma_u)}(x)| \leqslant |F_k(x) - F^{(\beta,\theta,\sigma_u)}(x)|$ for all $x \geqslant 0$,
- $|\max\{F_k(x), 1-\theta\} - F^{(\beta,\theta,\sigma_u)}(x)| \leqslant \theta$ for all $x \geqslant 0$.

Because of the last inequality, we add the scaling $1/\theta^2$ such that $\Delta^{[\text{CMmod2}]}(F_k, F^{(\beta,\theta,\sigma_u)}) \in [0,1]$. Note that the $\Delta^{[\text{CMmod2}]}(F_k, F^{(\beta,\theta,\sigma_u)})$ corresponds to the standard Cramér-von Mises distance between the distribution functions $F_{k,\theta}^{\max}(\cdot) = \max\left\{\frac{F_k(\cdot)-(1-\theta)}{\theta}, 0\right\}$ and $F_*^{(\beta,\theta,\sigma_u)}$.

As before, with some cumbersome but straightfoward calculations (see Appendix B.1), we receive

$$\begin{aligned}
\Delta^{[\text{CMmod2}]}(F_k, F^{(\beta,\theta,\sigma_u)}) \;&= \frac{1}{\theta^3 k}\sum_{i=l+1}^{k}\left(\frac{i-\frac{1}{2}}{k} - F^{(\beta,\theta,\sigma_u)}(T_{(i)}(u))\right)^2 \\
&\quad + \frac{k-l}{12\theta^3k^3} - \frac{(k(1-\theta))^3 - l^3}{3k^3\theta^3} \\
&\quad + \frac{(k(1-\theta))^2 - l^2}{k^2\theta^3} F^{(\beta,\theta,\sigma_u)}(T_{(l)}(u)) \\
&\quad - \frac{(k(1-\theta)) - l}{k\theta^3} F^{(\beta,\theta,\sigma_u)}(T_{(l)}(u))^2,
\end{aligned} \tag{4.18}$$

where $T_{(1)}(u) \leqslant \ldots \leqslant T_{(k)}(u)$ are the ordered IETs and $l := \lceil k(1-\theta)\rceil$, with $\lceil\cdot\rceil$ being the ceiling function.

In the case of $l = k$, which is the case if $\theta < \frac{1}{k}$, the distance function can be abbreviated in the following form (see Appendix B.2 for calculations). It then only

depends on the largest IET $T_{(k)}(u)$:

$$\begin{aligned}\Delta^{[\text{CMmod2}]}\left(F_k, F^{(\beta,\theta,\sigma_u)}\right) &\overset{\theta<\frac{1}{k}}{=} \frac{1}{3} - F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right) + F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right)^2 \\ &\to \frac{1}{3} \qquad \text{as} \quad \theta \to 0.\end{aligned} \tag{4.19}$$

This illustrates that, esspecially for small θ, the sample size, i.e., the number of exceedances k, must be large enough, meaning for $\theta \in [a, 1]$, $a > 0$, $k > \frac{1}{a}$. The heatmaps of the CMmod2 distance in Figure 4.3 show the usefulness of this modification.

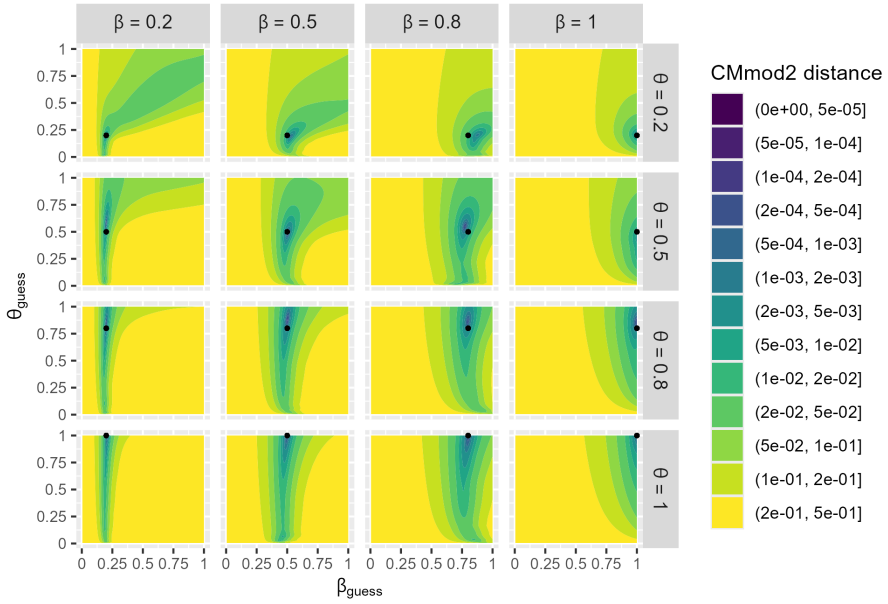

Figure 4.3: CMmod2 distances between a sample with true parameter values β, θ and $\sigma_u = p(u)^{-1/\beta}$ and $T_{(\beta_{\text{guess}},\theta_{\text{guess}},\sigma_u)}$. The black dots show where the distance should be minimal.

CMmod1 and CMmod2 estimator

Finally, we define the CMmod1 and CMmod2 estimators: For F_k, the e.c.d.f. of $T_1(u), \ldots, T_k(u)$, the CMmod1 estimator is defined as

$$\widehat{(\beta, \theta, \sigma_u)}(F_k) = \underset{(\beta,\theta,\sigma_u)\in[a,1]^2\times(0,\infty)}{\arg\inf} \Delta^{[\text{CMmod1}]}\left(F_k, F^{(\beta,\theta,\sigma_u)}\right) \tag{4.20}$$

and the CMmod2 estimator is

$$\widetilde{(\beta,\theta,\sigma_u)}(F_k) = \underset{(\beta,\theta,\sigma_u)\in[a,1]^2\times(0,\infty)}{\arg\inf} \Delta^{[\text{CMmod2}]}(F_k, F^{(\beta,\theta,\sigma_u)}), \tag{4.21}$$

respectively. If we seperate $\sigma_u = p(u)^{-1/\beta}L(1/p(u))$ in two parts and estimate $p(u)$ in advance as described at the beginning of this section, we can refer to

$$\widehat{(\beta,\theta,l_u)}(F_k) = \underset{(\beta,\theta,l_u)\in[a,1]^2\times(0,\infty)}{\arg\inf} \Delta^{[\text{CMmod1}]}(F_k, F^{(\beta,\theta,\hat{p}_n(u)^{-1/\beta}l_u)}) \tag{4.22}$$

and

$$\widetilde{(\beta,\theta,l_u)}(F_k) = \underset{(\beta,\theta,l_u)\in[a,1]^2\times(0,\infty)}{\arg\inf} \Delta^{[\text{CMmod2}]}(F_k, F^{(\beta,\theta,\hat{p}_n(u)^{-1/\beta}l_u)}) \tag{4.23}$$

as CMmod1 and CMmod2 estimators, respectively.

4.4 Conclusions and discussion

In this chapter, we addressed the challenge of estimating the parameters β, θ and σ_u of the asymptotic mixture distribution of IETs under the CTRE model. After reviewing established estimation methods for the PP, CPP and FPP submodels, and identifying the main difficulties in extending these approaches to the general FCPP model, we developed two modifications of the Cramér-von Mises estimator to fit the model parameters. In the following two chapters, we will analyse these methods both theoretically, in terms of weak consistency, and empirically, through simulation studies.

Alternative estimators, e.g., minimum distance estimators with a different criterion function than Cramér-von Mises oder estimators based on quantiles, on the (K-)gaps or on the censored and truncated estimators, have not yet been considered. These approaches represent promising directions for future research and comparison.

One important aspect not addressed in this chapter is the choice of the threshold u. Throughout this chapter and the thesis as a whole, u has been treated as fixed. However, the selection of an appropriate threshold is a central and challenging issue in extreme value theory. On the one hand, the threshold must be sufficiently high for the asymptotic approximation to be valid; on the other hand, it must not be so high that too few exceedances remain for reliable estimation. The literature offers a variety of methods for threshold selection, which could be incorporated and assessed in future work.

5 Theoretical results

This chapter presents the theoretical properties of the modified minimum distance estimator *CMmod1*. Similar results also apply to *CMmod2* and are presented in the Appendix to avoid repetitions and overlaps. The main results are the weak consistency of our estimation procedures under the simplification that the sample consists of stochastically independent IETs.

The chapter is organised as follows: In the first section, we summarise the well-known theoretical properties of the classical minimum distance method and briefly discuss why these cannot be easily transferred to our case. In the second and third sections, we work out the weak consistency of our estimation method step by step. For better readability, the corresponding proofs are provided in a fourth section. The chapter ends with a summary and discussion of the findings obtained and the implications of the restrictions made.

5.1 Known properties of the Cramér-von Mises estimators

Wolfowitz studied the minimum distance approach in various articles in the 1950s, e.g., Wolfowitz (1954) and Wolfowitz (1957). He established its strong consistency for several distance functions under the assumption of identifiable parameter values. From Wolfowitz (1957), we can derive that the classical Crámer-von Mises estimator is strongly consistent under general regularity conditions (see also Parr and Schucany, 1980; van der Vaart, 1998, Chapter 5). In the last decades, different modifications of the original Cramér-von Mises distance were investigated, for example the weighted Cramér-von Mises distance

$$\Delta_{\Psi_\phi}(G, F_\phi) = \int (G(x) - F_\phi(x))^2 \Psi_\phi(x) \mathrm{d}F_\phi(x), \tag{5.1}$$

where Ψ_ϕ is a non-negative weighting function (Parr and De Wet, 1981). Under fairly general conditions, the derived estimators are asymptotically normal and for appropriate weights also efficient.

Ötztürk and Hettmansperger (1997) investigated an even more general class of weighted Cramér-von Mises distance estimator by minimising

$$\Delta_{\Psi_\phi,H,\delta}(G, F_\phi) = \int H(\delta(x; G, F_\phi))\Psi_\phi(x)\mathrm{d}F_\phi(x), \tag{5.2}$$

where the function $\delta(x; G, F_\phi) = G(x) - F_\phi(x)$ quantifies the discrepancy between G and the distribution family $\Gamma = \{F_\phi \mid \phi \in \Phi\}$. The function H is nonnegative and continuous, satisying $H(0) = 0$, $H'(0) = 0$ and $H''(0) > 0$. They showed that, for suitable choices of H and Ψ_ϕ, the corresponding estimator is Fisher consistent, asymptotically normal and nearly efficient with desirable robustness properties. Choosing $H(x) = x^2$ gives the weighted Cramér-von Mises distance of (5.1).

Unfortunately, the consistency results from the literature cannot be directly transferred to our estimation setting, as they assume a fixed underlying distribution $G^{(n)} = G$ that does not change with increasing n. In these classical scenarios, either G belongs to the model class $\Gamma = \{F_\phi \mid \phi \in \Phi\}$, in which case consistency holds under regularity conditions, or G is not in Γ, implying a model misspecification. In the latter case, under the similar conditions, the estimator converges to the parameter value ϕ that best approximates G. In our setting, however, we face a triangular array structure where $G^{(n)}$ converges to F_ϕ, which is not covered by either of the classical cases. In particular, the standard Cramér-von Mises estimator seems not to be consistent in our situation (see Subsection 4.3.1).

5.2 Weak consistency of CMmod1 for a mixture distribution

In this and the next section, we work out the weak consistency for our estimation method. To do so, we first generalise our mixture distribution by allowing an arbitrary positive valued continuous distribution with parameters in a compact space that is continuous with respect to its parameters. The other distribution remains the Dirac measure at point zero. More precisely, let $T_{(\phi,\theta)}$ be a random variable on the non-negative real numbers with the mixture distribution

$$\mathbb{P}_{(\phi,\theta)} = (1 - \theta) \cdot \delta_0 + \theta \cdot \mu_\phi, \tag{5.3}$$

with weighting parameter $\theta \in [a, 1]$, $a > 0$, μ_ϕ a positive valued continuous distribution with parameter vector $\phi \in \Phi \subset \mathbb{R}^m$, Φ compact, and δ_0 the Dirac measure at point zero. It is allowed that the weighting parameter θ is also part of the parameter vector ϕ. $F^{(\phi,\theta)}$ denotes the c.d.f. of $\mathbb{P}_{(\phi,\theta)}$. $F_*^{(\phi)}$ and $f_*^{(\phi)}$ denote the c.d.f. and density function of μ_ϕ, respectively, that are continous with respect to their parameters.

The modified Cramér-von Mises distance between $F^{(\phi,\theta)}$ and another c.d.f. G of a positive valued distribution is now defined by

$$\Delta(G, F^{(\phi,\theta)}) \quad = \quad \int_0^\infty \left(G(x) - F^{(\phi,\theta)}(x)\right)^2 \cdot f_*^{(\phi)}(x)\,\mathrm{d}x \tag{5.4}$$

The minimum distance approximation of a c.d.f. G is then defined as

$$\widehat{(\phi,\theta)}(G) = \underset{(\phi,\theta)\in\Phi\times[a,1]}{\arg\min}\ \Delta(G, F^{(\phi,\theta)}) \tag{5.5}$$

and the CMmod1 estimator is $(\hat{\phi}_k, \hat{\theta}_k) := \widehat{(\phi,\theta)}(G_k)$, where G_k is the e.c.d.f. of G.

5.2.1 Known scale parameter

Let $T^{(n)}$ be a random variable on the positive real numbers with c.d.f. $F^{(n)}$ and

$$\frac{T^{(n)}}{\sigma_n} \xrightarrow{\mathrm{d}} T_{(\phi,\theta)}, \tag{5.6}$$

where $\sigma_n > 0$ is a known scale parameter. Without loss of generality, we set $\sigma_n = 1$ for all $n \in \mathbb{N}$. Let $T_{1,n}, \ldots, T_{k_n,n}$ be k_n i.i.d. copies of $T^{(n)}$ where $k_n \to \infty$, and F_{k_n} is the e.c.d.f., respectively. In addition, we require the following condition for the sequence $(k_n)_{n\in\mathbb{N}}$:

Condition 3. *Let $(k_n)_{n\in\mathbb{N}}$ be a sequence of positive integers with $k_n \to \infty$ such that*

$$\textit{for all } \varepsilon > 0: \quad \sum_{n=1}^{\infty} 2 \cdot \exp(-2k_n\varepsilon^2) < \infty. \tag{5.7}$$

The condition prevents the sample size k_n from growing too slowly. For example, this condition is fulfilled by $k_n = n$ or $k_n = n^{1/\alpha}, \alpha > 1$. The latter fulfills $k_n = o(n)$. In contrast, the condition is not met by $k_n = \log(n)$.

The objective of this subsection is to prove that CMmod1 for the parameter vector (ϕ, θ) is weakly consistent. For this we show two other results first. The first is similar to the Glivenko-Cantelli theorem and is essential for justifying the utilisation of the e.c.d.f. of the sample to describe the c.d.f. of the asymptotic mixture distribution. It states that the e.c.d.f. F_{k_n} converges towards the asymptotic c.d.f. $F^{(\phi,\theta)}$. The second result shows that the distance between the fitted c.d.f. $F^{(\hat{\phi}_{k_n},\hat{\theta}_{k_n})}$ and the theoretical c.d.f. $F^{(n)}$, as well as the distance between the fitted c.d.f. $F^{(\hat{\phi}_{k_n},\hat{\theta}_{k_n})}$ and the asymptotic c.d.f. $F^{(\phi,\theta)}$ converge to zero when measured using the modified Cramér-von Mises distance (5.4).

Theorem 2. *Let Condition 3 be fulfilled for $(k_n)_{n\in\mathbb{N}}$. Then*

$$\|F_{k_n} - F^{(n)}\|_\infty = \sup_x |F_{k_n}(x) - F^{(n)}(x)| \to 0 \quad \textit{almost surely.} \tag{5.8}$$

For all continuity points x of $F^{(\phi,\theta)}$, i.e. $\forall x > 0$, it applies

$$F_{k_n}(x) \underset{n\to\infty}{\longrightarrow} F^{(\phi,\theta)}(x) \qquad \textit{almost surely.} \tag{5.9}$$

If Condition 3 is not fulfilled, the above expressions converge, but possibly only in probability.

It follows that, under Condition 3,

$$\Delta(F_{k_n}, F^{(\phi,\theta)}) \to 0 \qquad \text{almost surely.} \tag{5.10}$$

The condition in Theorem 2 that x must be a continuity point of $F^{(\phi,\theta)}$ is necessary: For $\theta < 1$, the only discontinuity point of $F^{(\phi,\theta)}$ is zero and there (5.30) does not apply:

$$\left|F_{k_n}(0) - F^{(\phi,\theta)}(0)\right| = \left|0 - (1-\theta)\right| = 1 - \theta \quad \forall n \in \mathbb{N}. \tag{5.11}$$

Theorem 3. *Let $F^{(n)}(x) \underset{n\to\infty}{\longrightarrow} F^{(\phi,\theta)}(x)$, $(\phi,\theta) \in \Phi \times [a,1]$, for all $x > 0$ and let $(\hat{\phi}_{k_n}, \hat{\theta}_{k_n}) = \widehat{(\phi,\theta)}(F_{k_n})$, where the sequence $(k_n)_{n\in\mathbb{N}}$ fulfills Condition 3.*

a)

$$\Delta\left(F^{(n)}, F^{(\hat{\phi}_{k_n}, \hat{\theta}_{k_n})}\right) \underset{n\to\infty}{\longrightarrow} 0 \quad \textit{almost surely.} \tag{5.12}$$

b)

$$\Delta\left(F^{(\phi,\theta)}, F^{(\hat{\phi}_{k_n}, \hat{\theta}_{k_n})}\right) \underset{n\to\infty}{\longrightarrow} 0 \quad \textit{in probability.} \tag{5.13}$$

Using these results, we can show that the *CMmod1* estimator for (ϕ,θ) is weakly consistent.

Theorem 4. *Under the same conditions as in Theorem 3, the weak consistency applies, i.e.*

$$\widehat{(\phi,\theta)}(F_{k_n}) \to (\phi,\theta) \quad \textit{in probability.} \tag{5.14}$$

5.2.2 Unknown scale parameter

We are now looking at the same situation as in (5.6), such that

$$\frac{T^{(n)}}{\sigma_n} \xrightarrow{d} T_{(\phi,\theta)}, \tag{5.15}$$

but assume that the scale parameter $\sigma_n > 0$ is unknown.

If the sequence $(\sigma_n)_{n\in\mathbb{N}}$ converges towards a constant $\sigma > 0$, σ can be seen as part of the parameter vector ϕ, since then

$$T^{(n)} = \sigma_n \frac{T^{(n)}}{\sigma_n} \xrightarrow{d} \sigma\, T_{(\phi,\theta)}, \tag{5.16}$$

which corresponds to the situation from the subsection above.

If instead $\sigma_n \to \infty$, we cannot apply the above theorems since the parameter space of σ_n cannot be restricted to a compact set.

Instead, we assume that there is another weakly consistent estimator $\hat{\sigma}_n$ for σ_n, more precisely

$$\lim_{n\to\infty} |\sigma_n - \hat{\sigma}_n| = 0 \ \text{ in probability.}$$

Then, we get

$$\frac{\hat{\sigma}_n}{\sigma_n} \longrightarrow 1 \ \text{ in probability,}$$

since $\left\{\left|\frac{\hat{\sigma}_n}{\sigma_n} - 1\right| > \varepsilon\right\} \subseteq \left\{\left|\frac{\hat{\sigma}_n}{\sigma_n} - 1\right| > \frac{\varepsilon}{\sigma_n}\right\} = \{|\sigma_n - \hat{\sigma}_n| > \varepsilon\}$ for a large enough n. Using Slutskys Theorem, we get

$$\frac{T^{(n)}}{\hat{\sigma}_n} = \frac{\sigma_n}{\hat{\sigma}_n} \cdot \frac{T^{(n)}}{\sigma_n} \xrightarrow{\mathrm{d}} T_{(\phi,\theta)}. \tag{5.17}$$

Note that $T_{1,n}/\hat{\sigma}_n, \ldots, T_{k_n,n}/\hat{\sigma}_n$ may be dependent if $\hat{\sigma}_n$ uses the same sample, but the estimator $\widehat{(\phi,\theta)}(\hat{F}_{k_n})$ remains weakly consistent, where $\hat{F}_{k_n}$ is the e.c.d.f. of $T_{1,n}/\hat{\sigma}_n, \ldots, T_{k_n,n}/\hat{\sigma}_n$ if we can ensure that $\hat{F}_{k_n}(x) = F_{k_n}(x\hat{\sigma}_n) \to F^{(\phi,\theta)}(x)$ in probability for all continuity points x of $F^{(\phi,\theta)}$.

A sufficient but possibly not necessary condition would be that $T_{1,n}, \ldots, T_{k_n,n}$ are stochastically independent of $\hat{\sigma}_n$. We can accomplish this by estimating $\hat{\sigma}_n$ with a different, stochastically independent sample.

5.3 Application to IETs

We are now back in the situation from Chapter 3 and apply the results of the previous section to show that the introduced estimation method of Chapter 4 is in many situations weakly consistent regarding our model

$$\frac{T(u)}{b(1/p(u))} \xrightarrow{d} T_{(\beta,\theta,1)}. \tag{5.18}$$

For this, we restrict the parameter space of (β, θ) to a compact space $\Phi = [a, 1]^2$ with $a > 0$. If the scale parameter $\sigma_u = b(1/p(u))$ is known, we apply Theorem 4 directly. In general, however, we do not know the scale parameter, but we have further information about its structure:

$$\sigma_u = b(1/p(u)) = p(u)^{-1/\beta} \cdot L(1/p(u)). \tag{5.19}$$

Therefore, we first search for a consistent estimator for $p(u) = \mathbb{P}(X_1 > u)$ and afterwards we consider the slowly varying function L.

5.3.1 Estimation of $p(u)$

With the mixing condition $\Delta(u_n)$ introduced in Chapter 3 (see Condition 2), we can show that for the relative number of exceedances

$$\hat{p}_n(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i > u\} \tag{5.20}$$

it holds

$$\frac{p(u_{r_n})}{\hat{p}_n(u_{r_n})} \to 1 \quad \text{in probability,}$$

where r_n is a sequence of positive integers with $r_n = o(n)$.

Note that if a sequence (u_n) fulfills the $\Delta(u_n)$ condition, it also fulfills the $D(u_n)$ condition.

Theorem 5. *Let $(X_n)_{n\in\mathbb{N}}$ be stationary with extremal index $\theta \in (0, 1]$. Let there be a sequence of thresholds $(u_n)_{n\in\mathbb{N}}$ for which $\Delta(u_n)$ holds with $np(u_n) \to \nu \in (0, \infty)$. Let $(r_n)_{n\in\mathbb{N}}$ be a sequence of positive integers with $r_n = o(n)$. Then it holds*

$$\frac{p(u_{r_n})}{\hat{p}_n(u_{r_n})} \to 1 \quad \textit{in probability,} \tag{5.21}$$

where $p(u) = \mathbb{P}(X_1 > u)$ *and* $\hat{p}_n(u) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i > u\}$ *is the relative frequency of exceedances.*

Theorem 5 requires on the one hand that u must grow at a sufficiently fast rate so that $\hat{p}_n(u) \to 0$. On the other hand, u must grow at a sufficiently slow rate so that the expected number of exceedances grows towards infinity, since

$$\mathbb{E}(n\,\hat{p}_n(u_{r_n})) = \frac{n}{r_n} r_n p(u_{r_n}) \sim \frac{n}{r_n}\,\nu \to \infty. \tag{5.22}$$

This means that the expected number of exceedances $k_n = \mathbb{E}\left(\sum_{i=1}^{n} \mathbb{1}\{X_i > u\}\right)$ must not grow too quickly on the one hand, more precisely $k_n = o(n)$, and on the other hand it must not grow too slowly due to Condition 3, so that the results described above apply.

5.3.2 Convergence of L towards $l > 0$

Now, we take care of the slowly varying function L, the second part of the unknown scale parameter $\sigma_u = p(u)^{-1/\beta} \cdot L(1/p(u))$. It is well known and easy to show that every function that converges towards a positive value is slowly varying, but there are also slowly varying functions that do not converge, e.g., the logarithm function. In this subsection we require that $L(x) \to l$, $0 < l < \infty$, as $x \to \infty$. Then,

$$\frac{T(u)}{\hat{p}_n(u)^{-1/\beta}} = \frac{\sigma_u}{\hat{p}_n(u)^{-1/\beta}} \frac{T(u)}{\sigma_u} = \left(\frac{p(u)}{\hat{p}_n(u)}\right)^{-1/\beta} L(1/p(u)) \frac{T(u)}{\sigma_u} \xrightarrow{d} T_{(\beta,\theta,l)} \tag{5.23}$$

applies where the distribution of $T_{(\beta,\theta,l)}$ is $(1-\theta)\,\delta_0 + \theta\,\mathrm{ML}(\beta, \theta^{-1/\beta} l)$ so that

$$T(u) \overset{d}{\approx} (1-\theta)\,\delta_0 + \theta\,\mathrm{ML}(\beta, \theta^{-1/\beta} l\,\hat{p}_n(u)^{-1/\beta}). \tag{5.24}$$

By restricting the parameter space of μ to a compact space, we can now again apply Theorem 4, such that $\widehat{(\beta,\theta,l)}(F_{k_n})$ is a consistent estimator for (β, θ, l).

5.3.3 Tail parameter $\beta = 1$

In Chapter 3, we proved that if the waiting times W_n have finite mean, i.e. $\beta = 1$, the interexceedance time $T(u)$ is approximately mixture distributed with the Dirac

measure at point zero as one part and the exponential distribution as second part. More precisely,

$$\frac{T(u)}{\sigma_u} \xrightarrow{d} T_{(1,\theta,1)}, \tag{5.25}$$

with $\sigma_u = p(u)^{-1}\mathbb{E}(W_1)$, i.e., refering to (5.19), $L(1/p(u)) = \mathbb{E}(W_1) =: l$ is constant.

Using all the results of the previous subsections and assuming $\beta = 1$ is known, $\widehat{(\theta, l)}(F_{k_n})$ is a weakly consistent estimator of (θ, l).

Alternatively, we can estimate $\mathbb{E}(W_1)$ consistently in advance by $\bar{W} = \frac{1}{n}\sum_{i=1}^{n} W_i$. Under the same condition as in Theorem 5, $\hat{\sigma}_{u_{r_n}} = \bar{W}/\hat{p}_n(u_{r_n})$ is a consistent estimator for $\sigma_{u_{r_n}} = \mathbb{E}(W_1)/p(u_{r_n})$.

In this case, $\widehat{\theta}(F_{k_n})$ is a consistent estimator for the extremal index θ.

5.3.4 Tail parameter $\beta < 1$

If the tail parameter β is unknown, we lack specific information about the slowly varying function L. In the following, however, we consider three heavy-tailed distributions for the waiting times $(W_n)_{n\in\mathbb{N}}$ in the domain of attraction of D_β, a β-stable distribution with Laplace transform $\mathbb{E}\left(\exp(-sD_\beta)\right) = \exp(-s^\beta)$, i.e.

$$\frac{W_1 + \cdots + W_n}{b(n)} \xrightarrow{d} D_\beta. \tag{5.26}$$

For these three examples, the slowly varying function L is constant, i.e. convergent in particular, so we can apply the results of Subsection 5.3.2.

Waiting times with a stable distribution

Let $W_i/\varsigma \overset{d}{=} D_\beta$ be i.i.d. for $i = 1, \ldots, n$ with $\varsigma > 0$. Using the parameterisation of Samorodnitsky and Taqqu (1994), D_β is $S_\beta(\cos(\pi\beta/2)^{1/\beta}, 1, 0)$ distributed (see Subsection 2.1.2). Since

$$\frac{W_1 + \cdots + W_n}{n^{1/\beta}\,\varsigma} \overset{d}{=} D_\beta$$

(Samorodnitsky and Taqqu, 1994), $b(n) = n^{1/\beta}\,\varsigma$ and thus $L(n) = \varsigma$ applies.

Waiting times with a Mittag-Leffler distribution

Let $W_i \overset{d}{=} \mathrm{ML}(\beta, \varsigma)$ be i.i.d. for $i = 1, \ldots, n$. Then, $W_1/\varsigma \overset{d}{=} \mathrm{ML}(\beta, 1)$ with Laplace transform $\mathbb{E}\left(\exp(-s\, W_1/\varsigma)\right) = \exp\left(\frac{1}{1+s^\beta}\right)$. As stated in Subsection 2.1.3, the Laplace transform of $\frac{W_1+\cdots+W_n}{\varsigma n^{1/\beta}}$ converges towards the Laplace transform of D_β and therefore

$$\frac{W_1 + \cdots + W_n}{n^{1/\beta}\,\varsigma} \xrightarrow{d} D_\beta, \tag{5.27}$$

with $b(n) = n^{1/\beta}\,\varsigma$ and thus $L(n) = \varsigma$.

Waiting times with a Pareto distribution

Let $W_1, \ldots, W_n$ be i.i.d. with $\mathbb{P}(W_i > t) = Ct^{-\beta}$ where $C = \left(\varsigma^\beta \Gamma(1-\beta)\right)^{-1}$ and $t \geqslant \left(\varsigma\Gamma(1-\beta)^{1/\beta}\right)^{-1}$. With Meerschaert and Sikorskii (2012, Prop. 3.10 & Theorem 3.37) we can show that

$$\frac{W_1 + \cdots + W_n}{n^{1/\beta}\,\varsigma} \xrightarrow{d} D_\beta. \tag{5.28}$$

Therefore, $b(n) = n^{1/\beta}\,\varsigma$ and thus $L(n) = \varsigma$ applies.

5.4 Proofs

Proof of Theorem 2:

Theorem 2. *Let Condition 3 be fulfilled for $(k_n)_{n\in\mathbb{N}}$. Then*

$$\|F_{k_n} - F^{(n)}\|_\infty = \sup_x |F_{k_n}(x) - F^{(n)}(x)| \to 0 \quad \textit{almost surely.} \tag{5.29}$$

For all continuity points x of $F^{(\phi,\theta)}$, i.e. $\forall x > 0$, it applies

$$F_{k_n}(x) \underset{n\to\infty}{\longrightarrow} F^{(\phi,\theta)}(x) \qquad \textit{almost surely.} \tag{5.30}$$

If Condition 3 is not fulfilled, the above expressions converge, but possibly only in probability.

Proof. We use the Dvoretzky-Kiefer-Wolfowitz inequality saying that

$$\forall n \in \mathbb{N}, \quad \forall \varepsilon > 0: \quad P\left(\sup_x |G_n(x) - G(x)| > \varepsilon\right) \leqslant 2 \cdot \exp(-2n\varepsilon^2) \tag{5.31}$$

where G_n denotes the empirical c.d.f. for a sample of n i.i.d. random variables with c.d.f. G (Massart, 1990).

Then, for all $x > 0$,

$$\forall \varepsilon > 0 \; \exists n_x \in \mathbb{N} : \; \forall n \geqslant n_x, \; |F^{(n)}(x) - F^{(\phi,\theta)}(x)| < \varepsilon \tag{5.32}$$

applies since $T^{(n)} \xrightarrow{d} T_{(\phi,\theta)}$.

Let $\varepsilon > 0$ and $x > 0$ be fixed and $\delta_{n,x} := \varepsilon - |F^{(n)}(x) - F^{(\phi,\theta)}(x)|$. Then, $\delta_{n,x} \to \varepsilon$ for $n \to \infty$ and $\delta_{n,x} > 0$ for all $n \geqslant n_x$. Finally, since $k_n \to \infty$ and $\delta_{n,x}^2 \to \varepsilon^2$ we get

$$\begin{aligned}
\forall \varepsilon > 0 \; \forall n > n_0 : \quad & P(|F_{k_n}^{(n)}(x) - F^{(\phi,\theta)}(x)| > \varepsilon) \\
& \leqslant P(|F_{k_n}^{(n)}(x) - F^{(n)}(x)| + |F^{(n)}(x) - F^{(\phi,\theta)}(x)| > \varepsilon) \\
& = P(|F_{k_n}^{(n)}(x) - F^{(n)}(x)| > \delta_{n,x}) \\
& \leqslant P(\sup_y |F_{k_n}^{(n)}(y) - F^{(n)}(y)| > \delta_{n,x}) \\
& \leqslant 2 \cdot \exp(-2k_n\delta_{n,x}^2) \quad \overset{n\to\infty}{\longrightarrow} \quad 0 \; .
\end{aligned} \tag{5.33}$$

If we restrict the sequence $(k_n)_{n\in\mathbb{N}}$ such that

$$\forall \varepsilon > 0 : \quad \sum_{n=1}^{\infty} 2 \cdot \exp(-2k_n\varepsilon^2) < \infty, \tag{5.34}$$

we get complete convergence and thus the convergence even holds almost surely. This is achieved e.g. by $k_n = n^{1/\alpha}$, $\alpha \geqslant 1$: The condition $k_n = o(n)$ is obvious for $\alpha > 1$ and for all $z \geqslant 0$ and $m \geqslant 0$ we have $\exp(z) \geqslant (1 + \frac{z}{m})^m \geqslant \left(\frac{z}{m}\right)^m$, ergo $\exp(-z) \leqslant \left(\frac{m}{z}\right)^m$. Choosing $m = 2 \cdot \alpha$, we get

$$\sum_{n=1}^{\infty} 2 \cdot \exp(-2n^{1/\alpha}\varepsilon^2) \leqslant 2 \cdot \sum_{n=1}^{\infty} \left(\frac{2\alpha}{2n^{1/\alpha}\varepsilon^2}\right)^{2\alpha} = 2 \cdot \left(\frac{\alpha}{\varepsilon^2}\right)^{2\alpha} \sum_{n=1}^{\infty} n^{-2} < \infty. \tag{5.35}$$

□

For the proof of Theorem 3, we introduce two minor results in Corollary 1. They follow from the theorem of dominated convergence and Theorem 2:

Corollary 1. *If the sequence $(k_n)_{n\in\mathbb{N}}$ fulfills Condition 3, then*

$$\int_0^{\infty} \left|F_{k_n}(x) - F^{(\phi,\theta)}(x)\right| \cdot f_*^{(\tilde{\phi})}(x) \, \mathrm{d}x \xrightarrow[n\to\infty]{} 0 \tag{5.36}$$

and thus also

$$\int_0^\infty \left(F_{k_n}(x) - F^{(\phi,\theta)}(x)\right)^2 \cdot f_*^{(\tilde{\phi})}(x)\,\mathrm{d}x \underset{n\to\infty}{\longrightarrow} 0 \tag{5.37}$$

almost surely for abitrary $\tilde{\phi} \in \Phi$.

Proof of Theorem 3:

Theorem 3. *Let* $F^{(n)}(x) \underset{n\to\infty}{\longrightarrow} F^{(\phi,\theta)}(x)$, $(\phi,\theta) \in \Phi \times [a,1]$, *for all* $x > 0$ *and let* $(\hat{\phi}_{k_n}, \hat{\theta}_{k_n}) = \widehat{(\phi,\theta)}(F_{k_n})$, *where the sequence* $(k_n)_{n\in\mathbb{N}}$ *fulfills Condition 3.*

a)

$$\Delta\left(F^{(n)}, F^{(\hat{\phi}_{k_n},\hat{\theta}_{k_n})}\right) \underset{n\to\infty}{\longrightarrow} 0 \quad \textit{almost surely.} \tag{5.38}$$

b)

$$\Delta\left(F^{(\phi,\theta)}, F^{(\hat{\phi}_{k_n},\hat{\theta}_{k_n})}\right) \underset{n\to\infty}{\longrightarrow} 0 \quad \textit{in probability.} \tag{5.39}$$

Proof. For simplicity we denote $k = k_n$, $F_k = F_{k_n}$.

Part a)

$$\begin{aligned}
&\Delta\left(F^{(n)}, F^{(\hat{\phi}_k,\hat{\theta}_k)}\right) = \textstyle\int_0^\infty \left(F^{(n)}(x) - F^{(\hat{\phi}_k,\hat{\theta}_k)}(x)\right)^2 \cdot f_*^{(\hat{\phi}_k)}(x)\,\mathrm{d}x \\
&= \Delta\left(F_k, F^{(\hat{\phi}_k,\hat{\theta}_k)}\right) + \textstyle\int_0^\infty \left(F^{(n)}(x) - F_k(x)\right)^2 \cdot f_*^{(\hat{\phi}_k)}(x)\,\mathrm{d}x \\
&\quad +2 \cdot \textstyle\int_0^\infty \left(F_k(x) - F^{(\hat{\phi}_k,\hat{\theta}_k)}(x)\right)\left(F^{(n)}(x) - F_k(x)\right) \cdot f_*^{(\hat{\phi}_k)}(x)\mathrm{d}x \\
&\overset{(\star)}{\leqslant} \Delta\left(F_k, F^{(\phi,\theta)}\right) + 3 \cdot \|F^{(n)} - F_k\|_\infty \longrightarrow 0,
\end{aligned} \tag{5.40}$$

where $\|g\|_\infty := \sup_x |g(x)|$.

$(\star)$ Interim steps:

Since $(\hat{\phi}_k, \hat{\theta}_k) = \underset{(\phi,\theta)}{\arg\min}\, \Delta(F_k, F^{(\phi,\theta)})$, it holds

$$\Delta(F_k, F^{(\hat{\phi}_k,\hat{\theta}_k)}) \leqslant \Delta(F_k, F^{(\phi,\theta)}), \tag{5.41}$$

$$\begin{aligned}
&\int_0^\infty \left(F^{(n)}(x) - F_k(x)\right)^2 \cdot f_*^{(\hat{\phi}_k)}(x)\,\mathrm{d}x \\
&\leqslant \int_0^\infty \left|F^{(n)}(x) - F_k(x)\right| \cdot f_*^{(\hat{\phi}_k)}(x)\,\mathrm{d}x \\
&\leqslant \|F^{(n)} - F_k\|_\infty \cdot \int_0^\infty f_*^{(\hat{\phi}_k)}(x)\,\mathrm{d}x = \|F^{(n)} - F_k\|_\infty,
\end{aligned} \tag{5.42}$$

and

$$\begin{aligned}
&\int_0^\infty \left(F_k(x) - F^{(\hat{\phi}_k,\hat{\theta}_k)}(x)\right)\left(F^{(n)}(x) - F_k(x)\right) \cdot f_*^{(\hat{\phi}_k)}(x)\mathrm{d}x \\
&\leqslant \int_0^\infty \left|F_k(x) - F^{(\hat{\phi}_k,\hat{\theta}_k)}(x)\right| \cdot \left|F^{(n)}(x) - F_k(x)\right| \cdot f_*^{(\hat{\phi}_k)}(x)\mathrm{d}x \\
&\leqslant \int_0^\infty 1 \cdot \|F^{(n)} - F_k\|_\infty \cdot f_*^{(\hat{\phi}_k)}(x)\mathrm{d}x = \|F^{(n)} - F_k\|_\infty.
\end{aligned} \tag{5.43}$$

Then, the convergence follows from Corollary 1.

Part b) We cannot use the same reasoning for the second part, since

$$\left\|F^{(n)}(x) - F^{(\phi,\theta)}(x)\right\|_\infty \geqslant \left|F^{(n)}(0) - F^{(\phi,\theta)}(0)\right| = 1 - \theta \neq 0. \tag{5.44}$$

The same applies if we replace $F^{(n)}$ by F_k:

$$\left\|F_k(x) - F^{(\phi,\theta)}(x)\right\|_\infty = 1 - \theta \neq 0. \tag{5.45}$$

Instead, by using

$$\begin{aligned}
&0 < \Delta\left(F_k, F^{(\hat{\phi}_k,\hat{\theta}_k)}\right) \leqslant \Delta\left(F_k, F^{(\phi,\theta)}\right) \\
&\Longrightarrow \Delta\left(F_k, F^{(\hat{\phi}_k,\hat{\theta}_k)}\right) - \Delta\left(F_k, F^{(\phi,\theta)}\right) \leqslant 0
\end{aligned} \tag{5.46}$$

and

$$\begin{aligned}
&\left|\Delta\left(F^{(\phi,\theta)}, F^{(\tilde{\phi},\tilde{\theta})}\right) - \Delta\left(F_k, F^{(\tilde{\phi},\tilde{\theta})}\right)\right| \\
&= \left|\int_0^\infty \left(F^{(\phi,\theta)}(x) - F_k(x)\right)\left(F^{(\phi,\theta)}(x) + F_k(x) - 2F^{(\tilde{\phi},\tilde{\theta})}(x)\right) f_*^{(\tilde{\phi})}(x)\mathrm{d}x\right| \\
&\leqslant 2 \cdot \int_0^\infty \left|F_k(x) - F^{(\phi,\theta)}(x)\right| f_*^{(\tilde{\phi})}(x)\,\mathrm{d}x,
\end{aligned} \tag{5.47}$$

we obtain

$$\begin{aligned}
&\Delta\left(F^{(\phi,\theta)}, F^{(\hat{\phi}_k,\hat{\theta}_k)}\right) \\
&\quad\leqslant \Delta\left(F^{(\phi,\theta)}, F^{(\hat{\phi}_k,\hat{\theta}_k)}\right) - \Delta\left(F_k, F^{(\hat{\phi}_k,\hat{\theta}_k)}\right) + \Delta\left(F_k, F^{(\phi,\theta)}\right) \\
&\quad\leqslant 2 \cdot \textstyle\int_0^\infty \left|F_k(x) - F^{(\phi,\theta)}(x)\right| \cdot f_*^{(\hat{\phi}_k)}(x)\mathrm{d}x + \Delta\left(F_k, F^{(\phi,\theta)}\right) \\
&\quad\leqslant 2 \cdot \sup_{\tilde{\phi}\in\Phi} L_k(\tilde{\phi}) + \Delta\left(F_k, F^{(\phi,\theta)}\right).
\end{aligned} \tag{5.48}$$

By showing that

$$\sup_{\tilde{\phi}\in\Phi} L_k(\tilde{\phi}) \underset{n\to\infty}{\longrightarrow} 0 \quad \text{in probability,} \tag{5.49}$$

where

$$L_k(\tilde{\phi}) = \int_0^\infty \left|F_k(x) - F^{(\phi,\theta)}(x)\right| \cdot f_*^{(\tilde{\phi})}(x)\mathrm{d}x, \tag{5.50}$$

we receive the desired result.

Since due to Corollary 1, $L_k(\tilde{\phi}) \to 0$ for all $\phi \in \Phi$ and Φ compact, (5.49) is true if $L_k(\tilde{\phi})$ is continuous in $\tilde{\phi}$. As stated in Zorich (2016, p. 422), the mapping $\tilde{\phi} \mapsto \int_0^\infty \Phi_k(x, \tilde{\phi})\,\mathrm{d}x = L_k(\tilde{\phi})$ is continuous if the following two assumptions are fulfilled:

(i) The integrand $\Phi_k(x, \tilde{\phi})$ is continuous regarding $\tilde{\phi}$.

(ii) The integral $L_k(\tilde{\phi})$ converges uniformly on Φ, i.e., $\forall\, \varepsilon > 0\ \exists\, d_0 > 0\ \forall\, d \geqslant d_0$:

$$\left|L_k^{(d)}(\tilde{\phi})\right| := \left|\int_d^\infty \Phi_k(x, \tilde{\phi})\,\mathrm{d}x\right| < \varepsilon, \quad \forall\ \tilde{\phi} \in \Phi. \tag{5.51}$$

Assumption (i) is clearly fulfilled as $\Phi_k(x, \tilde{\phi})$ is a composition of continuous functions regarding $\tilde{\phi}$. For assumption (ii), we bound the function from above as follows:

$$\begin{aligned}
L_k^{(d)}(\tilde{\phi}) &= \textstyle\int_d^\infty \left|F_k(x) - F^{(\phi,\theta)}(x)\right| \cdot f_*^{(\tilde{\phi})}(x)\,\mathrm{d}x \\
&\leqslant \sup_{x\in[d,\infty)} \left|F_k(x) - F^{(\phi,\theta)}(x)\right| \cdot \textstyle\int_d^\infty f_*^{(\tilde{\phi})}(x)\,\mathrm{d}x \\
&\leqslant \sup_{x\in[d_0,\infty)} \left|F_k(x) - F^{(\phi,\theta)}(x)\right|
\end{aligned} \tag{5.52}$$

By choosing d_0 such that $\sup_{x\in[d_0,\infty)} \left|F_k(x) - F^{(\phi,\theta)}(x)\right| < \varepsilon$, the validity of assumption (ii) is established. □

Proof of Theorem 4:

Theorem 4. *Under the same conditions as in Theorem 3, the weak consistency applies, i.e.*

$$\widehat{(\phi,\theta)}(F_{k_n}) \to (\phi,\theta) \quad \textit{in probability.} \tag{5.53}$$

Proof. This proof is based on Theorem 5.9 of van der Vaart (1998). For simplicity, we denote $k = k_n$ and $(\hat{\phi}_k, \hat{\theta}_k)$ instead of $\widehat{(\phi,\theta)}(F_k)$. We define $M(\tilde{\phi},\tilde{\theta})$ as

$$M(\tilde{\phi},\tilde{\theta}) := \Delta(F^{(\phi,\theta)}, F^{(\tilde{\phi},\tilde{\theta})}). \tag{5.54}$$

We will show that

$$\forall \varepsilon > 0 : \inf_{(\tilde{\phi},\tilde{\theta})\in\Theta_\varepsilon} M(\tilde{\phi},\tilde{\theta}) > M(\phi,\theta) = 0, \tag{5.55}$$

where $\Theta_\varepsilon := \{(\tilde{\phi},\tilde{\theta}) \in \Phi \times [a,1] \mid \|(\tilde{\phi}-\phi, \tilde{\theta}-\theta)\| \geqslant \varepsilon\}$. This implies (5.53), since due to (5.55)

$$\forall \varepsilon > 0 \; \exists \delta > 0 : \; (\tilde{\phi},\tilde{\theta}) \in \Theta_\varepsilon \Rightarrow M(\tilde{\phi},\tilde{\theta}) > \delta \tag{5.56}$$

holds, i.e., $\forall \varepsilon > 0 \; \{\|(\hat{\phi}_k - \phi, \hat{\theta}_k - \theta)\| \geqslant \varepsilon\} \subseteq \{M(\hat{\phi}_k, \hat{\theta}_k) > \delta\}$. In Theorem 3 b), we have already shown that $M(\hat{\phi}_k, \hat{\theta}_k) \to 0$ in probability. Therefore, the set on the right hand side converges to zero in probability and so does the set on the left hand side.

To prove (5.55), we firstly show the following two conditions:

1. $M(\tilde{\phi},\tilde{\theta}) > 0 \;\; \forall\, (\tilde{\phi},\tilde{\theta}) \neq (\phi,\theta)$
2. $M(\cdot,\cdot)$ is continuous.

1. Let $(\tilde{\phi},\tilde{\theta}) \neq (\phi,\theta)$. Since the parameters of the family $\{\mathbb{P}_{(\phi,\theta)} \mid (\phi,\theta) \in \Phi \times [a,1]\}$ are identifiable, there exists $b > 0$ such that $\sup_{x>0} |F^{(\phi,\theta)}(x) - F^{(\tilde{\phi},\tilde{\theta})}(x)| > b$. Furthermore, $F^{(\phi,\theta)}$ and $F^{(\tilde{\phi},\tilde{\theta})}$ are continuous for all $x > 0$ and therefore

$$\forall\; 0 < c < b \;\exists\; B_c \in \mathcal{B}_{[0,\infty)} \text{ with } \mu_{\tilde{\phi}}(B_c) > 0 : \inf_{x\in B_c} |F^{(\phi,\theta)}(x) - F^{(\tilde{\phi},\tilde{\theta})}(x)| = c, \tag{5.57}$$

where $\mathcal{B}_{[0,\infty)}$ is the system of all Borel sets. This leads to

$$\begin{aligned} M(\tilde{\phi},\tilde{\theta}) &= \textstyle\int_0^\infty (F^{(\phi,\theta)}(x) - F^{(\tilde{\phi},\tilde{\theta})}(x))^2 \, f_*^{(\tilde{\phi})}(x)\,\mathrm{d}x \\ &\geqslant \textstyle\int_{B_c} (F^{(\phi,\theta)}(x) - F^{(\tilde{\phi},\tilde{\theta})}(x))^2 \, f_*^{(\tilde{\phi})}(x)\,\mathrm{d}x \\ &\geqslant c^2 \cdot \mu_{\tilde{\phi}}(B_c) > 0. \end{aligned} \tag{5.58}$$

2. Analogously to the previous proof of Theorem 3, we use Zorich (2016, p. 422), who states that the mapping $(\tilde{\phi}, \tilde{\theta}) \mapsto \int_0^\infty \Phi(x, \tilde{\phi}, \tilde{\theta})\,\mathrm{d}x = M(\tilde{\phi}, \tilde{\theta})$ is continuous if the following two assumptions are fulfilled:

(i) The integrand $\Phi(x, \tilde{\phi}, \tilde{\theta})$ is continuous regarding $\tilde{\phi}$ and $\tilde{\theta}$.

(ii) The integral $M(\tilde{\phi}, \tilde{\theta})$ converges uniformly on $\Phi \times [a, 1]$, i.e. $\forall\, \varepsilon > 0\ \exists\, d_0 > 0\ \forall\, d \geqslant d_0$:

$$\left|M^{(d)}(\tilde{\phi}, \tilde{\theta})\right| := \left| \int_d^\infty \Phi(x, \tilde{\phi}, \tilde{\theta})\,\mathrm{d}x \right| < \varepsilon, \quad \forall (\tilde{\phi}, \tilde{\theta}) \in \Phi \times [a, 1]. \tag{5.59}$$

Assumption (i) is clearly fulfilled as $\Phi(x, \tilde{\phi}, \tilde{\theta})$ is a composition of continuous functions. For assumption (ii), we bound the function $M^{(d)}$ from above as follows:

$$\begin{aligned} \int_d^\infty \Phi(x, \tilde{\phi}, \tilde{\theta})\,\mathrm{d}x &\leqslant \int_d^\infty f_*^{(\tilde{\phi})}(x)\,\mathrm{d}x = 1 - F_*^{(\tilde{\phi})}(d) \leqslant 1 - F_*^{(\tilde{\phi})}(d_0) \\ &\leqslant 1 - \inf_{\phi \in \Phi} F_*^{(\phi)}(d_0). \end{aligned} \tag{5.60}$$

By choosing d_0 such that $1 - \inf_{\phi \in \Phi} F_*^{(\phi)}(d_0) < \varepsilon$, the validity of assumption (ii) is established. Note that for each ε, there is a d_0, since Φ is compact, $F_*^{(\phi)}(x)$ is continuous in ϕ, and continous in x, monotone increasing as x rises and converges to 1, and therefore, $\inf_{\phi \in \Phi} F_*^{(\phi)}$ is also monotone increasing, continuous and converges to 1.

Lastly, we can show that (5.55) is true. Since Θ_ε is a compact set and M is continuous, the image $M(\Theta_\varepsilon)$ is also compact. Therefore, there is a $(\phi_*, \theta_*) \in \Theta_\varepsilon$ such that

$$\inf_{(\tilde{\phi}, \tilde{\theta}) \in \Theta_\varepsilon} M(\tilde{\phi}, \tilde{\theta}) = M(\phi_*, \theta_*) > 0. \tag{5.61}$$

□

Proof of Theorem 5:

Theorem 5. *Let $(X_n)_{n\in\mathbb{N}}$ be stationary with extremal index $\theta \in (0, 1]$. Let there be a sequence of thresholds $(u_n)_{n\in\mathbb{N}}$ for which $\Delta(u_n)$ holds with $np(u_n) \to \nu \in (0, \infty)$. Let $(r_n)_{n\in\mathbb{N}}$ be a sequence of positive integers with $r_n = o(n)$. Then it holds*

$$\frac{p(u_{r_n})}{\hat{p}_n(u_{r_n})} \to 1 \quad \textit{in probability}, \tag{5.62}$$

where $p(u) = \mathbb{P}(X_1 > u)$ and $\hat{p}_n(u) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i > u\}$ is the relative frequency of exceedances.

Proof.
Set $\nu_n = r_n\, p(u_{r_n})$. It holds $\nu_n \to \nu$. Further, let $\hat{\nu}_n = \frac{1}{q_n} \sum_{i=1}^{r_n \cdot q_n} \mathbb{1}\{X_i > u_{r_n}\}$ with $q_n = \lfloor n/r_n \rfloor$.

This can be interpreted as the sequence $X_1, \ldots, X_n$ being divided into q_n blocks of length r_n. Consequently, $\hat{\nu}_n$ is the mean of the number of exceedances in each block. It holds that $\hat{\nu}_n \to \nu$ in probability, see Robert (2009, Lemma 6.4).

Since $r_n \hat{p}_n(u_{r_n}) \sim \hat{\nu}_n$, we lastly get

$$\frac{\hat{p}_n(u_{r_n})}{p(u_{r_n})} = \frac{r_n\, \hat{p}_n(u_{r_n})}{r_n\, p(u_{r_n})} \sim \frac{\hat{\nu}_n}{\nu_n} \to 1 \quad \text{in probability as} \quad n \to \infty.$$

□

5.5 Conclusions and discussion

The aim of this chapter was to demonstrate the weak consistency of our estimation method *CMmod1*. To this end, we first generalised the modelling framework introduced in Chapter 3 to make the results applicable to a wide range of scenarios.

In Subsection 5.2.1, we initially assumed that the scale parameter is known. In Subsection 5.2.2, we then considered the situation where the scale parameter $\sigma_n > 0$ is unknown and diverges to infinity. We proved the weak consistency of the estimator for the distributional parameters, if a weakly consistent estimator of σ_n itself is available. Later on, we shifted our focus back to the specific estimation setting of interest. The results from the earlier sections can be transferred, if the parameter space of the parameter vector (β, θ) is restricted to a compact set. In a detailed analysis of the scale parameter $\sigma_u = p(u)^{-1/\beta} L(1/p(u))$, we were able to show that $\hat{p}_n(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leqslant u\}$ is a weakly consistent estimator for $p(u)$. Moreover, we showed that the slowly varying component $l_u = L(1/p(u))$ converges to a constant μ in certain cases. In such cases, the triplet $\widehat{(\beta, \theta, l_u)}(F_k)$ constitutes a weakly consistent estimator for (β, θ, l_u).

In addition to assuming a compact parameter space, the proof relies on further assumptions, which we will discuss again in the following.

Since the estimator is based on minimising the distance between the empirical and the asymptotic distribution function, it is crucial that the empirical distribution function accurately reflects the theoretical one. To ensure this, we assume that the sample $T_{1,n}, \ldots, T_{k_n,n}$ consists of i.i.d. random variables. However, this assumption is generally not fulfilled in our application setting as described in Chapter 3, where we consider a stationary process $(X_n)_{n\in\mathbb{N}}$. Although the original waiting times $(W_n)_{n\in\mathbb{N}}$ are independent and also independent of the magnitudes $(X_n)_{n\in\mathbb{N}}$, this

independence transfers to the IETs $(T_k(u))_{k\in\mathbb{N}}$ only if the sequence $(X_n)_{n\in\mathbb{N}}$ consists of independent random variables. However, a small simulation study, see Appendix C.2, indicates that dependent IETs provide results in a similar good approximation of the true distribution function as artificially generated independent IETs. In the special case where the magnitudes are independent, the IETs are asymptotically Mittag-Leffler distributed and the sample $T_1(u), \dots, T_{k_n}(u)$ consists of independent random variables. This implies $\theta = 1$.

In Subsection 5.2.2, we assumed that σ_n is unknown and can be estimated weakly consistently by $\hat{\sigma}_n$ in advance. To ensure that $\widehat{(\phi,\theta)}(\hat{F}_{k_n})$ remains weakly consistent, $\hat{F}_{k_n}(x) = F_{k_n}(x\hat{\sigma}_n) \to F^{(\phi,\theta)}(x)$ must hold for all continuity points x of $F^{(\phi,\theta)}$. This requirement may be violated, when we estimate σ_n with the same sample that we use for the e.c.d.f. F_{k_n} since in general then $T_{1,n}/\hat{\sigma}_n, \dots, T_{1,k_n}/\sigma_n$ are dependent. However, we would like to point out that in our application (see Section 5.3) we estimate $p(u_n)$ consistently by $\hat{p}_n(u_n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i > u_n\}$ in advance using the magnitudes $X_1, \dots, X_n$, while for the estimation of (β, θ, l_u) the CMmod1 estimator uses the IETs $T_1(u_n), \dots, T_{k_n}(u_n)$, i.e. different samples $X_1, \dots, X_n$ and $T_1(u_n), \dots, T_{k_n}(u_n)$, respectively. Nevertheless, $X_1, \dots, X_n$ and $T_1(u_n), \dots, T_{k_n}(u_n)$ are not stochastically independent since we use the exceedances of $X_1, \dots, X_n$ to calculate the IETs $T_1(u_n), \dots, T_{k_n}(u_n)$. To guarantee the weak consistency, different samples for $\hat{p}_n(u_n)$ and F_{k_n} should be used.

A particular challenge in the proof was the assumption that the sample $T_{1,n}, \dots, T_{k_n,n}$ originates from a continuous distribution on the positive real line and converges only asymptotically to the mixture distribution $\mathbb{P}_{(\phi,\theta)} = (1-\theta)\,\delta_0 + \theta\,\mu_\phi$. This raises the question of whether the estimation method remains consistent if we instead observe samples that already follow a mixture distribution of the form $(1-\theta)\,\delta_0 + \theta\,\mu_\phi$. In this case, the original Cramér-von Mises estimator is weakly consistent, as the sample-independent discrepancy

$$|F_k(0) - (1-\theta)| = |0 - 1 + \theta| = 1 - \theta \tag{5.63}$$

would vanish. Our modified estimator CMmod1 also remains a weakly consistent estimator in this setting. In contrast, we would not recommend using the modification CMmod2 without further investigation in this scenario, as the truncation was specifically added to deal with equation (5.63).

6 Simulation study

This chapter is dedicated to present the simulation results analysing the performance of the minimum distance estimators CMmod1 and CMmod2. All statistical computations are done with R (R Core Team, 2024).

This chapter is organised as follows: We begin by outlining the general approach and describing the simulation settings under consideration. In the second section, we conduct an empirical evaluation of the two minimum distance estimators, CMmod1 and CMmod2. We investigate not only various parameter settings for β and θ, but also examine the impact of the sample size n, the choice of the number of exceedances k, and different data-generating processes, all of which are constructed to satisfy the model assumptions. In the following section, we compare our estimators with established methods in situations where the data-generating process corresponds to one of the special cases: CPP or FPP. Additionally, we analyse the behaviour of the estimators as β approaches one. A subsequent section is devoted to the evaluation of bootstrap procedures, which are applied to real data in the following chapter. Here, we assess their reliability and suitability for practical inference. The chapter concludes with a summary of the key findings and a discussion.

6.1 Simulation setup

To assess the estimation performance, a wide range of settings can be varied. In the following sections, we focus on one aspect at a time while keeping the other parameters fixed or varying them only slightly. This approach helps to limit the overall number of scenarios under consideration.

Unless stated otherwise in the subsequent sections, the simulation data are generated as follows: We simulate sequences $(X_i, W_i)_{i=1}^n$ of length $n = 10\,000$, where the magnitudes X_i follow an MAR process as defined in Subsection 2.2.3. For $\beta < 1$, the waiting times W_i are i.i.d. one-sided stable, as specified in Subsection 5.3.4, and for $\beta = 1$ the waiting times are equidistant equal to one. This setup ensures that the distributional parameters β, θ and $l_u = L(1/p(u))$ are known exactly. We vary the tail parameter β and the extremal index θ across the values 0.2, 0.5, 0.8 and 1, while keeping $l_u = 1$ fixed. The threshold u is selected for each sequence such that there are exactly $k = 200$ exceedances. This leads to a constant estimate of the exceedance probability $\hat{p}_n(u) = k/n = 0.02$. In the context of extreme value theory,

it is standard practice to fix the number of exceedances k rather than the threshold u directly.

For each setting we generate $N = 1000$ simulation runs and report the empirical bias and the empirical root of the mean-square error (RMSE):

$$\text{bias} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\vartheta}_i - \vartheta\right) \qquad \text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\vartheta}_i - \vartheta\right)^2},$$

where $\hat{\vartheta}_i$ is the i-th estimate of a real parameter ϑ.

To minimise the CMmod1 and CMmod2 distances, we use the standard optimisation algorithm *L-BFGS-B* based on quasi-Newton with several starting points (Byrd et al., 1995). The search space for the parameter vector (β, θ, l_u) is restricted to the compact set $[0.05, 1] \times [0.05, 1] \times [0.05, 10]$. As starting points, we use $\{0.25,\ 0.55,\ 0.85\} \times \{0.25,\ 0.55,\ 0.85\} \times \{t_l\}$, where t_l is chosen as follows: If the log-moment estimator of the Mittag-Leffler distribution, introduced in Subsection 3.2.3, lies within $[0.05, 10]$, it is used as t_l; otherwise, t_l is set to 1. Since there is a risk that a local minimum exists near the true parameter values, but a smaller (possibly global) minimum may be located at the lower boundary of the search space, we apply the optimisation procedure for all 9 combinations of starting points. From these, we select the result corresponding to the smallest objective function value, provided that the optimum is not located at the boundary of the search space. If all optimisation runs converge to a parameter vector located at the lower boundary of the search space, we select the one with the smallest objective function value and retain it. In this case, however, the result should be interpreted with caution.

6.2 Performance of CMmod1 and CMmod2

First, we examine the general behaviour of CMmod1 and CMmod2, and compare their performances. In this subsection, the sample size is fixed at $n = 10\,000$, and the number of exceedances is set to $k = 200$.

As previously described, we optimise the three parameters β, θ, and l_u over a compact search space. In Chapter 4, we have already shown through calculations and heatmaps that CMmod1 in particular can struggle with the lower bound of the search space regarding the parameter θ. This is likely due to the existence of a global minimum at the boundary, which may prevent the optimisation algorithm from identifying a local minimum closer to the true parameter values. These observations are confirmed by our simulation results: Figure 6.1 displays, for parameters β and θ, the relative frequency of simulation runs in which the estimated optimum lies at the lower boundary of the search space, i.e. at 0.05. For parameter l_u, the figure shows the

relative frequency of runs in which the estimate lies at either boundary of the search space, 0.05 or 10. In the latter case, the global minimum mostly corresponds to the upper bound. We observe that both estimators handle the parameter β reasonably well regarding the lower boundary. This also applies for CMmod2 regarding the parameter θ, while CMmod1 in some scenarios, when the true parameter value for θ is 0.2, struggles. However, estimating l_u appears to be challenging for both methods, with CMmod1 performing worse than CMmod2 in this regard. It is worth noting that CMmod1 might also perform better in terms of θ if we were to restrict the search range further, e.g. to $[0.1, 1]$.

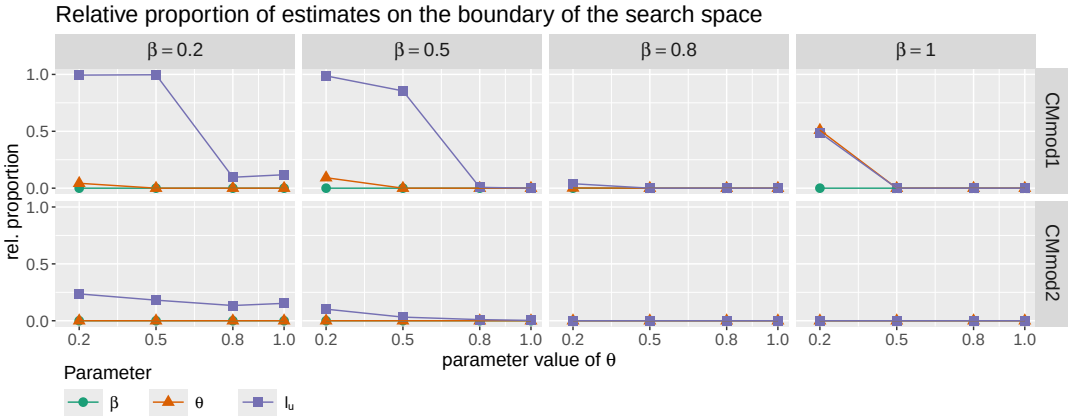

Figure 6.1: Relative frequency of simulation runs in which the estimate lies at the (lower) boundary of the search space for different parameter values with sample size $n = 10\,000$ and number of exceedances $k = 200$. For β and θ, the lower boundary is 0.05. For the scale parameter l_u, the boundaries are 0.05 and 10.

Next, we turn to the RMSE and bias of the two minimum distance estimators. Figure 6.2 shows RMSE values of CMmod2 plotted against those of CMmod1. Each point corresponds to one simulation scenario. If a point lies below the grey diagonal, CMmod2 outperforms CMmod1. Conversely, points above the diagonal indicate superior performance by CMmod1. We find that CMmod2 outperforms CMmod1 in the majority of scenarios. Nonetheless, there are scenarios in which CMmod1 achieves better results for at least one parameter. However, these points tend to lie close to the diagonal, meaning that CMmod1's RMSE is only marginally better or occur when both RMSE values are low. In contrast, in some cases, CMmod1's RMSE is more than twice as large as that of CMmod2. In addition, for both parameters β and θ, CMmod2's RMSE is alway lower than 0.1.

We now additionally examine Figure 6.3 to analyse the behaviour of the two estimators with respect to the true parameter values of β and θ. It becomes evident that CMmod2 is less sensitive to changes in parameter settings. In many cases, both RMSE and bias decrease with increasing values of β and θ. However, there are instances where

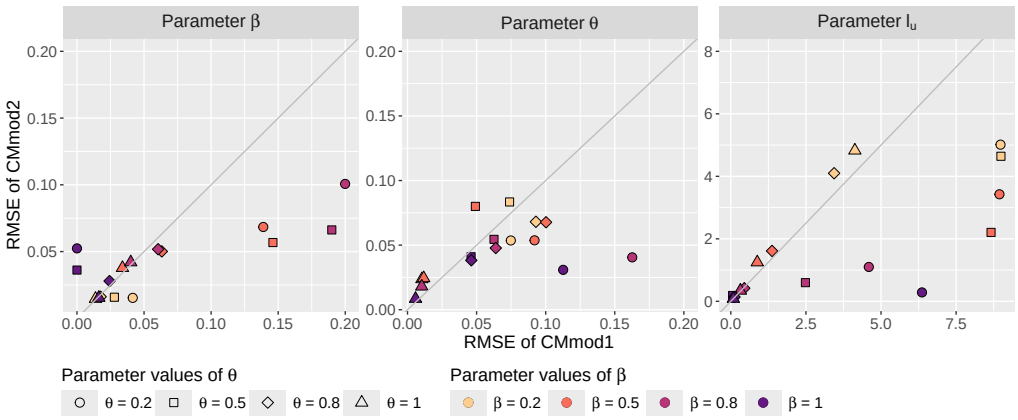

Figure 6.2: Scatter plot comparing the RMSEs of CMmod1 (x-axis) and CMmod2 (y-axis). Each point represents one simulation scenario with sample size $n = 10\,000$ and number of exceedances $k = 200$. Points below the diagonal indicate superior performance of CMmod2.

RMSE and bias are small for both low values (0.2) and high values (1), but higher for intermediate values (e.g., 0.5 or 0.8). For $\beta = 0.8$ and small values of θ, the bias with respect to parameter β is around 0.2. This implies that when β is close to 1 and θ is low, β tends to be overestimated reaching the upper limit of 1. However, CMmod2 also exhibits a notable positive bias, particularly for the parameter θ, as well as for the scale parameter l_u when β takes small values.

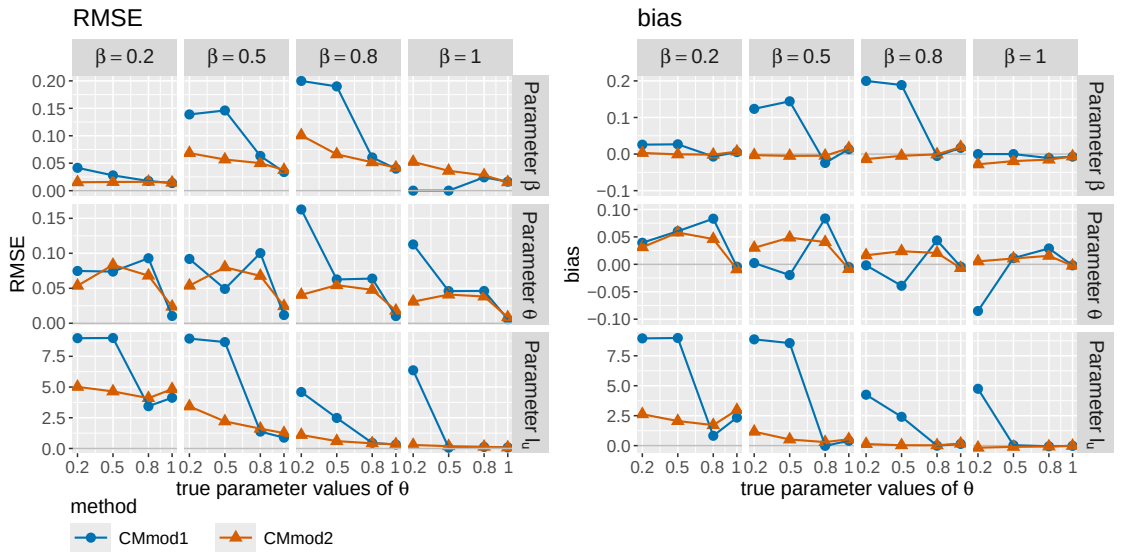

Figure 6.3: Comparison of the RMSE (left) and bias (right) of the CMmod1 and CMmod2 estimators for the parameters β, θ and l_u for different parameter values with sample size $n = 10\,000$ and number of exceedances $k = 200$.

In summary, we observe a clear advantage of the estimator CMmod2. In the following section, we will further investigate this by considering additional sample sizes n and numbers of exceedances k.

6.2.1 Sample size n and number of exceedances k

In the previous section, we fixed the sample size at $n = 10\,000$ and used the same number of exceedances, $k = 200$, across all scenarios. Assuming, on average, hourly observations, this would correspond to a data series spanning approximately 14 months. If extreme events, defined as the 200 largest values, occurred at regular intervals over time, this would imply one extreme event every 50 hours. If instead we had daily observations, the dataset would cover a period of roughly 28 years, with an extreme event occurring approximately every 50 days under the same assumption of regular temporal occurrence.

Thus, although a sample size of 10 000 may initially appear large, it is in fact a realistic choice. Moreover, such a large sample is necessary because we subsequently thin the data using the POT method. The proportion of exceedances must remain sufficiently low to ensure the validity of the asymptotic results. At the same time, however, we require a sufficiently large number of exceedances to reliably estimate the empirical distribution function.

In this section, our focus is on the sample size and the number of exceedances. To this end, we vary both parameters as follows:

- $n \in \{5\,000, 10\,000, 20\,000, 40\,000\}$,
- $k \in \{50, 100, 200, 400, 800, 1600\}$.

Note that changing k also implicitly alters the threshold u. For a fixed sample size n, a smaller k corresponds to a higher threshold u.

First of all, we can report that the results of the previous section remain true for other sample sizes n and number of exceedances k: CMmod1 struggles more with finding the local minimum (see Figure 6.4) especially in case of low true parameter values for β and θ. Moreover, CMmod2 outperforms CMmod1 in most scenarios (see Figure 6.5).

Figure 6.4 displays the relative frequencey of simulation runs in which the estimated optimum lies at the (lower) boundary of the search space across all parameter settings corresponding to Figure 6.1 of the previous section. For the CMmod2 estimator, we observe a clear improvement with increasing sample size n and a suitable choice of k, the number of exceedances. In contrast, for CMmod1, the sample size appears to have a smaller effect; however, the relative frequency of boundary estimates decreases as the number of exceedances increases. For both estimation approaches and across all sample sizes, the simulation results demonstrate that the choice of k has a substantial

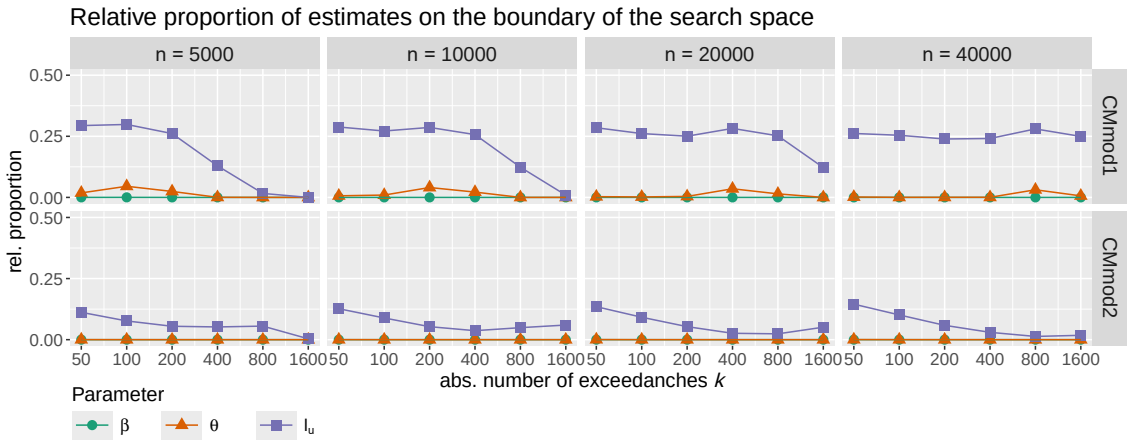

Figure 6.4: Relative frequency of simulation runs in which the estimate lies at the (lower) boundary of the search space dependent of the sample size n and number of exceedances k, and across all parameter settings. For β, θ and l_u the lower boundary is 0.05. For the scale parameter l_u, the upper boundary is 10.

influence on the likelihood of boundary estimates. A high relative proportion of boundary estimates indicates that, in these cases, the optimisation procedure fails to find a local minimum near the true parameter values, resulting instead in estimates at the boundary of the search space. Based on these findings, for real-data applications, it is advisable to apply multiple choices of k when fitting the model, particularly if an initial estimate lies at the boundary of the search space. Such boundary estimates may indicate that the search space has been chosen too restrictively or that the optimisation procedure has failed to locate the local optimum.

We now turn to Figure 6.5, which presents the RMSE and bias across all parameter settings, with a particular focus on the influence of the sample size n and the number of exceedances k. The results clearly show that CMmod2 outperforms CMmod1. For both estimation approaches, we generally observe a reduction in both RMSE and bias as the sample size increases. However, the choice of k also has a decisive impact on the performance. In fact, for smaller sample sizes, a favourable choice of k can result in noticeably lower RMSE compared to larger sample sizes with an unfavourable choice of k. This effect is particularly pronounced for CMmod1, suggesting that CMmod2 is more robust with respect to the choice of the number of exceedances k.

In the following, we summarise several noteworthy and, in some cases, on first sight surprising findings from the analysis:

- Across all parameter settings, the bias for all three parameters is predominantly positive. This indicates a general tendency towards overestimation rather than underestimation. However, it is important to note that Figure 6.5 shows RMSE

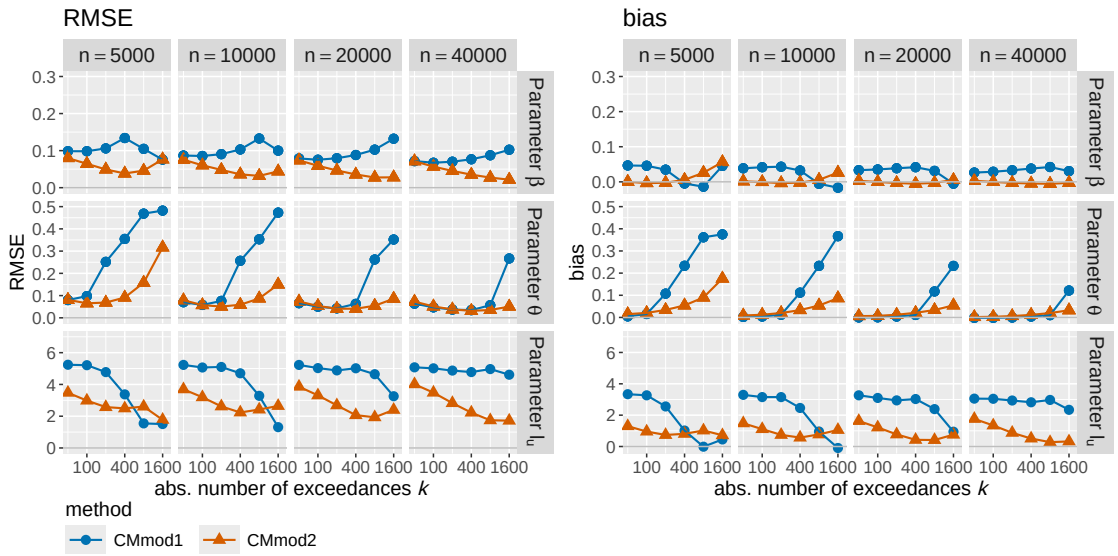

Figure 6.5: Comparison of the RMSE (left) and bias (right) of the CMmod1 and CMmod2 estimators for the parameters β, θ and l_u dependent of the sample size n and number of exceedances k, and across all parameter settings.

values averaged over different parameter settings, and for individual settings, the empirical bias can indeed be negative. One such example is the bias for parameter β in the case $\beta = 1$.

- For CMmod1, the RMSE does not behave consistently across the three parameters considered. While the RMSE for β and θ increases with larger k, the RMSE for the scale parameter l_u decreases with increasing k. A similar pattern is observed for the bias. This highlights the inherent difficulty in selecting an appropriate number of exceedances, or equivalently, a suitable threshold.
- Also for CMmod1, the RMSE and bias for the scale parameter l_u do not decrease with increasing sample size n, as would generally be expected. Comparing the curves in Figure 6.5 with the corresponding curves in Figure 6.4, a clear relationship becomes apparent: the RMSE appears to be strongly influenced by simulation runs where the estimated optimum for l_u lies at the upper boundary of the search space, i.e., at 10. A similar relationship can be observed for the CMmod2 estimator; however, the proportion of boundary estimates is considerably lower in this case. This circumstance may also explain the contrasting behaviour described earlier.

Thus far, we have either examined the influence of the true parameter values under fixed sample size n and number of exceedances k, or the influence of n and k averaged across all true parameter values. However, as shown in Section 5, the true parameter values themselves have a substantial impact on the performance metrics under consideration (see Figures 6.1 to 6.3). We therefore compare the RMSE and

bias again, this time for two specific scenarios: a smaller sample size of $n = 10\,000$ with $k = 200$ exceedances, and a larger sample size of $n = 40\,000$ with $k = 400$ exceedances. These scenarios are chosen such that the number of exceedances k increases with n, while the relative proportion k/n decreases. This comparison, illustrated in Figure 6.6, reveals that for the CMmod2 estimator, the RMSE and bias decrease with increasing sample size, or at least are already close to zero for both sample sizes, which is in line with the theoretical findings in Section 5. For the CMmod1 estimator, a similar trend is observed, with the exception of a few scenarios where $\theta = 0.2$ or $\theta = 0.5$.

In conclusion, the analysis confirms the robustness and superior performance of CMmod2 while also emphasising the critical, yet challenging, task of choosing an appropriate number of exceedances k. Moreover, these findings are consistent with the theoretical results presented in Chapter 5.

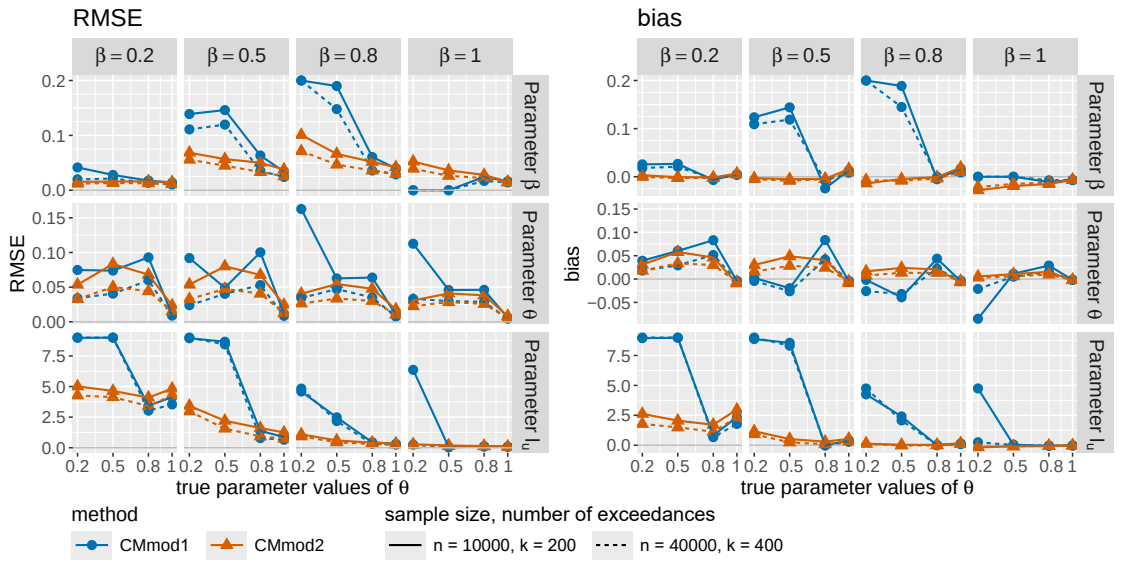

Figure 6.6: Comparison of the RMSE (left) and the bias (right) of the CMmod1 and CMmod2 estimators for the parameters β, θ and l_u, for different parameter values, and as dependent on the sample size n and the number of exceedances k. Considered are $n = 10\,000, k = 200$ (solid line) and $n = 40\,000, k = 400$ (dashed line).

6.2.2 Magnitude processes and waiting time distributions

In the previous subsections, we consistently generated our data using the same data-generating process. In this section, we aim to introduce and compare different stationary processes for the magnitudes $(X_n)_{n\in\mathbb{N}}$ and various distributions for the waiting times $(W_n)_{n\in\mathbb{N}}$. When selecting these processes and distributions, we are

constrained by the requirement that the true parameter values for β, θ and l_u must be known.

Stationary processes with extremal index θ

Up to this point, we have used the MAR process with a Fréchet marginal distribution for the magnitudes, as introduced in Subsection 2.2.3. We now extend our analysis to include the class of *moving maxima processes* (MM).

Let $Z_{-m}, Z_{-m+1}, \ldots, Z_0, Z_1, \ldots$ be i.i.d. Fréchet random variables, and let $\alpha_j \geqslant 0$ for $j = 0, \ldots, m$ be coefficients satisfying $\sum_{i=0}^{m} \alpha_j = 1$.

The MM process $(X_n)_{n\in\mathbb{N}}$ is then defined as

$$X_i = \max_{0\leqslant j\leqslant m} \alpha_j Z_{i-j} \quad \text{for all} \quad i \in \mathbb{N}. \tag{6.1}$$

The extremal index of this process is given by $\theta = \max_{0\leqslant j\leqslant m} \alpha_j$ (see, e.g., Holešovský and Fusek, 2022).

We choose the coefficients α_j to be monotonically decreasing, as this ensures that the process satisfies the local dependence condition $D^{(2)}(u_n)$. As described in Subsection 4.1.2, the $D^{(k)}(u_n)$ condition says, that within a cluster an exceedance is almost always followed by another exceedance within $k-1$ consecutive observations. This condition is not further discussed in this work, as it is not required for our estimators CMmod1 and CMmod2. However, it is relevant for certain other estimators of the extremal index in the special case $\beta = 1$, as explained in Subsection 4.1.2. The MAR process also satisfies this condition. Note that for $\theta = 1$, the MAR and MM process are identical, since the the random variables $(X_n)_{n\in\mathbb{N}}$ are then stochastically independent, standard Fréchet distributed.

The specific choice of coefficients is provided in Table 6.1.

Table 6.1: Specification of the coefficients α_j for the MM process dependent on the extremal index θ.

	m	coefficients $\alpha_0, \ldots, \alpha_m$
$\theta = 1$	0	$\alpha_0 = 1$
$\theta = 0.8$	1	$\alpha_0 = 0.8, \alpha_1 = 0.2$
$\theta = 0.5$	3	$\alpha_0 = 0.5, \alpha_1 = \alpha_2 = 0.2, \alpha_3 = 0.1$
$\theta = 0.2$	6	$\alpha_0 = \alpha_1 = \alpha_2 = 0.2, \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0.1$

Positive valued distributions with tail parameter β

For the waiting times, we previously used a degenerate distribution at 1 for $\beta = 1$, i.e., $W_n \equiv 1$ for all $n \in \mathbb{N}$, and for $\beta < 1$ we employed the one-sided stable distribution, as described in Subsection 5.3.4. We now extend the set of distributions as follows:

For $\beta = 1$, we consider the following distributions, all with an expected value of 1:

- The Dirac measure at point one,
- the exponential distribution,
- the Pareto distribution with shape parameter $\alpha = 1.5$, where only the first moment exists,
- the Pareto distribution with shape parameter $\alpha = 2.5$, where both the first and second moments exist.

For $\beta < 1$, we additionally consider the Mittag-Leffler distribution and the Pareto distribution as described in Subsection 5.3.4, with $\varsigma = 1$. Figure 6.7 illustrates the probability density functions or probability mass functions of the respective distributions.

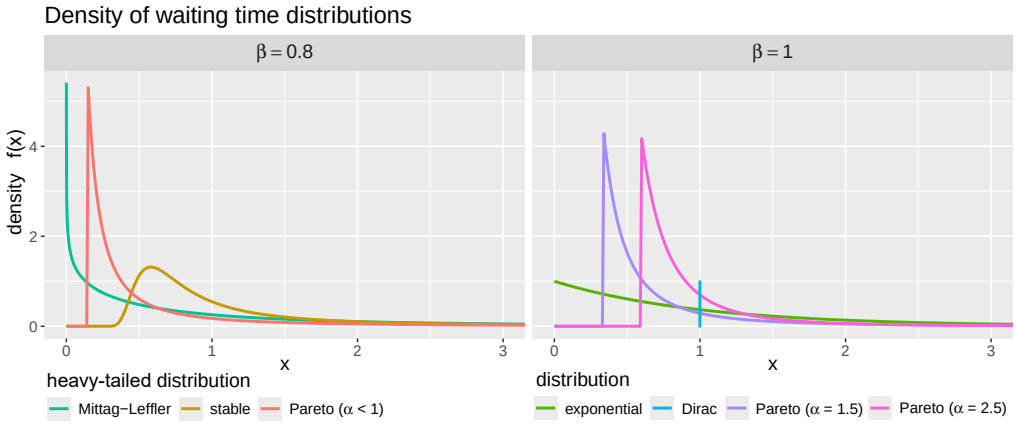

Figure 6.7: Density of the continuous waiting time distributions with tail parameter $\beta = 0.8$ (left) or tail parameter $\beta = 1$ (right). In case of the Dirac measure, it is the probability mass function.

The results of the analysis indicate that the choice of the stationary process for the magnitudes and the distribution for the waiting times does not have a structural influence on the estimation results. Across all considered settings, the overall behaviour of the estimators remains qualitatively unchanged, regardless of the specific distribution chosen, see Figures D.1 and D.2 in the Appendix. This suggests that the proposed estimation procedures are robust with respect to the distributional assumptions for the waiting times.

6.3 Comparison with established estimators

This section investigates the performance of our estimators in situations where the data-generating process corresponds exactly to one of the two submodels: the CPP or the FPP. In these settings, we compare CMmod1 and CMmod2 with established estimators for θ and β, respectively, as introduced in Section 4.1. If our estimators yield comparable results in these special cases, this would suggest that explicit model selection may not be necessary in practice. Instead, one could rely on the generalised FCPP model and estimate both parameters jointly using our proposed methods.

It is important to emphasise that the benchmark estimators used for comparison assume knowledge of the underlying submodel and therefore estimate only a one of both parameters β and θ together with the scale parameter. In contrast, CMmod1 and CMmod2 jointly estimate both θ and β without requiring prior model specification. However, if it is known in advance that the data follow one of the submodels, our estimation procedures can be adapted accordingly by restricting the optimisation to the unknown parameter only. This simplification may yield even more accurate estimates, as it reduces the dimensionality of the estimation problem and avoids potential identifiability issues.

Throughout this section, we use a fixed sample size of $n = 10\,000$ and a number of exceedances $k = 200$. We again consider the MAR and MM processes for the magnitudes, along with the different waiting time distributions discussed previously. Since the magnitude-generating processes (MAR vs. MM) showed no structural effect on estimation performance, we focus here exclusively on results based on the MAR process. In the CPP model with $\theta < 1$, corresponding results for the MM process are shown in additional figures in Appendix D. For the FPP model, where $\theta = 1$, the MAR and MM process are identical. In the CPP model, we vary the extremal index $\theta \in \{0.2, 0.3, \ldots, 0.9, 1\}$. In contrast, under the FPP model, we vary the tail parameter $\beta \in \{0.2, 0.3, \ldots, 0.9, 0.95, 0.99\}$. We deliberately exclude the case $\beta = 1$ from the FPP model, as some of the heavy-tailed distributions used are only defined for $\beta < 1$. Instead, we examine the limiting behaviour as $\beta \to 1$.

6.3.1 Submodel CPP

In the CPP submodel with $\beta = 1$, we compare our estimators CMmod1 and CMmod2 with three established estimators of the extremal index θ, all based on interexceedance times:

- the interval estimator of Ferro and Segers (2003),
- the K-gaps estimator of Süveges and Davidson (2010) with auxility parameter $K = 2$, and

- the censored estimator of Holešovský and Fusek (2020) with auxility parameter $D = 2$.

For the implementation of the K-gaps and censored estimators, adapted versions of the functions `kgaps` and `dgaps` from the R-package `exdex` (Northrop and Christodoulides, 2023) were used. The interval estimator was implemented independently. To accommodate the setting with non-equidistant waiting times, the terms $T_i(u) - 1$ and $T_i(u) - 2$ in the original formula, see Equation (4.4), were replaced by $\max\{T_i(u) - 1, 0\}$ and $\max\{T_i(u) - 2, 0\}$, respectively.

Since for $\beta = 1$ it holds that $\sigma_u = \frac{\mathbb{E}(W_1)}{p(u)}$, we have $l_u = \mathbb{E}(W_1)$. This allows for a separate estimation of l_u, as described in Subsection 4.1.1, via the following expression:

$$\widehat{l_u} = \frac{k}{n} \frac{1}{k} \sum_{i=1}^{k} T_i(u). \tag{6.2}$$

where $T_1(u), \ldots, T_k(u)$ denote the sample of IETs.

The results of this comparison are presented in Figure 6.8. In addition to the findings from the previous section, we observe that our estimators, CMmod1 and CMmod2, produce RMSE and bias values that are comparable to, and sometimes even superior to, those of the established methods for the extremal index θ. The RMSE of the interval estimator is consistently higher than that of CMmod2 across all settings. The same is true for CMmod1, with a few exceptions in scenarios where $\theta < 0.4$. These exceptions likely coincide with an increased RMSE for the scale parameter l_u, which may negatively influence the joint estimation. The interval estimator performs particularly poorly when the waiting times follow a Pareto distribution with shape parameter $\alpha = 1.5$. A likely explanation is that, as discussed in Subsection 4.1.2, the interval estimator relies on the existence of the first two moments of the underlying mixture distribution. However, the Pareto distribution used in this setting only possesses a finite first moment.

The K-gaps and censored estimators behave quite similarly overall. However, while the RMSE of the K-gaps estimator increases for $\theta > 0.8$, the RMSE of the censored estimator continues to decrease. Both estimators perform particularly well in the case of equidistant waiting times (i.e., the Dirac measure). In contrast, they face greater challenges when the waiting times are random (i.e., exponentially or Pareto distributed), especially for small values of θ, where the bias increases as $\theta \to 0$. It is worth noting that both estimators require the specification of an auxiliary parameter K or D, respectively. In our setup, both were fixed at $K = D = 2$. This choice is reasonable in the absence of prior information, as the dependence condition $D^{(k_0=2)}(u_n)$ is satisfied and the condition $K = D \geqslant 1 = k_0 - 1$ therefore holds. Nevertheless, alternative choices for these parameters could potentially improve estimation accuracy. This highlights a general limitation of both methods: the

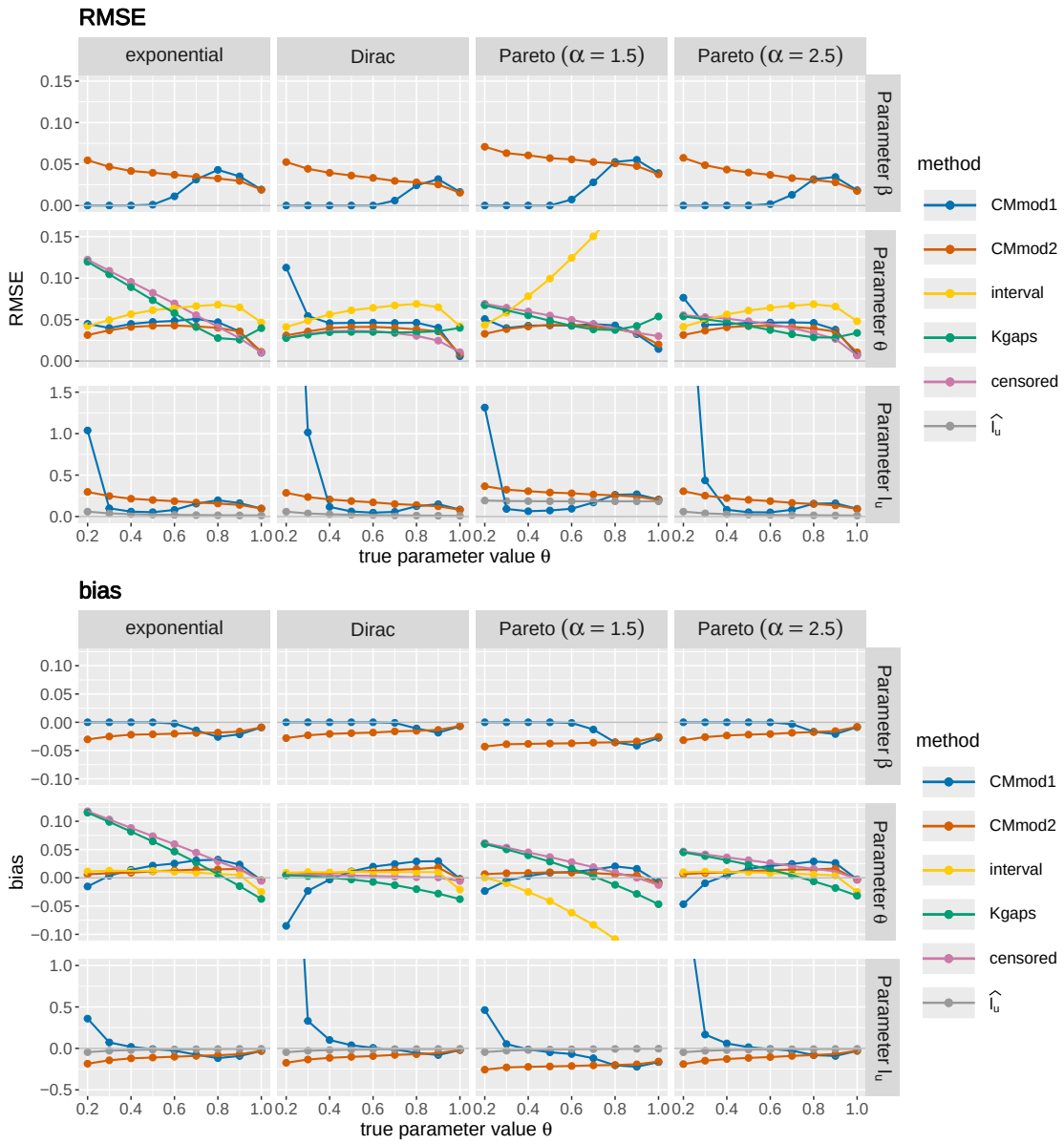

Figure 6.8: Comparison of the RMSE (top) and bias (bottom) of CMmod1 and CMmod2 with different established estimators for the parameters (β,) θ and l_u for $\beta = 1$ and different parameter values of θ, and for the MAR process and a variety of waiting time distributions using sample size $n = 10\,000$ and number of exceedances $k = 200$.

necessity of choosing an auxiliary parameter that directly influences estimation accuracy, while the corresponding dependence condition $D^{(k)}(u_n)$ is often difficult to verify in practice.

When comparing RMSE and bias values for the scale parameter l_u, we find that the

separate estimator $\widehat{l_u}$ performs better than both CMmod1 and CMmod2. However, this is not surprising, as l_u exploits the prior knowledge that $\beta = 1$. Still, our joint estimators also achieve sufficiently good performance in estimating l_u.

Overall, the analysis demonstrates that CMmod2, and CMmod1 for $\theta \geqslant 0.4$, can effectively compete with specialised estimators of the extremal index, even though they simultaneously estimate an additional parameter.

6.3.2 Submodel FPP

In the FPP submodel with $\theta = 1$, we compare our estimators CMmod1 and CMmod2 with the two commonly used estimators for the Mittag-Leffler distribution:

- the log-moment estimator of Cahoy et al. (2010), and
- the maximum likelihood estimator.

Both are implemented in the R-package `MittagLeffleR` (Gill and Straka, 2017). As the implementation permits values of β greater than one, we impose an upper bound of one to ensure interpretability and consistency with the model assumptions.

Since all four estimation methods behave particularly differently for Pareto distributed waiting times as $\beta \to 1$, we begin by discussing the results for the remaining scenarios and address the Pareto case separately in more detail afterwards.

Figure 6.9 presents the results for scenarios with Mittag-Leffler and stable distributed waiting times. The maximum likelihood estimator (mle) achieves the best performance in almost all settings considered. This finding supports the results of Hees et al. (2021), who showed that the maximum likelihood estimator often outperforms the log-moment estimator. Both estimators perform better under Mittag-Leffler distributed waiting times, whereas stable distributions lead to increased bias and, consequently, higher RMSE values.

Our estimators CMmod1 and CMmod2 show little sensitivity to the type of waiting time distribution. Between the two, CMmod1 generally yields slightly lower RMSE and bias values and thus outperforms CMmod2 in most scenarios. Overall, both estimators perform similarly well to the log-moment estimator, and in some settings even slightly better, despite the additional task of estimating the parameter θ simultaneously.

We now turn our attention to the case of Pareto-distributed waiting times. To set the context, we first examine how all three considered waiting time distributions behave as the tail parameter β approaches 1. Figure 6.10 shows the probability densities of the Mittag-Leffler, stable, and Pareto distributions for $\{0.9, 0.95, 0.99\}$. It is evident from the figure that the Mittag-Leffler distribution, as described in Subsection 2.1.3, converges to the exponential distribution. Similarly, the stable distribution increasingly concentrates mass around 1 as β increases, implying convergence to a Dirac

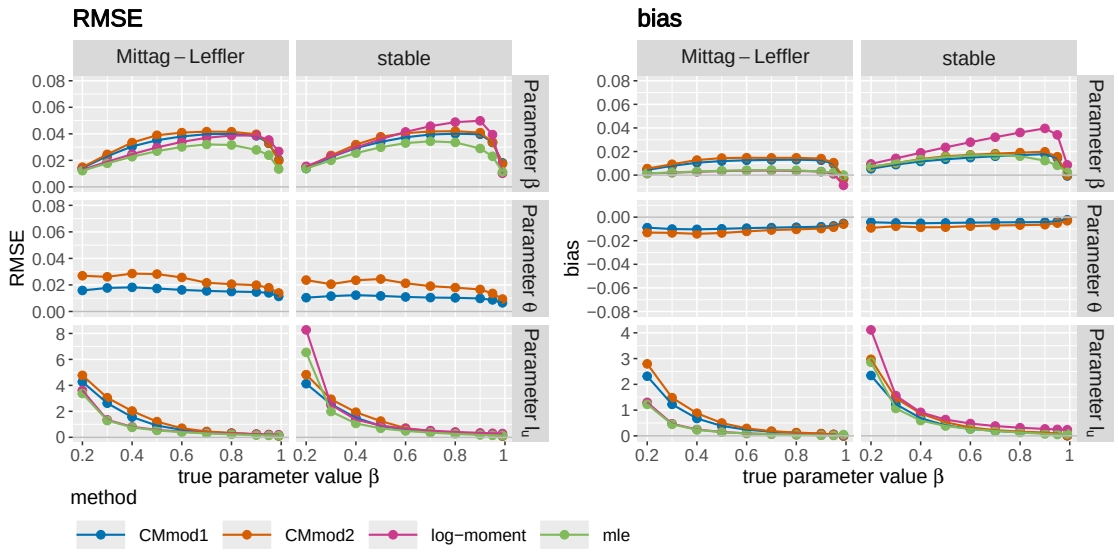

Figure 6.9: Comparison of the RMSE (left) and bias (right) of CMmod1 and CMmod2 with the maximum likelihood and log-moment estimators for the parameters $\beta(, \theta)$ and l_u for $\theta = 1$ and different parameter values of β, and for Mittag-Leffler and stable distributed waiting times using sample size $n = 10\,000$ and number of exceedances $k = 200$.

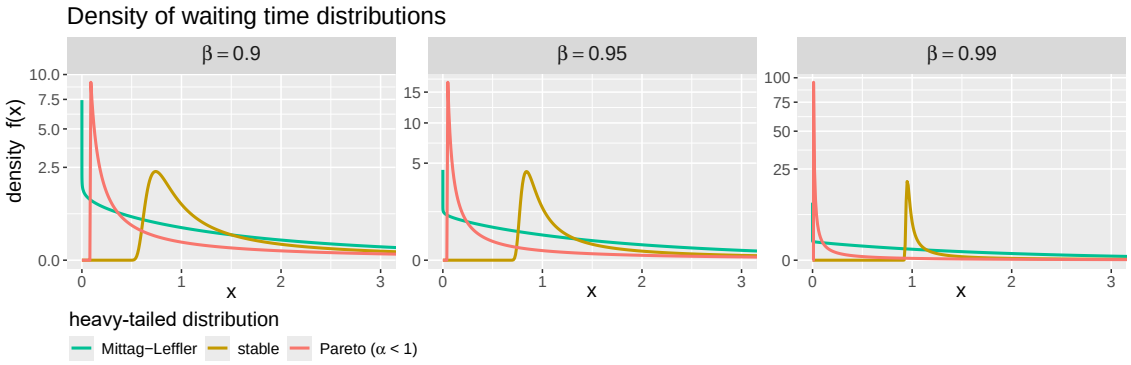

Figure 6.10: Density of the continuous heavy-tailed waiting time distributions with tail parameter $\beta = 0.9$ (left), $\beta = 0.95$ (middle) and $\beta = 0.99$ (right). Note that the y-axis is log-transformed.

measure at 1. Both of these limiting distributions have a positive expectation and are consistent with the model assumptions. In fact, we have already analysed them in the case $\beta = 1$.

In contrast, the Pareto distribution behaves fundamentally differently: as $\beta \to 1$, it concentrates more and more probability mass near zero, converging to a Dirac measure at 0. This limit has expectation zero and therefore violates the model assumptions. This likely explains the increase in RMSE and the emergence of

negative bias for all four estimators of β as $\beta \geqslant 0.8$, as shown in Figure 6.11 (left). At the same time, the scale parameter l_u is increasingly underestimated and tends towards values close to zero. Both effects increase the probability of small IETs under the FCPP model.

The RMSE and bias of CMmod1 further show that, in the extreme case $\beta = 0.99$, the extremal index θ is underestimated and often hits the lower boundary of the search space. This implies that the Dirac component in the mixture distribution receives high weight, even though its true weight in this setting is zero. Interestingly, in the same scenario, the estimate of β improves in accuracy, while l_u is estimated at the upper boundary of the search space. CMmod2 exhibits a similar, though less pronounced, behaviour for $\beta = 0.99$: the estimate for θ is slightly biased downward, while the RMSE for β decreases. The scale parameter l_u, however, continues to be underestimated.

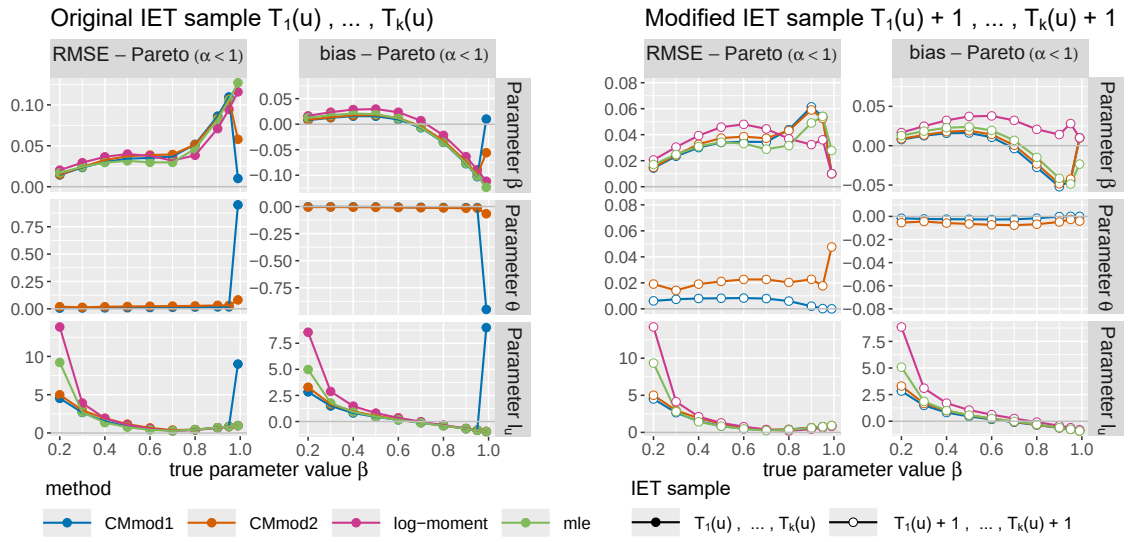

Figure 6.11: Comparison of the RMSE and bias of CMmod1 and CMmod2 with the maximum likelihood and log-moment estimators for the parameters $\beta($, $\theta)$ and l_u for $\theta = 1$ and different parameter values of β, and for Pareto waiting times using sample size $n = 10\,000$ and number of exceedances $k = 200$. On the left, the original IET sample $T_1(u), \dots T_k(u)$, and on the right side, the modified sample $T_1(u) + 1, \dots T_k(u) + 1$ is used.

To improve the estimation in this problematic setting, we propose using a modified sample of interexceedance times $S_1(u), \dots, S_k(u)$, defined by adding 1 to each observation in the original sample: $S_i(u) = T_i(u) + 1$ for all $i = 1, \dots, k$. This modification is theoretically justified: if Theorem 1 holds for $T(u)$, then it also holds

for $S(u) = T(u) + 1$, since

$$\frac{S(u)}{b(1/p(u))} = \frac{T(u)+1}{b(1/p(u))} = \frac{T(u)}{b(1/p(u))} + o(1) \xrightarrow{d} T_{(\beta,\theta,1)}. \tag{6.3}$$

Figure 6.11 (right) shows the results obtained using the modified sample. For $\beta < 0.8$, the effect of the modification is negligible. However, for $\beta \geqslant 0.8$, the RMSE for all estimators is substantially reduced, demonstrating the practical benefit of this simple adjustment.

These findings confirm that CMmod1 and CMmod2 remain competitive even when the model reduces to a submodel, making them suitable in settings where the model structure is unknown.

6.3.3 Computing time

Overall, CMmod1 and CMmod2 show quite satisfactory performance even in both special cases, despite needing to estimate one parameter more than the competitors which are designed for these scenarios. A drawback is their relatively high computational costs. Numerical optimisation is needed to determine the triplet $(\hat{\beta}, \hat{\theta}, \hat{l}_u)$ that minimises the distance function. Because of a possible global minimum at the search boundary or multiple local minima, we used several initialisations $\{0.25, 0.55, 0.85\}^2 \times \{t_l\}$ which further increased computation time. The computing time seems to be linear in the number of exceedances k, as illustrated in Figure 6.12. For comparison, Figure 6.12 also includes the runtime of the maximum likelihood method in the special case $\theta = 1$, for which the computing time is also much higher than that of the other estimators.

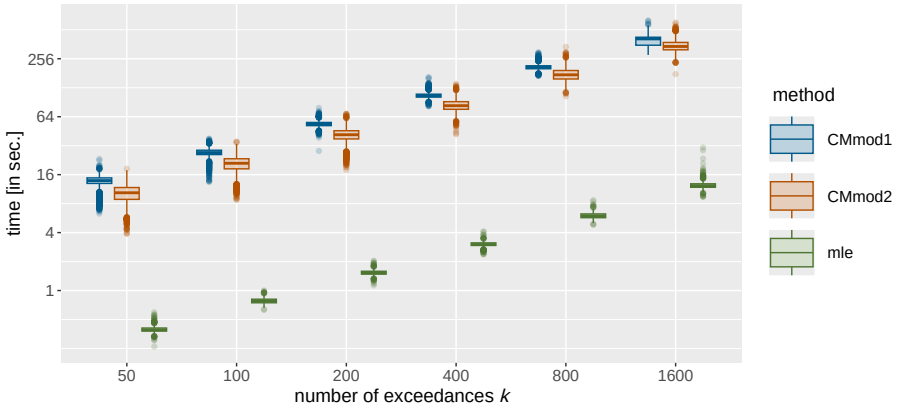

Figure 6.12: Computing time of CMmod1, CMmod2 and the Mittag-Leffler maximum likelihood estimator for an increasing number of exceedances $k = 50, 100, 200, 400, 800, 1600$. Note that both the x-axis and y-axis are log-transformed.

6.4 Parametric bootstrap

In the data analysis in Chapter 7, we use parametric bootstrap methods for statistical inference. We now examine these in a small simulation study to demonstrate their reliability. In the following, we focus on CMmod2, since we use this method in the data analysis. Due to the high computation time required, we choose $N = 100$ simulation runs and for each simulation run with parameter estimates $(\hat{\theta}, \hat{\beta}, \hat{l}_u)$, we then generate $B = 100$ bootstrap samples of the same size from the FCPP with these parameter values and compare the estimates for the original data with those obtained.

For a significance test of the null hypothesis $\beta = 1$ ($\theta = 1$) at a given significance level such as $\alpha = 0.05$, we reject this hypothesis if less than $100 \cdot \alpha\%$ of the B bootstrap estimates for this parameter equal one.

Figure 6.13 displays the estimated rejection rates of the bootstrap tests for the two null hypotheses $H_0 : \beta = 1$ and $H_0 : \theta = 1$. In case of the test for the hypothesis $\beta = 1$, the empirical rejection rates are below 9% under the null hypothesis in all scenarios considered here. The rejection rate increases quickly as β decreases. In case of the test for the hypothesis $\theta = 1$, a more stringent significance level of 3% is maintained in the scenarios considered here. The rejection rate increases moderately at first, but for $\theta \leqslant 0.8$, the test rejects the null hypothesis reliably.

These results highlight the trustworthiness of the classification obtained in the next chapter for the real data (see Figure 7.3). In a similar manner, approximate two-sided confidence intervals could be calculated by using the $\alpha/2$ and the $1 - \alpha/2$ quantile of the estimates obtained for the bootstrap samples as boundaries for each parameter.

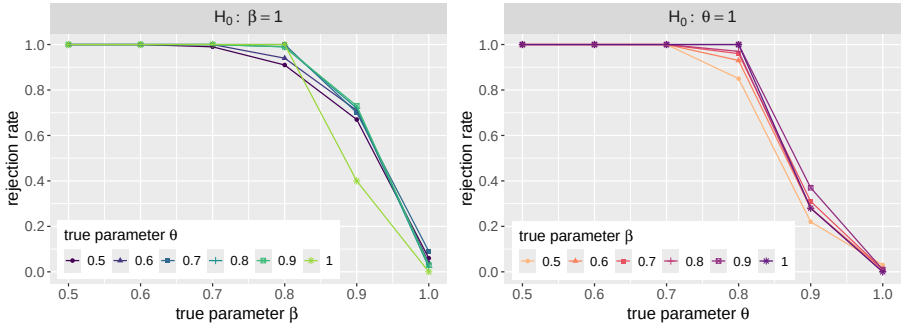

Figure 6.13: Rejection rates of the bootstrap-tests for the hypothesis $\beta = 1$ (left) and $\theta = 1$ (right) with $n = 10\,000$ and $k = 200$.

Since we do not know the true standard errors of our estimators, we estimate them in the data analysis from bootstrap samples. Figure 6.14 shows, for three selected parameter combinations, similar to those for the three locations in the data analysis, that the bootstrap procedure yields reasonable estimates, with the

boxplots representing the distribution of the bootstrap standard errors. The blue dot representing the standard error estimated from $N = 100$ simulation runs performed with the true parameter values.

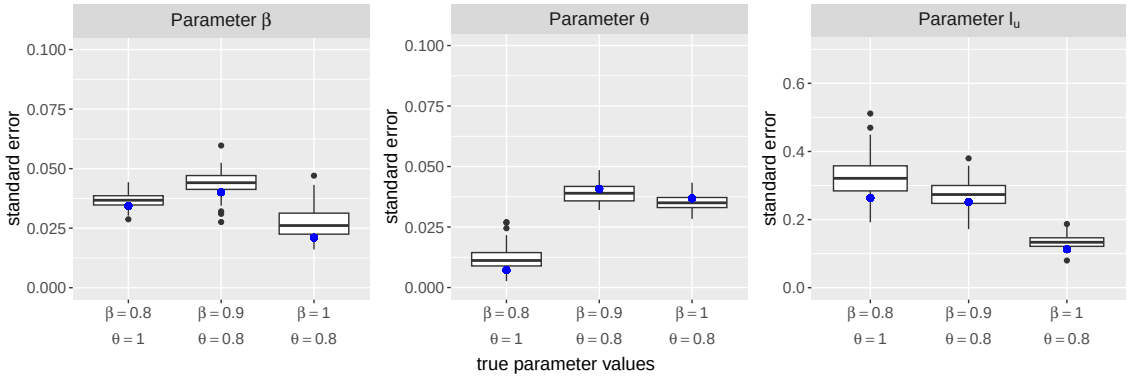

Figure 6.14: Comparison of the bootstrap standard errors (boxplot) with standard errors estimated by simulations (blue points) for three scenarios.

6.5 Conclusions and discussion

Following the previous chapter, in which the theoretical result of weak consistency was established for CMmod1, with minor modifications also applicable to CMmod2 (see Appendix C.1), this chapter has investigated the practical properties of both estimators through an extensive simulation study.

The study provides a comprehensive assessment of the proposed estimators CMmod1 and CMmod2 across a wide range of scenarios. The results show that both estimators perform well under varying sample sizes, thresholds, parameter configurations, and data-generating mechanisms. In particular, CMmod2 demonstrates consistently strong performance and relatively low sensitivity to the choice of the number of exceedances k, making it a robust candidate for practical applications.

However, the study highlights that the number of exceedances k plays a critical role in estimation quality. While estimation accuracy generally improves with increasing sample size, a poor choice of k can still result in boundary estimates or increased RMSE, even in large samples. This underlines the practical importance of carefully selecting the threshold or the number of exceedances, respectively, or performing sensitivity analyses across multiple values of k.

In the analysis of different magnitude processes and waiting time distributions, no clear influence on estimation accuracy was initially observed. However, a more detailed investigation of heavy-tailed Pareto-distributed waiting times revealed that all estimators, including maximum likelihood and log-moment methods in the special

case $\theta = 1$, struggle when the tail parameter β is large. This behaviour is likely due to violations of model assumptions in the limiting distribution as $\beta \to 1$. To address this issue, a simple data transformation of the interexceedance times was proposed, which substantially improved estimation accuracy for $\beta \geqslant 0.8$.

The comparison with established estimators in the CPP and FPP submodels shows that CMmod1 and CMmod2 yield largely comparable, and in some cases even competitive results, despite the added complexity of jointly estimating θ and β. This supports the use of the generalised FCPP model and suggests that, in applications where the underlying extremal structure is unknown, explicit model selection may not be necessary. Only the large computing time might be a noticeable disadvantage.

Finally, a supplementary simulation study demonstrated the practical reliability of parametric bootstrap procedures. The accuracy of bootstrap-based significance tests and standard error estimates was evaluated and found to be satisfactory, supporting their use in the real-data analysis presented in the next chapter.

7 Case study: mid-latitude winter cyclones

In this chapter, we apply the FCPP model proposed in Chapter 3 with the estimation method introduced in Chapter 4 to a real-world dataset.

The case study is motivated by Blender et al. (2015), who presume that due to their cluster behaviour the IETs of extreme mid-latitude cyclones are heavy-tailed. Therefore, they suggest the Mittag-Leffler distribution for fitting. Since they do not consider other explanations for the temporal clustering, we fit our generalised model to the data and compare the results with the Mittag-Leffler distribution and the other submodels.

7.1 Introduction

In meteorology and climate research, the modelling and prediction of extreme events is of great interest. For instance, identifying regions with elevated risk can support civil protection measures. Accurate modelling is also essential for the insurance sector, as extreme weather events may result in significantly more frequent and costly damages. For instance, the damage caused by winter storm Kyrill in January 2007 was estimated to be between 4.7 and 7 billion euros and could have been as high as 40 billion euros if the wind speed had been higher (Alert Air Worldwide, 2007). In the recent past, winter storms Dudley and Eunice (in Germany also known as Ylenia and Zeynep) occurred in quick succession in February 2022 and caused damage of between 3 billion and 5 billion euros (Alert Air Worldwide, 2022). In many applications, such as storms or heavy rainfall, a temporally clustered occurrence of events has been observed. Capturing this clustering within models is crucial, as the impacts and damages are often considerably greater when extreme events occur in close succession.

In this study, we consider mid-latitude cyclones, which strongly affect weather conditions, e.g., temperature, wind, precipitation and cloud cover, and are therefore of great meteorological interest (Dacre and Pinto, 2020).

Mailier et al. (2006) analysed the temporal clustering of mid-latitude cyclones by calculating the dispersion statistic as introduced in (3.2), since it measures the degree

of deviation from a PP. Their results indicate temporal clustering at the west coast of Europe, where the exit region of storm tracks on the North Atlantic is located, while they occur more regularly in the entry region on the east coast of North America. This pattern has been reproduced in our data (see Figure E.1) as well as in other studies (Dacre and Pinto, 2020).

Blender et al. (2015) suggest the application of FPPs to model the clustering behaviour, with IETs following a Mittag-Leffler instead of an exponential distribution. However, since software such as the Mittag-LeffleR R-package (Gill and Straka, 2017) was not available at that time, they instead used the simpler Weibull distribution and identified the shape parameter k of the Weibull distribution for $k \leqslant 1$ with the tail parameter β of the Mittag-Leffler distribution. For $k < 1$ the Weibull distribution is heavy-tailed in the sense that the distribution tails decrease slower than those of the exponential distribution, but for every k the expected value and variance exist. The CPP model, i.e. mixture distributed IETs with the Dirac measure at point zero and the exponential distribution as parts, was not considered as another explanation of the serial clustering.

Given these findings, we apply our generalised model to the occurrences of mid-latitude cyclones in Europe and compare it with its special cases PP, FPP and CPP. This allows us to determine whether the IETs can be better described by the exponential, by the Mittag-Leffler distribution, by a mixture distribution with an exponential component, or by a combination of both.

7.2 Data and Method

To ensure comparability with the results of Blender et al. (2015), we use the same data source and method for identifying extreme cyclones. However, there are some differences: While Blender et al. (2015) used the ERA-Interim reanalysis dataset, which includes meteorological data from 1979 onwards, we use its successor, ERA5, which begins in 1940 and offers a higher horizontal resolution. Further differences are described in the following sections.

7.2.1 ERA5 reanalysis data

The ERA5 reanalysis data (Hersbach et al., 2023) are provided by the European Centre for Medium-Range Weather Forecast (ECMWF). *Reanalysis* means that model data are combined with observations from across the world in a complete and consistent dataset (Copernicus Climate Change Service, Climate Data Store, 2023). They have the great advantage that we do not have to handle missing or wrong data.

The position of cyclones in the northern hemisphere are typically identified by the maxima of the relative vorticity or the minima of mean sea-level pressure in a given area at a certain time (see, e.g., Neu et al., 2013). As Blender et al. (2015), we use the variable *relative vorticity* at 850 hPa pressure level. The pressure level indicates the altitude of the measurement.

7.2.2 Analysis procedure

In this analysis we included data from Winter 1940/41 to 2022/23 in 6-hourly time steps with a horizontal resolution of 1° on the North Atlantic Area (30°N - 60°N, 20°E - 80°W). The extreme cyclones are determined by the peak-over-threshold method as all cyclones exceeding the 99% quantile of the relative vorticity calculated separately at each grid point (Figure E.2 in the Appendix illustrates the corresponding thresholds). We only analyse winter data from December, January and February (DJF) due to the different climate and weather dynamics in the other seasons of the year. This is a standard approach in meteorological studies (see, e.g., Blender et al., 2015; Neu et al., 2013) and also justifies the assumption of (approximately) stationary magnitudes. A disadvantage is that we can only consider IETs within the same winter, which cannot get larger than 90 days. Therefore, we retain the last IET even if the associated extreme event no longer falls within the winter period. Otherwise, it would be very difficult to argue that all IETs follow the same distribution.

First, we compare the results across the entire region regarding the generalised model and the submodels. We use our CMmod2 estimator to fit the FCPP parameters since the estimates were more stable and reliable than those of the CMmod1 estimator in the simulation study in Chapter 6. For better comparability we use the CMmod2 estimator for the submodels as well. Note that the CMmod2 estimator for fixed $\theta = 1$ corresponds to the original CM estimator. In order to quantify the uncertainty associated with our estimates, we apply a bootstrap procedure to our data. Therefore, for each grid point we generate 100 new data samples of a data generating process with the estimates as parameter values. With these bootstrap samples we estimate the standard error. In addition, we use the bootstrap for classification: If the tail parameter β is estimated to be smaller than one in 95% or more of the bootstrap samples, then we assume $\beta < 1$ to be true for that location. Similarly, we assume $\theta < 1$ to be true if 95% or more of the sample estimates $\hat{\theta}$ are smaller than one.

In a second step, we examine three specific grid points on the exit region of the North Atlantic storm track with different parameter estimations to analyse the differences in depth. These locations are chosen such that regarding the CMmod2 estimation method Location A (in the interior of France) belongs to the CPP model, Location B (west coast of U.K.) belongs to the FCPP model, and Location C (west coast of the Netherlands) belongs to the FPP model. Figure 7.1 plots the magnitudes of the relative vorticities observed at these three locations for the last five winters.

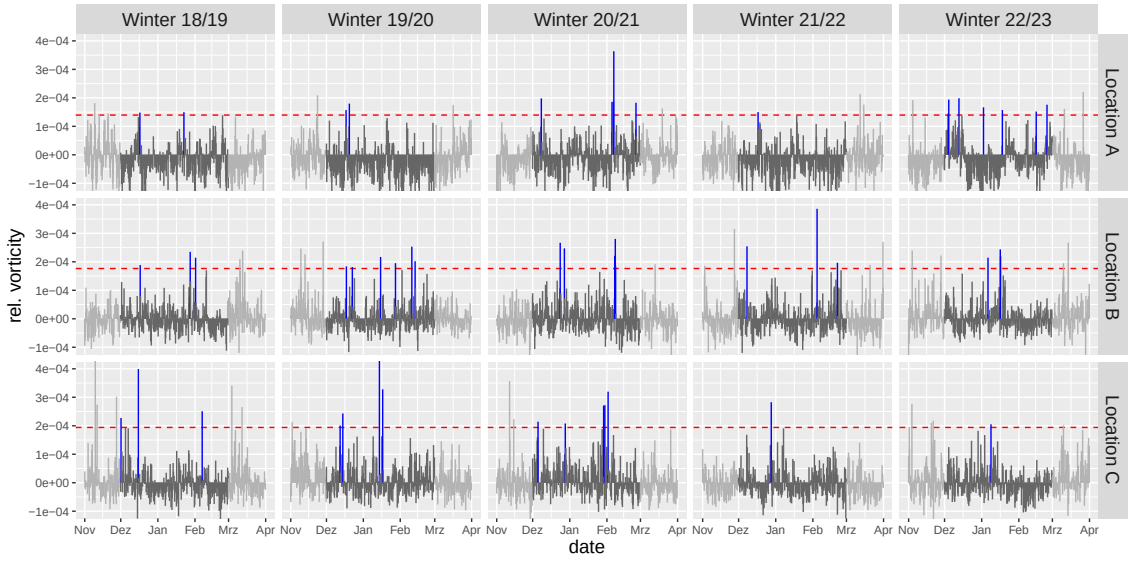

Figure 7.1: Time series of the last five winters for three locations.

The extreme events crossing the 99% quantile as threshold value show clustering behaviour at all three locations.

7.3 Results

Our results focus on the two parameters β and θ since they capture the clustering behaviour. All three models (see Figure 7.2) fit well to the general pattern that temporal clustering occurs at the exit region of storm tracks to the west of Europe, while they occur more regularly at the American east coast (Dacre and Pinto, 2020). Moreover, the mountains in southern Europe seem to have a large influence on the IET distribution, as we find the strongest clustering behaviour regarding both parameters β and θ there. Comparing the results for the FCPP and the FPP model (see Figure 7.2, middle and right column), we see that the tail parameter β is generally estimated larger in the FCPP. Except for the storm track exit region over the north Atlantic and European mountain areas the tail parameter β is estimated mostly close to one, while the FPP model suggests lower values of β. The reason for the difference between the results for the FCPP and FPP model is that the FCPP is more flexible and explains the temporal clustering via both effects, the mixture component and heavy tails, so that a larger value of β is compensated in the FCPP model by an extremal index θ less than 1 in these regions. When comparing the results of the FCPP model to the CPP model (see Figure 7.2, middle and left column), we observe fewer differences concerning the extremal index θ. This is due to the tail parameter β being estimated to be close to one at most locations, which puts us in the special case of the CPP. In regions where β is estimated to be clearly less than one, the estimate

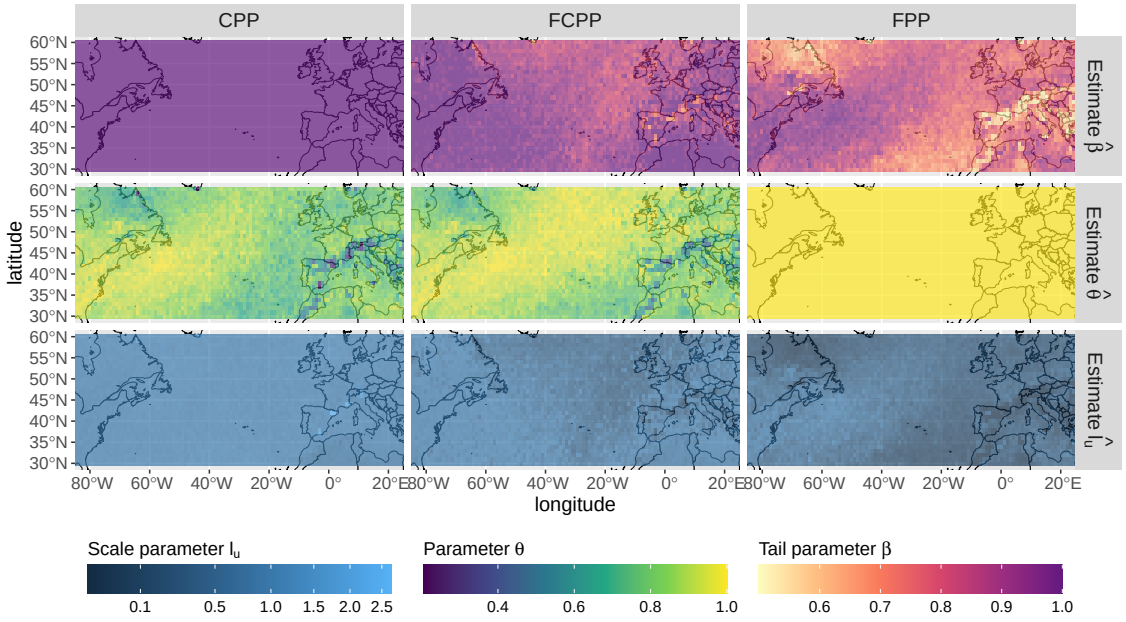

Figure 7.2: Estimations of tail parameter β, extremal index θ and scale parameter l_u in case of the CPP (left), the FCPP (middle) and FPP (right). The tail parameter is equal to 1 in the CPP, while this applies to θ in the FPP.

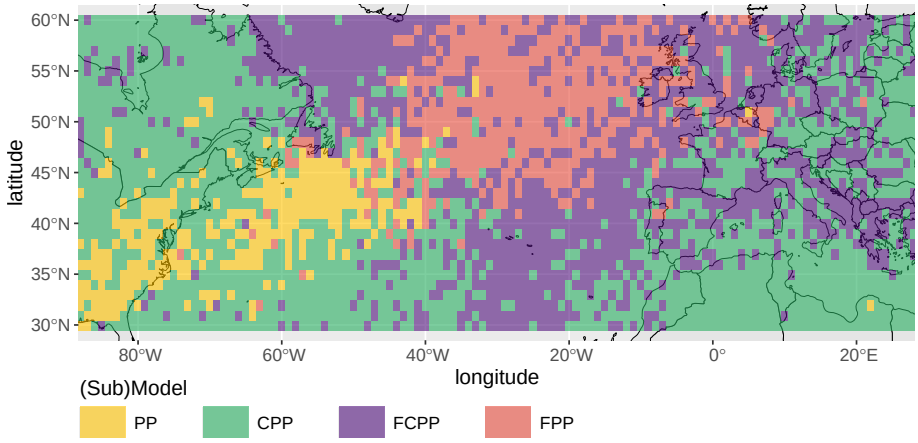

Figure 7.3: Bootstrap classification of extreme mid-latitude cyclones into the four (sub)models.

of θ is higher than in the CPP model. The results of the bootstrap classification (see Figure 7.3) underline these results.

Now we have a closer look at the three specific locations. Figure 7.4 shows histograms and QQ plots of the IETs with densities and theoretical quantiles, respectively, fitted using the exponential distribution. We observe that all three locations are poorly described by an exponential distribution. Furthermore, the locations differ from each other. Compared to the exponential distribution, all three locations have a higher probability mass on very small values, although this mass is spread over a larger range at Location C than at Locations A and B. Additionally, the figure suggests that the right tail of the distributions for Locations B and C is heavier than what the exponential distribution would suggest.

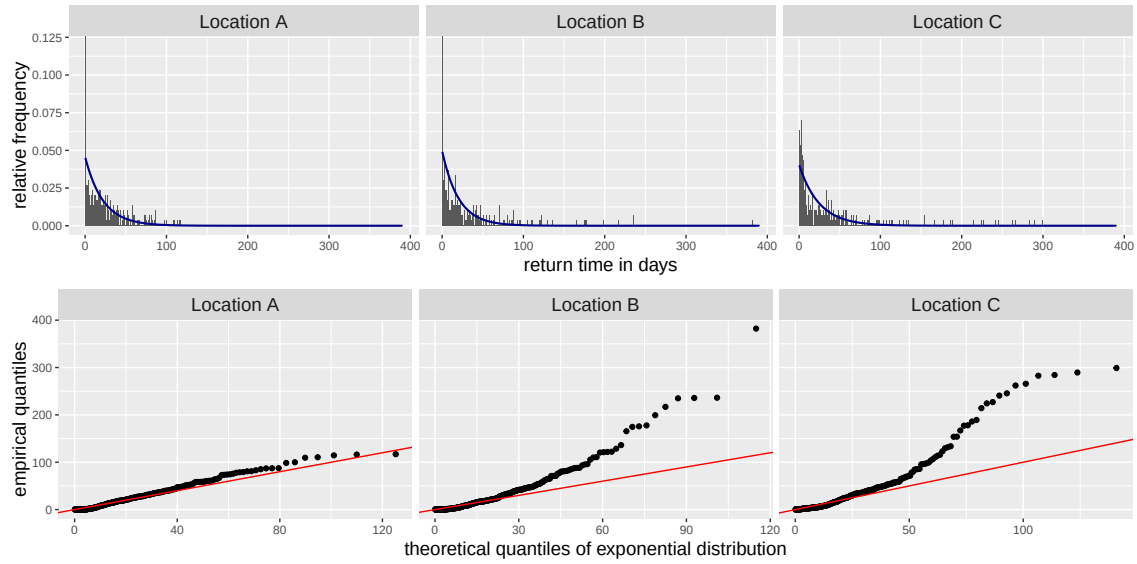

Figure 7.4: Histogram with bar width of one day (top) and QQ plots (bottom) of the IETs of location A, B and C with densities and theoretical quantiles, respectively, fitted using the exponential distribution.

The three illustrative examples have been selected in such a manner that Location A corresponds to the CPP submodel, Location C to the FPP submodel, and Location B to the general FCPP model according to CMmod2. Table 7.1 shows the parameter estimations of the FCPP model for Locations A, B and C along with estimated standard errors by a parametric bootstrap procedure. The top row of Figure 7.5 shows the histograms of the three locations with a bar width of one day. The fitted densities of the three models CPP, FCPP and FPP are included for comparison. We can see the differences between the locations concerning the parameter estimation of β and θ. At Locations B and C, β is estimated to be less than one. The right tail of the distribution is apparently heavier there than at Location A, where β is estimated to be equal to one. At Locations A and B, θ is estimated to be less than

Table 7.1: Parameter estimations for the three Locations A, B and C using CMmod2. Marked in grey are the submodels that are equal to the estimation values of FCPP. Second rows report the estimated standard errors by a parametric bootstrap procedure.

	Location A			Location B			Location C		
	β	θ	l_u	β	θ	l_u	β	θ	l_u
FCPP	1.00	0.83	0.96	0.88	0.84	0.49	0.77	1.00	0.23
	(.020)	(.031)	(.106)	(.035)	(.030)	(.115)	(.025)	(.010)	(.062)
CPP	1.00	0.83	0.96	1.00	0.76	1.00	1.00	0.74	1.23
		(.033)	(.026)		(.034)	(.030)		(.032)	(.042)
FPP	0.81	1.00	0.26	0.72	1.00	0.12	0.77	1.00	0.23
	(.027)		(.062)	(.028)		(.040)	(.026)		(.060)

one. This is due to the high number of very small IETs that do not exceed one day. In the middle row of Figure 7.5, QQ plots are shown. Especially the upper quantiles of Location C, adjusted for the FCPP, seem to fit worse than those of the CPP. To answer the question of why the FCPP did not estimate $\beta = 1$ and $\theta < 1$ in this case, we examine the same QQ plots and focus on the small IETs (see Figure 7.5, bottom row). Here, we can see that the FCPP best describes the empirical quantiles at all three locations. It is noteworthy that over 90% of the IETs at all three locations are shorter than 100 days. Thus, the CMmod estimation method gives more weight to smaller observations in the estimation.

We now examine the impact of the various models on the probabilities associated with the IETs using Location B as illustrative example. Table 7.2 presents the probabilities of the IETs not exceeding 1, 2, 7, 30, 100, and 365 days, respectively.

Table 7.2: Estimated probabilities (in percentage) of IETs not exceeding 1, 2, 7, 30, 100, and 365 days at Location B for the four (sub)models.

t (days)	1	2	7	30	100	365
FCPP	17.67	18.80	23.60	39.48	65.86	91.94
CPP	24.58	25.15	27.96	39.58	64.62	95.34
FPP	4.86	7.85	17.96	41.32	67.38	87.78
PP	1.23	2.45	8.32	31.08	71.08	98.92

Since we do not know the truth, we cannot conclusively determine which model is the best. However, we observe great differences between the estimates, particularly for the very short times. For intermediate IETs (100 days), the estimates from the four models are more similar, but they diverge again for even longer times. The compound models, CPP and FCPP, in particular predict a high probability, around 19% and 25% respectively, of another extreme event under two days. In contrast, the FPP model estimates this probability to be below 8%. Only at less than 7 days does the probability increase clearly to 18%, even with the FPP.

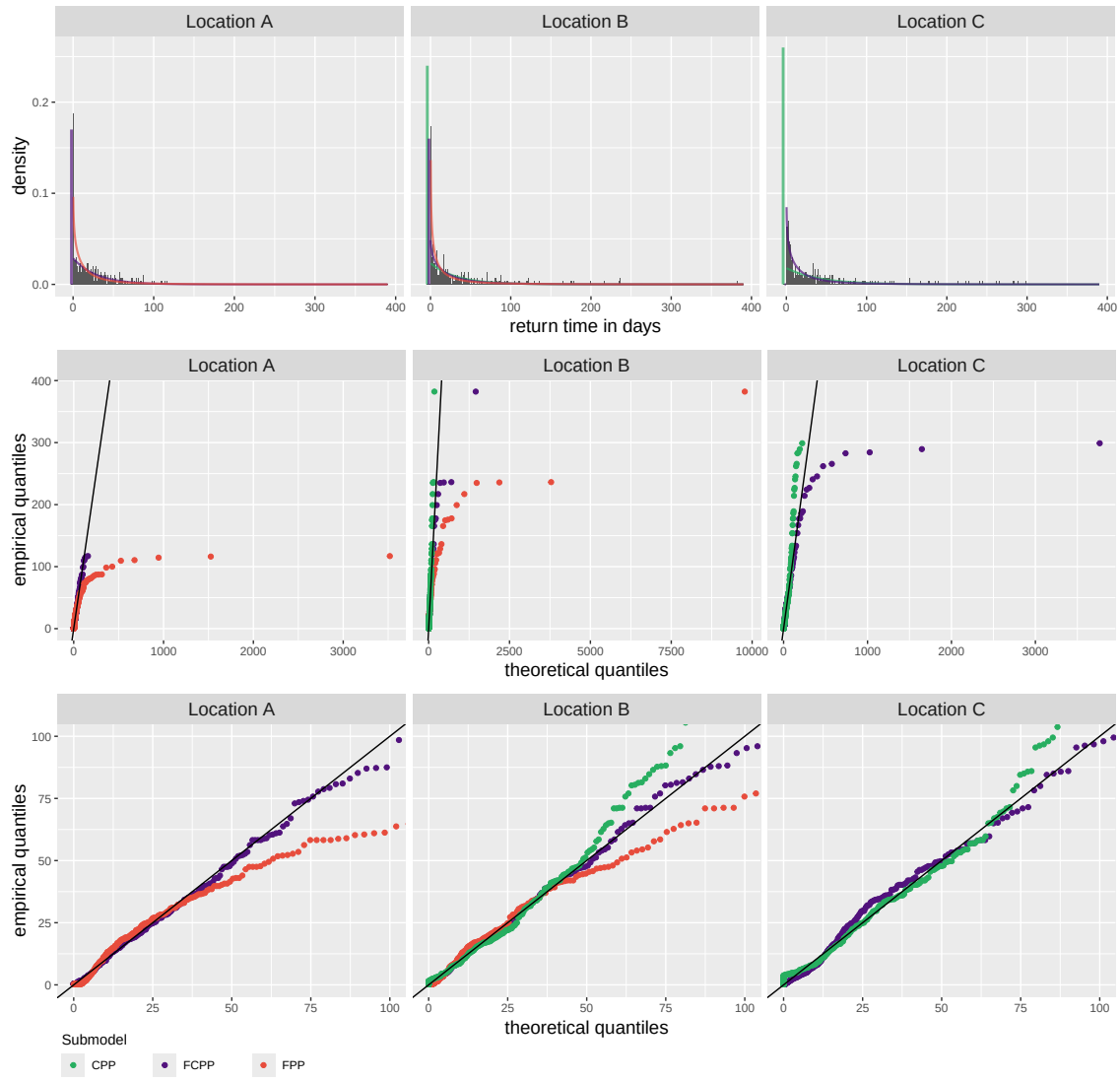

Figure 7.5: Histogram (top) and QQ plots (middle, bottom) of the IETs of Locations A, B and C with densities and theoretical quantiles, respectively, fitted using the CPP, the FCPP and FPP model. At the bottom row the QQ plot focuses on small IETs up to 100 days.

Lastly, we consider the conditional probability of the time until the next exceedance, given that t_0 time has already elapsed since the last exceedance:

$$\mathbb{P}(T(u) > x + t_0 \mid T(u) > t_0) = \frac{\mathbb{P}(T(u) > x + t_0)}{\mathbb{P}(T(u) > t_0)}. \tag{7.1}$$

While the distributions in the PP and CPP are independent of the time t_0, in the FPP and FCPP models, the probability of short remaining IETs decreases as t_0 increases, because the Mittag-Leffler distribution is not memoryless. Figure 7.6 illustrates this behaviour for Location B.

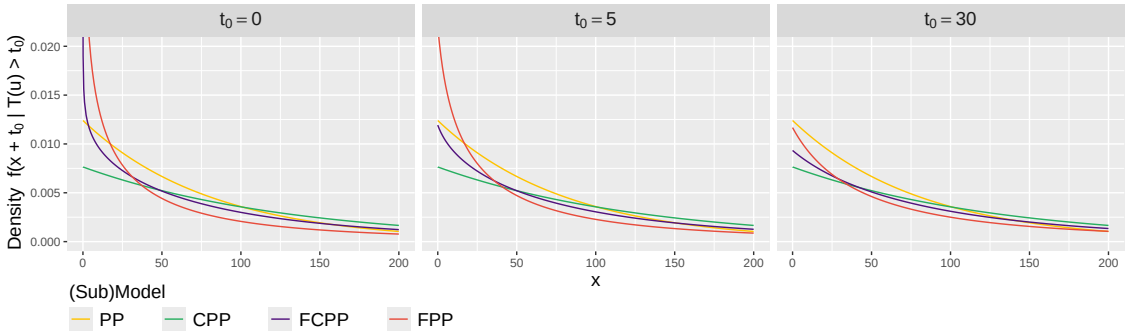

Figure 7.6: Conditional probability for all four (sub)models regarding three starting points $t_0 = 0, 5, 30$ for Location B.

7.4 Conclusions and discussion

In this chapter, we applied the FCPP model and its three submodels, PP, CPP, and FPP, to the real-world dataset of extreme mid-latitude cyclones. Our aim was to investigate the hypothesis proposed by Blender et al. (2015), who suggest that the temporal clustering of cyclones can be explained by heavy-tailed IET distributions, specifically proposing the Mittag-Leffler distribution. We selected three specific locations, each showing different temporal clustering, for further analysis. By doing so, we aimed to gain a deeper understanding of varying cluster behaviours. Additionally, it revealed further properties of our estimator, CMmod2.

In general, our results support the hypothesis of Blender et al. (2015) that the deviation of the dispersion from zero in the exit region of the North Atlantic storm track can, at least partially, be described by the Mittag-Leffler distribution. Furthermore, our findings suggest that the use of the generalised mixture distribution model could provide even better fits in some regions. For the area under study, we identified regions for all four models using a bootstrap-based classification method.

A detailed examination of the three specific locations, chosen for their estimated parameter values, allowed us to visually observe, via histograms and QQ plots, that the distributions differed from one another. For Location B, both the tail parameter β and the extremal index θ were estimated to be smaller than one, meaning it did not belong to either the CPP or FPP submodel. We used this location to compare the different model fits. The calculation of various probabilities revealed that these strongly depend on the chosen model. Therefore, it is not irrelevant which model is used to describe cluster behaviour. This underscores the utility of our generalised FCPP model.

However, the examination of the QQ plots also revealed that, although the FCPP model fitted with our estimation method accurately described the small IETs, the right tail of the distribution was occasionally too heavy. This could be due to the fact that the CMmod2 approach places more weight on small IETs. This makes sense to some extent, as high quantiles, due to their heavy tails, exhibit high volatility. Alternatively, the Mittag-Leffler distribution, which has no finite moments for $\beta < 1$, might be too heavy-tailed for this data. The Weibull distribution, which Blender et al. (2015) used to replace the Mittag-Leffler distribution due to its unavailability, is also heavy-tailed when compared to the exponential distribution, but its moments are always finite. Therefore, the Weibull distribution, as a continuous component of a mixture model, could potentially provide better fits for long IETs as well. However, there is no underlying asymptotic theory for this model.

In summary, the application of the FCPP model and its submodels to extreme mid-latitude cyclones has provided valuable insights into the temporal clustering of these events and has demonstrated the utility of the generalised mixture model for describing such clustering behaviour. However, the potential for model improvement, particularly with respect to long IETs, warrants further investigation in future studies.

8 Summary

This dissertation addresses the modelling of interarrival times of temporally clustered extreme events. Thereby, an event is classified as extreme if it exceeds a chosen threshold u, according to the peaks-over-threshold approach. Accordingly, the time intervals between two extremes are termed interexceedance times.

The starting point of this work is the consideration of two competing models within the framework of uncoupled marked point processes, both capable of explaining temporal clustering of extremes. The first model assumes stationary magnitudes occurring at deterministic, equidistant time points. The second model, in contrast, involves independent magnitudes separated by random, heavy-tailed waiting times. This thesis proposes a generalisation of these models by combining stationary magnitudes with random, positive waiting times. This leads to a setting in which the interexceedance times, when the threshold is sufficiently high, are approximately governed by a mixture distribution consisting of a Dirac measure at zero and a Mittag-Leffler distribution. Under additional assumptions, the resulting model corresponds to a fractional compound Poisson process (FCPP) with the cluster size as marks. The two initial models, namely the compound Poisson process (CPP) and the fractional Poisson process (FPP), emerging as special cases within this general framework. Interexceedance times that asymptotivally follow such a mixture distribution can be interpreted in the following way: intracluster times, i.e., intervals between extremes within the same cluster, are very small and converge to zero, while intercluster times, i.e., intervals between extremes in different clusters, follow the Mittag-Leffler distribution. The derivation of this generalised model is presented in the first main chapter of this thesis.

The mixture distribution of the FCPP model is characterised by three parameters: the tail parameter $\beta \in (0, 1]$, the extremal index $\theta \in (0, 1]$ and a scale parameter $\sigma_u > 0$. To fit this distribution to observed data, a suitable estimation method is required. This challenge is addressed in Chapters 4, 5, and 6 of this thesis: At the heart of the problem lies the difficulty of finding an appropriate estimator. On the one hand, the Mittag-Leffler distribution exhibits heavy tails; on the other hand, the theoretical mixture distribution places positive probability mass at zero, whereas the observed interexceedance times are strictly positive and serve as the empirical basis for estimation. As a solution, in Chapter 4 two variants of minimum distance estimators based on modified Cramér–von Mises distances are proposed: CMmod1 and CMmod2. The following two chapters develop and analyse the theoretical and

practical properties of these estimators. Under the additional assumptions of a compact parameter space and a stochastically independent sample, both estimators are shown to be weakly consistent. A large-scale simulation study further evaluates their finite-sample performance. The results indicate that CMmod2 performs particularly robustly with respect to the threshold choice, i.e., the number of exceedances, and across various parameter constellations. CMmod1 performs similarly well in many settings and even slightly better in some cases. However, it shows greater sensitivity to the number of exceedances and exhibits noticeably poorer performance for small values of β and θ. Moreover, both estimators were evaluated in the two submodels (CPP and FPP) and compared to established methods specific to those frameworks. Encouragingly, CMmod1 and CMmod2 produce comparable, and in some scenarios even competitive results, despite the added complexity of estimating an additional parameter. This advantage, however, comes at the cost of significantly higher computational time. Overall, the analysis suggests that applying the FCPP model to temporally clustered data may eliminate the need for prior model selection.

Although the main focus of this dissertation lies on the mathematical and statistical theory of models with any temporal clustered extreme events, the underlying motivation stems from a specific practical application in meteorology and climate science: Mid-latitude cyclones along the North Atlantic storm track tend to occur in temporal clusters along the west coast of Europe, while their occurrence along the east coast of North America appears more regular. Blender et al. (2015) hypothesise that this behaviour is driven by heavy-tailed interexceedance times and propose the use of the Mittag-Leffler distribution for modelling purposes. However, the CPP, which is commonly used in extreme value theory to model clustering, was not considered in their approach. The FCPP model developed in this thesis provides a framework that unifies both perspectives, allowing for a joint consideration of heavy-tailed waiting times and temporal clustering. The final substantive chapter of this dissertation is therefore devoted to analysing these meteorological data using the proposed FCPP model. The results indicate that, in certain regions, clustering behaviour can indeed be at least partially explained by the Mittag-Leffler component. However, for all considered (sub)models, distinct spatial patterns were observed, and suitable models could be identified for each region. Moreover, the data analysis demonstrated the practical relevance of model choice by illustrating how estimated quantities, such as particularly short IETs, can differ considerably depending on the selected (sub)model.

In summary, this dissertation contributes to the more accurate and flexible modelling of temporally clustered extreme events in the peak-over-threshold framework by introducing a generalised model and presenting suitable estimation methodology.

This dissertation addresses only certain aspects of the analysis of clustered extreme events and therefore offers substantial potential for further research.

Let us first consider modelling within our CTRE framework: A key assumption of

our model is that the underlying data follow an uncoupled marked point process, i.e. magnitudes and waiting times are assumed to be independent. This assumption appears implausible in many real-world applications. Thus, studying and deriving limit theorems for a coupled marked point process would constitute an interesting continuation of this work. In particular, within the context of meteorology and climate science, the assumption of stationary magnitudes deserves critical discussion. Future work should examine how our model could be applied or extended to incorporate seasonality and trends. Another line of research could address spatial extensions. In the data analysis about mid-latitude cyclones, we treated each grid point independently. Incorporating spatial structure or jointly analysing homogeneous or neighbouring locations could improve estimation precision and reduce variance. Here, the inclusion of covariates would again play a role. Beyond our framework, there are additional stochastic processes capable of describing temporal clustering, such as Hawkes processes or log-Gaussian Cox processes. A comparison with our model would be highly insightful.

Within the developed model, further investigation of the estimation method is also of interest. In this work, we considered two variants of a minimum distance estimator based on the Cramér–von-Mises criterion. The data analysis revealed that this method places relatively high weight on the fit to short interexceedance times. Depending on the application, this may not always be desirable. Therefore, future studies could explore replacing the Cramér–von-Mises distance with alternative criterion functions. Moreover, the construction of other estimation methods, such as those based on maximum likelihood or quantiles, would be a promising direction. This would also allow for a fair comparison of different approaches.

Bibliography

Alert Air Worldwide. (2007). *European winter storm Kyrill* [Accessed 05 June 2025]. https://alert.air-worldwide.com/extratropical-cyclone/2007/european-winter-storm-kyrill/

Alert Air Worldwide. (2022). *Storms Dudley and Eunice* [Accessed 05 June 2025]. https://alert.air-worldwide.com/extratropical-cyclone/2022/storms-dudley-and-eunice/

Balkema, A., & De Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, *2*. https://doi.org/10.1214/aop/1176996548

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004). *Statistics of extremes - theory and applications*. John Wiley & Sons.

Bingham, N. H., Goldie, C. M., & Teugels, J. L. (1987). *Regular variation*. Cambridge University Press.

Blender, R., Raible, C. C., & Lunkeit, F. (2015). Non-exponential return time distributions for vorticity extremes explained by fractional Poisson processes. *Quarterly Journal of the Royal Meteorological Society*, *141*, 249–257. https://doi.org/10.1002/qj.2354

Borovkov, A. A. (1987). *Compound renewal processes*. Cambridge University Press. https://doi.org/10.1017/9781009093965

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*(5), 1190–1208. https://doi.org/10.1137/0916069

Cahoy, D. O. (2013). Estimation of Mittag-Leffler parameters. *Communications in Statistics - Simulation and Computation*, *42*(2), 303–315. https://doi.org/10.1080/03610918.2011.640094

Cahoy, D. O., Uchaikin, V., & Woyczynski, W. (2010). Parameter estimation for fractional Poisson processes. *Journal of Statistical Planning and Inference*, *140*, 3106–3120. https://doi.org/10.1016/j.jspi.2010.04.016

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer-Verlag. https://doi.org/10.1007/978-1-4471-3675-0

Copernicus Climate Change Service, Climate Data Store. (2023). *Era5 hourly data on pressure levels from 1940 to present* [Accessed 20 November 2024]. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.bd0915c6

Cramér, H. (1928). On the composition of elementary errors: Second paper: Statistical applications. *The Annals of Probability*, *1*, 141–180. https://doi.org/10.1080/03461238.1928.10416872

Dacre, H. F., & Pinto, J. G. (2020). Serial clustering of extratropical cyclones: A review of where, when and why it occurs. *npj Climate and Atmospheric Science*, *3*(48). https://doi.org/10.1038/s41612-020-00152-9

de Haan, L., & Ferreira, A. (2006). *Extreme value theory: An introduction*. NY: Springer-Verlag.

Drossos, C. A., & Philippou, A. N. (1980). A note on minimum distance estimates. *Annals of the Institute of Statistical Mathematics*, *32*, 121–123. https://doi.org/10.1007/BF02480318

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer-Verlag.

Ferreira, M. (2018). Heuristic tools for the estimation of the extremal index: A comparison of methods. *REVSTAT – Statistical Journal*, *16*(1), 115–136. https://doi.org/10.57805/revstat.v16i1.235

Ferreira, M. (2024). Clustering of extreme values: Estimation and application. *ASta Advances in Statistical Analysis*, *108*, 101–125. https://doi.org/10.1007/s10182-023-00474-y

Ferro, C., & Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*, 545–556. https://doi.org/10.1111/1467-9868.00401

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(2), 180–190. https://doi.org/10.1017/S0305004100015681

Foss, S., Korshunov, D., & Zachary, S. (2013). *An introduction to heavy-tailed and subexponential distributions*. Springer.

Gill, G., & Straka, P. (2017). *MittagLeffler: Using the Mittag-Leffler distributions in R*. https://strakaps.github.io/MittagLeffleR

Gnedenko, B. V. (1983). On limit theorems for a random number of random variables. In J. V. Prokhorov & K. Itô (Eds.), *Probability theory and mathematical statistics* (pp. 167–176). Springer Berlin Heidelberg. https://doi.org/10.1007/BFb0072914

Gnedenko, B. W. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, *44*(3), 423–453. http://www.jstor.org/stable/1968974

Gut, A., & Hüsler, J. (1999). Extreme shock models. *Extremes*, *2*, 295–307. https://doi.org/10.1023/A:1009959004020

Haubold, H. J., Mathai, A. M., & Saxena, R. K. (2011). Mittag-Leffler functions and their applications. *Journal of Applied Mathematics*, *2011*, 1–51. https://doi.org/10.1155/2011/298628

Hees, K., Nayak, S., & Straka, P. (2021). Statistical inference for inter-arrival times of extreme events in bursty time series. *Computational Statistics & Data Analysis*, *155*, 107096. https://doi.org/10.1016/j.csda.2020.107096

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., & Thépaut, J.-N. (2023). *Era5 hourly data on pressure levels from 1940 to present* [Accessed 21 July 2023]. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.bd0915c6

Holešovský, J., & Fusek, M. (2020). Estimation of the extremal index using censored distributions. *Extremes*, *23*(2), 197–213. https://doi.org/10.1007/s10687-020-00374-3

Holešovský, J., & Fusek, M. (2022). Improved interexceedance-times-based estimator of the extremal index using truncated distribution. *Extremes*, *25*, 695–720. https://doi.org/10.1007/s10687-022-00444-8

Hsing, T., Hüsler, J., & Leadbetter, M. (1988). On the exceedance point process for stationary sequence. *Probability Theory and Related Fields*, *78*, 97–112. https://doi.org/10.1007/BF00718038

Kozubowski, T. J. (2001). Fractional moment estimation of Linnik and Mittag-Leffler parameters. *Mathematical and Computer Modelling*, *34*(9), 1023–1035. https://doi.org/10.1016/S0895-7177(01)00115-7

Laskin, N. (2003). Fractional Poisson process. *Communications in Nonlinear Science and Numerical Simulation*, *8*(3-4), 201–213. https://doi.org/10.1016/S1007-5704(03)00037-6

Laub, P. J., Lee, Y., & Taimre, T. (2021). *The elements of Hawkes processes*. Springer.

Leadbetter, M. R., Lindgren, G., & Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer.

Mailier, P. J., Stephenson, D. B., Ferro, C. A. T., & Hodges, K. I. (2006). Serial clustering of extratropical cyclones. *Monthly Weather Review*, *134*, 2224–2240. https://doi.org/10.1175/MWR3160.1

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, *18*(3), 1269–1283. https://doi.org/10.1214/aop/1176990746

Mathieu, C. (2025). *R-package: FCPP*. https://github.com/CMeschede/FCPP

Mathieu, C., Hees, K., & Fried, R. (2025). Models for temporal clustering of extreme events with applications to mid-latitude winter cyclones. *arXiv*. https://doi.org/10.48550/arXiv.2308.14625

Meerschaert, M. M., & Sikorskii, A. (2012). *Stochastic models for fractional calculus*. De Gruyter.

Meschede, C. (2018). *Quantifizierung von Risiken extremer Ereignisse, die stoßweise und in Clustern auftreten*. TU Dortmund University.

Mittag-Leffler, G. M. (1903a). Sur la nouvelle fonction $E_\alpha(x)$. *Comptes Rendus de l'Académie des Sciences*, *137*, 554–558.

Mittag-Leffler, G. M. (1903b). Une generalisation de l'integrale de laplace-abel. *Comptes Rendus de l'Académie des Sciences Série II*, *137*, 537–539.

Møller, J., Syversveen, A. R., & Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, *25*(3), 451–482. https://doi.org/https://doi.org/10.1111/1467-9469.00115

Neu, U., Akperov, M. G., N., Bellenbaum, Benestad, R. S., Blender, R., R., C., Cocozza, A., Dacre, H. F., Feng, Y., Fraedrich, K., Grieger, J., Gulev, S., Hanley, J., Hewson, T., Inatsu, M., Keay, K., Kew, S. F., Kindem, I., Leckebusch, G. C., ... Wernli, H. (2013). IMILAST a community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bulletin of the American Meteorological Society*, *94*, 529–547. https://doi.org/10.1175/BAMS-D-11-00154.1

Northrop, p. J., & Christodoulides, C. (2023). *exdex: Estimation of the extremal index* [R package version 1.2.3]. https://CRAN.R-project.org/package=exdex

O'Brien, G. L. (1987). Extreme values for stationary and markov sequences. *The Annals of Probability*, *15*(1), 281–291. http://www.jstor.org/stable/2243956

Ötztürk, Ö., & Hettmansperger, T. P. (1997). Generalised wighted Cramér-von Mises distance estimators. *Biometrika*, *84*(2), 283–294. https://doi.org/10.1093/biomet/84.2.283

Parr, W. C. (1981). Minimum distance estimation: A bibliography. *Communications in Statistics - Theory and Methods*, *10*(12), 1205–1224. https://doi.org/10.1080/03610928108828104

Parr, W. C., & De Wet, T. (1981). On Cramér-von Mises-norm parameter estimation. *Communications in Statistics - Theory and Methods*, *10*(12), 1149–1166. https://doi.org/10.1080/03610928108828100

Parr, W. C., & Schucany, W. R. (1980). Mnimum distance and robust estimation. *Journal of the American Statistical Association*, *75*(371), 616–624. https://doi.org/10.2307/2287658

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*(1), 119–131. https://doi.org/10.1214/aos/1176343003

Pillai, R. N. (1990). On Mittag-Leffler functions and related distributions. *Annals of the Institute of Statistical Mathematics*, *42*(1), 157–161. https://doi.org/10.1007/BF00050786

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Reiss, D.-R. (2012). *A course on point processes*. Springer.

Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*. Springer.

Reynkens, T., & Verbelen, R. (2020). *Reins: Functions from "Reinsurance: Actuarial and statistical aspects"* [R package version 1.0.10]. https://CRAN.R-project.org/package=ReIns

Richter, W. (1965). Limit theorems for sequences of random variables with sequences of random indeces. *Theory of Probability & Its Applications*, *10*(1), 74–84. https://doi.org/10.1137/1110007

Robert, C. Y. (2009). Inference for the limiting cluster size distribution of extreme values. *The Annals of Statistics*, *37*(1), 271–310. https://doi.org/10.1214/07-AOS551

Samorodnitsky, G., & Taqqu, M. S. (1994). *Stable non-Gaussian random processes: Stochastic models with infinite variance*. Taylor & Francis.

Segers, J. (2002). *Extreme events: Dealing with dependence*. EURANDOM report 2002-036.

Seneta, E. (1976). *Regularly varying functions*. Springer.

Süveges, M. (2007). Likelihood estimation of the extremal index. *Extremes*, *10*, 41–55. https://doi.org/10.1007/s10687-007-0034-2

Süveges, M., & Davidson, A. C. (2010). Model misspecification in peaks over threshold analysis. *Ann. Appl. Stat*, *4*(1), 203–221. https://doi.org/10.1214/09-AOAS292

van der Vaart, A. (1998). M- and Z-estimators. In *Asymptotic statistics* (pp. 41–84). Cambridge University Press. https://doi.org/10.1017/CBO9780511802256.006

von Mises, R. (1931). *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*. Leipzig: Deuticke.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wolfowitz, J. (1954). Estimation by the the minimum distance method in nonparametric stochastic difference equations. *The Annals of Mathematical Statistics*, *25*(2), 203–217.

Wolfowitz, J. (1957). The minimum distance method. *The Annals of Mathematical Statistics*, *28*(1), 75–88.

Wuertz, D., Maechler, M., & core team members., R. (2016). *stabledist: Stable distribution functions* [R package version 0.7-1]. https://CRAN.R-project.org/package=stabledist

Zorich, V. A. (2016). *Mathematical analysis II* (2nd ed.). Springer.

A Appendix to Chapter 3

This chapter presents the proof of the following theorem from Subsection 3.2.2. The proof was originally provided by the author of this dissertation as part of her Master's thesis (Meschede, 2018). It is based on the proof by Ferro and Segers (2003) and has been expanded with additional details to improve accessibility for undergraduate students. For the present work, the proof has been translated into English and adapted to the notation used throughout this thesis.

Theorem A.1. *Let $T_{(\beta,\lambda)}$ be a random variable with mixture distribution*

$$(1-\theta)\,\delta_0 + \theta\, Exp(\theta\lambda), \tag{A.1}$$

where δ_0 is the Dirac measure at point zero, $Exp(\theta\lambda)$ is the exponential distribution with rate $\theta\lambda$, and $\theta \in (0,1]$ and $\lambda > 0$.

Let $(X_n)_{n\in\mathbb{N}}$ be a strictly stationary process of magnitudes in the max-domain of attraction of a GEV distribution with extremal index $\theta \in (0,1]$, x_R is the right endpoint of the distribution of X_1, and $p(u) = \mathbb{P}(X_1 > u)$ is the excess probability.

Let $(W_n)_{n\in\mathbb{N}}$ be corresponding waiting times such that $W_n \equiv 1$ for all $n \in \mathbb{N}$, i.e. the events occur equidistant every time unit.

Without loss of generality, we set $X_0 > u$. The interexceedance time regarding the processes $(X_n)_{n\in\mathbb{N}}$ and a threshold u is defined as

$$T(u) = \min\{k \geqslant 1 \mid X_k > u\}.$$

it holds

$$p(u)T(u) \xrightarrow{d} T_{(\theta,1)} \quad as \quad u \to x_R. \tag{A.2}$$

From the definition of $T(u)$ we can derive

$$\mathbb{P}(T(u) > n) = \mathbb{P}(M_n \leqslant u \mid X_0 > u). \tag{A.3}$$

The theorem implies that X_1 is in the max-domain of attraction of a GEV and there is a sequence of thresholds $(u_n)_{n\in\mathbb{N}}$ such that the $D(u_n)$ condition holds as well as

$\lim_{n\to\infty} n\,p(u_n) = \nu$ and $\lim_{n\to\infty} \mathbb{P}(M_n \geqslant u_n) = \exp(-\nu\theta)$ with $\nu \in (0,\infty)$. The following proof uses a slightly stronger mixing condition than the $D(u_n)$ condition:

Condition 4. *Let $\mathcal{F}_{j,k}(u_n) = \sigma(\{X_i > u_n\} \mid j \leqslant i \leqslant k)$ be a σ-algebra generated by $\{X_i > u_n\}$, $i = j,\ldots,k$, with integers $j < k$. For all $A_1 \in \mathcal{F}_{1,l}(u_n)$, with $\mathbb{P}(A) > 0$ and $A_2 \in \mathcal{F}_{l+s,n}(u_n)$ and $1 \leqslant l \leqslant n-s$,*

$$|\mathbb{P}(A_2 \mid A_1) - \mathbb{P}(A_2)| \leqslant \alpha(n,s),$$

and $\alpha(n,s_n) \to 0$ as $n \to \infty$ for some positive integer s_n with $s_n = o(n)$.

Proof. In the following, we aim to show that $\lim_{n\to\infty} \mathbb{P}(p(u_n)T(u_n) > t) = \theta\exp(-\theta\,t)$ holds for all $t > 0$.

Let $M_n = \max\{X_1,\ldots,X_n\}$ and $M_{k,l} = \max\{X_{k+1},\ldots,X_l\}$ for $n,k,l \in \mathbb{N}$ with $k < l$. We assume that the threshold sequence $(u_n)_{n\in\mathbb{N}}$ and the sequence of positive integers $(s_n)_{n\in\mathbb{N}}$ are chosen such that Condition 4 holds, i.e., $s_n = o(n)$ and $\alpha(n,s_n) \to 0$. We can then find a sequence of positive integers $(r_n)_{n\in\mathbb{N}}$ satisfying $s_n = o(r_n)$ and $n\cdot\alpha(n,s_n) = o(r_n)$. Such a sequence could, for example, be defined as $r_n = \left\lfloor \max\{\sqrt{n\cdot s_n}\,,\; n\cdot\sqrt{\alpha(n,s_n)}\}\right\rfloor$. According to O'Brien (1987, Corollary 2.3), for any sequence $(k_n)_{n\in\mathbb{N}}$ of positive integers with $k_n \to \infty$, it holds that

$$\mathbb{P}\left(M_{k_n} \leqslant u_n\right) \sim (1-p(u_n))^{k_n\cdot\theta_n^R(u_n)}, \tag{A.4}$$

where for two sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$, the notation $a_n \sim b_n$ means $\lim_{n\to\infty}\frac{a_n}{b_n} = 1$, and where $\theta_n^R(u_n) := \mathbb{P}(M_{r_n} \leqslant u_n \mid X_0 > u_n) = \mathbb{P}(T(u_n) > r_n)$. Additionally, we can derive from O'Brien (1987) that $\theta_n^R(u_n) \to \theta$ as $n \to \infty$ (see Beirlant et al., 2004, Chapter 10).

Since by assumption $s_n = o(r_n)$, there exists a positive integer $n_0 \in \mathbb{N}$ such that $s_n < r_n$ for all $n > n_0$.Thus, using Condition 4 and stationarity, we obtain

$$\begin{aligned} &\mathbb{P}\left(M_{r_n,r_n+s_n} > u_n \mid X_0 > u_n\right) \sim \mathbb{P}\left(M_{r_n,r_n+s_n} > u_n\right) \\ &= \mathbb{P}\left(M_{s_n} > u_n\right) \leqslant s_n\,p(u_n) \xrightarrow[n\to\infty]{} 0. \end{aligned} \tag{A.5}$$

Here, we use that

$$\mathbb{P}\left(M_{s_n} > u_n\right) = \mathbb{P}\left(\bigcup_{i=1}^{s_n}\{X_i > u_n\}\right) \leqslant \sum_{i=1}^{s_n}\mathbb{P}\left(X_i > u_n\right).$$

It follows that

$$\mathbb{P}\left(M_{r_n,r_n+s_n} \leqslant u_n \mid X_0 > u_n\right) \to 1 \quad \text{as} \quad n \to \infty.$$

This means that the probability that the maximum of a sequence of random variables exceeds the threshold, conditional on a random variable r_n positions earlier exceeding the threshold, becomes negligible for large n.

Next, we define $l_n = \lfloor t/p(u_n) \rfloor$. We observe that $l_n = \lfloor t(n\,p(u_n))^{-1} n \rfloor \sim t\nu^{-1} n$. There exists a positive integer $m_0 \in \mathbb{N}$ with $m_0 \geqslant n_0$ and $l_n > r_n + s_n$ for all $n \geqslant m_0$, as $r_n = o(n)$, $s_n = o(n)$, and $\lim\limits_{n\to\infty} l_n/n = t\nu^{-1}$ as $n \to \infty$. Thus, using Condition 4, we obtain

$$\begin{aligned} &\mathbb{P}\left(M_{r_n+s_n,l_n} \leqslant u_n \mid M_{r_n} \leqslant u_n, X_0 > u_n\right) \\ &\quad \sim \mathbb{P}\left(M_{r_n+s_n,l_n} \leqslant u_n\right) = \mathbb{P}\left(M_{l_n-(r_n+s_n)} \leqslant u_n\right). \end{aligned} \tag{A.6}$$

We now set $k_n = l_n - (r_n + s_n)$. It follows that $k_n \sim t\nu^{-1} n$ and thus $k_n \to \infty$ as $n \to \infty$. Applying the results from above, namely equations (A.4) to (A.6), and using $(1 - p(u_n))^n \to \exp(-\nu)$, we obtain

$$\begin{aligned} \mathbb{P}\left(p(u_n)\,T(u_n) > t\right) &= \mathbb{P}\left(T(u_n) > l_n\right) = \mathbb{P}\left(M_{l_n} \leqslant u_n \mid X_0 > u_n\right) \\ &= \mathbb{P}\left(M_{r_n} \leqslant u_n, M_{r_n,r_n+s_n} \leqslant u_n, M_{r_n+s_n,l_n} \leqslant u_n \mid X_0 > u_n\right) \\ &\overset{(A.5)}{=} \mathbb{P}\left(M_{r_n} \leqslant u_n, M_{r_n+s_n,l_n} \leqslant u_n \mid X_0 > u_n\right) + o(1) \\ &= \mathbb{P}\left(M_{r_n+s_n,l_n} \leqslant u_n \mid M_{r_n} \leqslant u_n, X_0 > u_n\right) \mathbb{P}\left(M_{r_n} \leqslant u_n \mid X_0 > u_n\right) + o(1) \\ &\overset{(A.6)}{\sim} \mathbb{P}\left(M_{k_n} \leqslant u_n\right) \cdot \theta_n^R(u_n) + o(1) \\ &\overset{(A.4)}{\sim} (1 - p(u_n))^{k_n \theta_n^R(u_n)} \theta_n^R(u_n) + o(1) \\ &\sim \left((1 - p(u_n))^n\right)^{\theta_n^R(u_n)\,t\nu^{-1}} \theta_n^R(u_n) + o(1) \\ &\xrightarrow[n\to\infty]{} \exp(-\nu)^{\theta t \nu^{-1}} \theta = \theta \exp(-\theta t). \end{aligned}$$

Overall, we conclude that $\mathbb{P}\left(p(u_n)\,T(u_n) \leqslant t\right) \xrightarrow[n\to\infty]{} 1 - \theta \exp(-\theta t)$. □

B Appendix to Chapter 4

This chapter presents some calculations of Chapter 4.

B.1 Sum representation of the CMmod1 and CMmod2 distances

First, we show that

$$\begin{aligned}\Delta^{[\mathrm{CMmod1}]}(F_k, F^{(\beta,\theta,\sigma_u)}) &= \int_0^\infty \left(F_k(x) - F^{(\beta,\theta,\sigma_u)}(x)\right)^2 f_*^{(\beta,\theta,\sigma_u)}(x)\,\mathrm{d}x \\ &= \frac{1}{\theta k}\sum_{i=1}^{k}\left(\frac{i-\frac{1}{2}}{k} - F^{(\beta,\theta,\sigma_u)}(T_{(i)}(u))\right)^2 + \frac{1}{12\theta k^2} - \frac{1}{3}\frac{(1-\theta)^3}{\theta},\end{aligned}$$

and

$$\begin{aligned}\Delta^{[\mathrm{CMmod2}]}(F_k, F^{(\beta,\theta,\sigma_u)}) &= \frac{1}{\theta^2}\int_0^\infty \left(\max\{F_k(x), 1-\theta\} - F^{(\beta,\theta,\sigma_u)}(x)\right)^2 f_*^{(\beta,\theta,\sigma_u)}(x)\,\mathrm{d}x \\ &= \frac{1}{\theta^3 k}\sum_{i=l+1}^{k}\left(\frac{i-\frac{1}{2}}{k} - F^{(\beta,\theta,\sigma_u)}(T_{(i)}(u))\right)^2 \\ &\quad + \frac{k-l}{12\theta^3k^3} - \frac{(k(1-\theta))^3 - l^3}{3k^3\theta^3} \\ &\quad + \frac{(k(1-\theta))^2 - l^2}{k^2\theta^3}F^{(\beta,\theta,\sigma_u)}(T_{(l)}(u)) \\ &\quad - \frac{(k(1-\theta)) - l}{k\theta^3}F^{(\beta,\theta,\sigma_u)}(T_{(l)}(u))^2,\end{aligned}$$

where F_k is the empirical c.d.f. of the ordered observations $T_{(1)}(u) < \cdots < T_{(k)}(u)$ and $l := \lceil k(1-\theta)\rceil$ with $\lceil\cdot\rceil$ being the ceiling function.

Since for this cumbersome but straight forward calculations only the weighting parameter θ matters, we write F^θ for $F^{(\beta,\theta,\sigma_u)}$ and F for the $F_*^{(\beta,\theta,\sigma_u)}$.

We begin with some partial calculations (I) - (IX) where X and X_θ are random variables with c.d.f.s F and F^θ, respectively :

$$(I)\quad 1-F(t)=\frac{1-1+\theta-\theta F(t)}{\theta}=\frac{1-F^\theta(t)}{\theta}$$

$$\begin{aligned}(II)\quad \mathbb{E}\left(F(X)\mathbb{1}\{X\geqslant t\}\right)&=\mathbb{E}\left(F(X)\mid X\geqslant t\right)\mathbb{P}\left(X\geqslant t\right)\\&=\frac{1+F(t)}{2}\left(1-F(t)\right)=\frac{1-F(t)^2}{2}\end{aligned}$$

$$\begin{aligned}(III)\quad &\int F_k(x)^2\,\mathrm{d}F(x)=\mathbb{E}(F_k(X)^2)\\&=\frac{1}{k^2}\sum_{i=1}^k\sum_{j=1}^k\mathbb{E}(\mathbb{1}\{T_i(u)\leqslant X\}\mathbb{1}\{T_j(u)\leqslant X\})\\&=\frac{1}{k^2}\sum_{i=1}^k\mathbb{E}(\mathbb{1}\{T_i(u)\leqslant X\})+\frac{2}{k^2}\sum_{j<i}\mathbb{E}(\mathbb{1}\{T_i(u)\leqslant X\}\mathbb{1}\{T_j(u)\leqslant X\})\\&=\frac{1}{k^2}\sum_{i=1}^k(1-F(T_i(u)))+\frac{2}{k^2}\sum_{j<i}\mathbb{E}(\mathbb{1}\{T_{(i)}(u)\leqslant X\})\\&\overset{(I)}{=}\frac{1}{\theta}\frac{1}{k^2}\sum_{i=1}^k(1-F^\theta(T_i(u)))+\frac{2}{k^2}\sum_{i=1}^k(i-1)\cdot(1-F(T_{(i)}(u)))\\&\overset{(I)}{=}\frac{1}{\theta}\frac{1}{k^2}\sum_{i=1}^k(1-F^\theta(T_i(u)))+\frac{1}{\theta}\frac{2}{k^2}\sum_{i=1}^k(i-1)\cdot(1-F^\theta(T_{(i)}(u)))\\&=\frac{1}{\theta}\frac{2}{k^2}\sum_{i=1}^k\left(i-\frac{1}{2}\right)\cdot(1-F^\theta(T_{(i)}(u)))\\&=\frac{1}{\theta}\frac{2}{k^2}\sum_{i=1}^k\left(i-\frac{1}{2}\right)-\frac{2}{k}\sum_{i=1}^k\frac{i-\frac{1}{2}}{k}\cdot F^\theta(T_{(i)}(u))\\&=\frac{1}{\theta}-\frac{1}{\theta}\frac{2}{k}\sum_{i=1}^k\frac{i-\frac{1}{2}}{k}\cdot F^\theta(T_{(i)}(u))\end{aligned}$$

$$
\begin{aligned}
(IV)\quad & 2\int F_k(x)\cdot F^\theta(x)\,\mathrm{d}F(x) \\
&= 2\cdot\mathbb{E}(F_k(X)F^\theta(X)) \\
&= 2\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}(\mathbb{1}\{T_i(u)\leqslant X\}F^\theta(X)\mid\mathbb{1}\{T_i(u)\leqslant X\})\mathbb{P}(T_i(u)\leqslant X) \\
&= 2\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}(F^\theta(X)\mid\mathbb{1}\{T_i(u)\leqslant X\})(1-F(T_i(u))) \\
&\overset{(I)}{=} 2\frac{1}{\theta}\frac{1}{k}\sum_{i=1}^{k}\left(1-\theta+\theta\cdot\mathbb{E}(F(X)\mid\mathbb{1}\{T_i(u)\leqslant X\})\right)(1-F^\theta(T_i(u))) \\
&= 2\frac{1}{\theta}\frac{1}{k}\sum_{i=1}^{k}\left(1-\theta+\theta\cdot\frac{1+F(T_i(u))}{2}\right)(1-F^\theta(T_i(u))) \\
&= \frac{1}{\theta}\frac{1}{k}\sum_{i=1}^{k}\left(1+F^\theta(T_i(u))\right)(1-F^\theta(T_i(u))) \\
&= \frac{1}{\theta}\frac{1}{k}\sum_{i=1}^{k}\left(1-F^\theta(T_i(u))^2\right) = \frac{1}{\theta}-\frac{1}{\theta}\frac{1}{k}\sum_{i=1}^{k}F^\theta(T_i(u))^2
\end{aligned}
$$

$$
\begin{aligned}
(V)\quad \int F^\theta(x)^2\,\mathrm{d}F(x) &= \mathbb{E}(F^\theta(X)^2) \\
&= (1-\theta)^2+2\cdot\theta(1-\theta)\cdot\mathbb{E}(F(X))+\theta^2\cdot\mathbb{E}(F(X)^2) \\
&= (1-\theta)^2+2\cdot\theta(1-\theta)\cdot\frac{1}{2}+\theta^2\cdot\frac{1}{3} = \frac{1}{3}\theta^2+(1-\theta)
\end{aligned}
$$

$$
\begin{aligned}
(VI)\quad \{1-\theta>F_k(X)\} &\Leftrightarrow \left\{k(1-\theta)>\sum_{i=1}^{n}\mathbb{1}\{T_{(i)}(u)\leqslant X\}\right\} \\
&\Leftrightarrow \left\{l>\sum_{i=1}^{n}\mathbb{1}\{T_{(i)}(u)\leqslant X\}\right\} \\
&\Leftrightarrow \left\{T_{(l)}(u)>X\right\} \\
&\Rightarrow \{1-\theta\leqslant F_k(X)\}\Leftrightarrow\left\{T_{(l)}(u)\leqslant X\right\}
\end{aligned}
$$

$$
\begin{aligned}
(VII)\ \ & \mathbb{E}\left(F_k(X)^2 \mathbb{1}\{X \geqslant T_{(l)}(u)\}\right) \\
&= \frac{1}{k^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left(\mathbb{1}\{T_{(i)}(u) \leqslant X\}\mathbb{1}\{T_{(j)}(u) \leqslant X\}\mathbb{1}\{T_{(l)}(u) \leqslant X\}\right) \\
&= \frac{1}{k^2}\sum_{i=1}^{n}\mathbb{E}\left(\mathbb{1}\{T_{(i)}(u) \leqslant X\}\mathbb{1}\{T_{(l)}(u) \leqslant X\}\right) \\
&\quad + \frac{2}{k^2}\sum_{i>j}^{n}\mathbb{E}\left(\mathbb{1}\{T_{(i)}(u) \leqslant X\}\mathbb{1}\{T_{(l)}(u) \leqslant X\}\right) \\
&= \frac{2}{k^2}\sum_{i=1}^{n}\left(i-\frac{1}{2}\right)\mathbb{E}\left(\mathbb{1}\{T_{(i)}(u) \leqslant X\} \cdot \mathbb{1}\{T_{(l)}(u) \leqslant X\}\right) \\
&= \frac{2}{k^2}\sum_{i=1}^{l}\left(i-\frac{1}{2}\right)\mathbb{E}\left(\mathbb{1}\{T_{(l)}(u) \leqslant X\}\right) \\
&\quad + \frac{2}{k^2}\sum_{i=l+1}^{k}\left(i-\frac{1}{2}\right)\mathbb{E}\left(\mathbb{1}\{T_{(i)}(u) \leqslant X\}\right) \\
&= \frac{2}{k^2}\sum_{i=1}^{l}\left(i-\frac{1}{2}\right)\left(1-F\left(T_{(l)}(u)\right)\right) + \frac{2}{k^2}\sum_{i=l+1}^{k}\left(i-\frac{1}{2}\right)\left(1-F(T_{(i)}(u))\right) \\
&\overset{(I)}{=} \frac{2}{k^2}\frac{1}{\theta}\sum_{i=1}^{l}\left(i-\frac{1}{2}\right)\left(1-F^{\theta}\left(T_{(l)}(u)\right)\right) \\
&\quad + \frac{2}{k^2}\frac{1}{\theta}\sum_{i=l+1}^{k}\left(i-\frac{1}{2}\right)\left(1-F^{\theta}(T_{(i)}(u))\right) \\
&= \frac{2}{k^2}\frac{1}{\theta}\sum_{i=1}^{k}\left(i-\frac{1}{2}\right) - \frac{2}{k^2}\frac{1}{\theta}\sum_{i=1}^{l}\left(i-\frac{1}{2}\right)F^{\theta}\left(T_{(l)}(u)\right) \\
&\quad - \frac{2}{k^2}\frac{1}{\theta}\sum_{i=l+1}^{k}\left(i-\frac{1}{2}\right)F^{\theta}\left(T_{(i)}(u)\right) \\
&= \frac{1}{\theta} - \frac{l^2}{k^2\theta}F^{\theta}\left(T_{(l)}(u)\right) - \frac{2}{k^2}\frac{1}{\theta}\sum_{i=l+1}^{k}\left(i-\frac{1}{2}\right)F^{\theta}\left(T_{(i)}(u)\right)
\end{aligned}
$$

$$
\begin{aligned}
(VIII)\;\; & 2\,\mathbb{E}\left(F_k(X)F^{\theta}(X)\mathbb{1}\{T_{(l)}(u)\leqslant X\}\right)\\
&= \frac{2}{k}\sum_{i=1}^{n}\mathbb{E}\left(F^{\theta}(X)\mathbb{1}\{T_i(u)\leqslant X\}\mathbb{1}\{T_{(l)}(u)\leqslant X\}\right)\\
&= \frac{2}{k}\sum_{i=1}^{l}\mathbb{E}\left(F^{\theta}(X)\mathbb{1}\{T_{(l)}(u)\leqslant X\}\right)+\frac{2}{k}\sum_{i=l+1}^{k}\mathbb{E}\left(F^{\theta}(X)\mathbb{1}\{T_{(i)}(u)\leqslant X\}\right)\\
&= \frac{2(1-\theta)}{k}l\left(1-F\left(T_{(l)}(u)\right)\right)+\frac{2(1-\theta)}{k}\sum_{i=l+1}^{k}\left(1-F\left(T_{(i)}(u)\right)\right)\\
&\quad +\frac{2\theta}{k}l\mathbb{E}\left(F(X)\mathbb{1}\{T_{(l)}(u)\leqslant X\}\right)+\frac{2\theta}{k}\sum_{i=l+1}^{k}\mathbb{E}\left(F(X)\mathbb{1}\{T_{(i)}(u)\leqslant X\}\right)\\
&\overset{(I)\&(II)}{=} \frac{2(1-\theta)}{k\theta}l\left(1-F^{\theta}\left(T_{(l)}(u)\right)\right)+\frac{2(1-\theta)}{k\theta}\sum_{i=l+1}^{k}\left(1-F^{\theta}\left(T_{(i)}(u)\right)\right)\\
&\quad +\frac{\theta}{k}l\left(1-F\left(T_{(l)}(u)\right)^2\right)+\frac{\theta}{k}\sum_{i=l+1}^{k}\left(1-F\left(T_{(i)}(u)\right)^2\right)\\
&= \frac{2(1-\theta)}{\theta}-\frac{2(1-\theta)}{k\theta}lF^{\theta}\left(T_{(l)}(u)\right)-\frac{2(1-\theta)}{k\theta}\sum_{i=l+1}^{k}F^{\theta}\left(T_{(i)}(u)\right)\\
&\quad +\frac{1}{k\theta}l\left(2\theta-1+2(1-\theta)F^{\theta}(T_{(l)}(u))-F^{\theta}\left(T_{(l)}(u)\right)^2\right)\\
&\quad +\frac{1}{k\theta}\sum_{i=l+1}^{k}\left(2\theta-1+2(1-\theta)F^{\theta}(T_{(i)}(u))-F^{\theta}\left(T_{(i)}(u)\right)^2\right)\\
&= \frac{2(1-\theta)}{\theta}+\frac{2\theta-1}{\theta}-\frac{1}{k\theta}lF^{\theta}\left(T_{(l)}(u)\right)^2-\frac{1}{k\theta}\sum_{i=l+1}^{k}F^{\theta}\left(T_{(i)}(u)\right)^2\\
&= \frac{1}{\theta}-\frac{1}{k\theta}lF^{\theta}\left(T_{(l)}(u)\right)^2-\frac{1}{k\theta}\sum_{i=l+1}^{k}F^{\theta}\left(T_{(i)}(u)\right)^2
\end{aligned}
$$

$$
\begin{aligned}
(IX)\ \ & \mathbb{E}\left(F^{\theta}(X)^2\,\mathbb{1}\{X \geqslant T_{(l)}(u)\}\right) \\
&= (1-\theta)^2\mathbb{E}\left(\mathbb{1}\{X \geqslant T_{(l)}(u)\}\right) + 2\theta(1-\theta)\mathbb{E}\left(F(X)\,\mathbb{1}\{X \geqslant T_{(l)}(u)\}\right) \\
&\quad + \theta^2\mathbb{E}\left(F(X)^2\,\mathbb{1}\{X \geqslant T_{(l)}(u)\}\right) \\
&\overset{(I)\&(II)}{=} \frac{(1-\theta)^2}{\theta}\left(1-F^{\theta}\left(T_{(l)}(u)\right)\right) + \theta(1-\theta)\left(1-F\left(T_{(l)}(u)\right)^2\right) \\
&\quad + \frac{1}{3}\theta^2\left(1-F\left(T_{(l)}(u)\right)^3\right) \\
&= \frac{(1-\theta)^2}{\theta}\left(1-F^{\theta}\left(T_{(l)}(u)\right)\right) \\
&\quad + \frac{1-\theta}{\theta}\left(2\theta-1+2(1-\theta)F^{\theta}\left(T_{(l)}(u)\right)-F^{\theta}\left(T_{(l)}(u)\right)^2\right) \\
&\quad + \frac{1}{3}\theta^2\left(1-F\left(T_{(l)}(u)\right)^3\right) \\
&= 1-\theta+\frac{(1-\theta)^2}{\theta}F^{\theta}\left(T_{(l)}(u)\right)-\frac{1-\theta}{\theta}F^{\theta}\left(T_{(l)}(u)\right)^2-\frac{1}{3}\theta^2F\left(T_{(l)}(u)\right)^3+\frac{1}{3}\theta^2
\end{aligned}
$$

This gives us the following for the first distance function:

$$
\begin{aligned}
\Delta^{[\mathrm{CMmod1}]}(F_k,F^{\theta}) &= (III)-(IV)+(V) \\
&= \frac{1}{\theta}-\frac{2}{k}\frac{1}{\theta}\sum_{i=1}^{k}\frac{i-\frac{1}{2}}{k}\cdot F^{\theta}(T_{(i)}(u))-\frac{1}{\theta}+\frac{1}{\theta}\frac{1}{k}\sum_{i=1}^{n}F^{\theta}(T_i(u))^2+\frac{1}{3}\theta^2+(1-\theta) \\
&= \frac{1}{3}\theta^2-\theta+1+\frac{1}{\theta}\frac{1}{k}\sum_{i=1}^{k}\left(F^{\theta}(T_{(i)}(u))^2-2\cdot\frac{i-\frac{1}{2}}{k}\cdot F^{\theta}(T_{(i)}(u))+\left(\frac{i-\frac{1}{2}}{k}\right)^2\right) \\
&\quad -\frac{1}{\theta}\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i-\frac{1}{2}}{k}\right)^2 \\
&= \frac{1}{3}\theta^2-\theta+1+\frac{1}{3\theta}+\frac{1}{\theta}\frac{1}{k}\sum_{i=1}^{k}\left(F^{\theta}(T_{(i)}(u))-\frac{i-\frac{1}{2}}{k}\right)^2+\frac{1}{12}\frac{1}{\theta}\frac{1}{k^2} \\
&= -\frac{1}{3}\frac{1}{\theta}(1-\theta)^3+\frac{1}{\theta}\frac{1}{k}\sum_{i=1}^{k}\left(F^{\theta}(T_{(i)}(u))-\frac{i-\frac{1}{2}}{k}\right)^2+\frac{1}{12}\frac{1}{\theta}\frac{1}{k^2}
\end{aligned}
$$

$$
\begin{aligned}
&\Delta^{[\mathrm{CMmod2}]}(F_k, F^\theta) \\
&\quad= \frac{1}{\theta^2}\mathbb{E}\left(\left(\max\{F_k(X), 1-\theta\} - F^\theta(X)\right)^2\right) \\
&\quad= \frac{1}{\theta^2}\mathbb{E}\left(\left(1-\theta-F^\theta(X)\right)^2 \mathbb{1}\{1-\theta > F_k(X)\}\right) \\
&\qquad\quad + \frac{1}{\theta^2}\mathbb{E}\left(\left(F_k(X)-F^\theta(X)\right)^2 \mathbb{1}\{1-\theta \leqslant F_k(X)\}\right) \\
&\quad\overset{(VI)}{=} \mathbb{E}\left(F(X)^2 \mid X < T_{(l)}(u)\right)\cdot F(T_{(l)}(u)) \\
&\qquad\quad + \frac{1}{\theta^2}\,\mathbb{E}\left(\left(F_k(X)-F^\theta(X)\right)^2 \mathbb{1}\{X \geqslant T_{(l)}(u)\}\right) \\
&\quad= \frac{1}{3}\cdot F(T_{(l)}(u))^3 + \frac{1}{\theta^2}\left((VII)-(VIII)+(IX)\right) \\
&\quad= \frac{1}{\theta^3} - \frac{l^2}{k^2\theta^3}F^\theta\left(T_{(l)}(u)\right) - \frac{2}{k^2}\frac{1}{\theta^3}\sum_{i=l+1}^{k}\left(i-\frac{1}{2}\right)F^\theta\left(T_{(i)}(u)\right) \\
&\qquad\quad - \frac{1}{\theta^3} + \frac{l}{k\theta^3}F^\theta\left(T_{(l)}(u)\right)^2 + \frac{1}{k\theta^3}\sum_{i=l+1}^{k}F^\theta\left(T_{(i)}(u)\right)^2 \\
&\qquad\quad + \frac{1-\theta}{\theta^2} + \frac{(1-\theta)^2}{\theta^3}F^\theta\left(T_{(l)}(u)\right) - \frac{1-\theta}{\theta^3}F^\theta\left(T_{(l)}(u)\right)^2 + \frac{1}{3} \\
&\quad= \frac{1}{3} + \frac{1-\theta}{\theta^2} \\
&\qquad\quad - \frac{l^2-(k(1-\theta))^2}{k^2\theta^3}F^\theta\left(T_{(l)}(u)\right) + \frac{l-k(1-\theta)}{k\theta^3}F^\theta\left(T_{(l)}(u)\right)^2 \\
&\qquad\quad + \frac{1}{k\theta^3}\sum_{i=l+1}^{k}\left(F^\theta(T_{(i)}(u)) - \frac{i-\frac{1}{2}}{k}\right)^2 - \frac{1}{k^3\theta^3}\sum_{i=l+1}^{k}\left(i-\frac{1}{2}\right)^2 \\
&\quad= \frac{1}{k\theta^3}\sum_{i=l+1}^{k}\left(F^\theta(T_{(i)}(u)) - \frac{i-\frac{1}{2}}{k}\right)^2 + \frac{1}{3} + \frac{1-\theta}{\theta^2} \\
&\qquad\quad - \frac{l^2-(k(1-\theta))^2}{k^2\theta^3}F^\theta\left(T_{(l)}(u)\right) + \frac{l-k(1-\theta)}{k\theta^3}F^\theta\left(T_{(l)}(u)\right)^2 \\
&\qquad\quad + \frac{1}{12}\frac{1}{k^3\theta^3}(k-l) - \frac{1}{3}\frac{1}{k^3\theta^3}\left(k^3-l^3\right) \\
&\quad= \frac{1}{k\theta^3}\sum_{i=l+1}^{k}\left(F^\theta(T_{(i)}(u)) - \frac{i-\frac{1}{2}}{k}\right)^2 + \frac{1}{12}\frac{k-l}{k^3\theta^3} + \frac{1}{3}\frac{l^3-(k(1-\theta))^3}{k^3\theta^3} \\
&\qquad\quad - \frac{l^2-(k(1-\theta))^2}{k^2\theta^3}F^\theta\left(T_{(l)}(u)\right) + \frac{l-k(1-\theta)}{k\theta^3}F^\theta\left(T_{(l)}(u)\right)^2
\end{aligned}
$$

B.2 Additional calculations

Now, we show that

$$\lim_{\theta \to 0} \Delta^{[\text{CMmod1}]}(F_k, F^{(\beta,\theta,\sigma_u)}) = 0.$$

and that for θ such that $\lceil k(1-\theta) \rceil = k$ holds:

$$\Delta^{[\text{CMmod2}]}(F_k, F^{(\beta,\theta,\sigma_u)}) = \frac{1}{3} - F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right) + F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right)^2.$$

Let $X_\beta \sim \text{ML}(\beta, 1)$. Then,

$$X_{\beta,\theta,\sigma_u} := \theta^{-1/\beta}\sigma_u X_\beta \sim \text{ML}(\beta, \theta^{-1/\beta}\sigma_u)$$

and

$$F_*^{(\beta,\theta,\sigma_u)}(x) = F_*^{(\beta,1,1)}(\theta^{1/\beta}\sigma_u^{-1}x).$$

Remember that $F^{(\beta,\theta,\sigma_u)}$ is the c.d.f. of the mixture distribution and $F_*^{(\beta,\theta,\sigma_u)}$ the c.d.f. of the continuous part, the Mittag-Leffler distribution. It holds that $F_*^{(\beta,\theta,\sigma_u)}(X_{\beta,\theta,\sigma_u}) \overset{d}{=} U$ is standard uniformly distributed.

For all $k \in \mathbb{N}$, it holds

$$\begin{aligned}
&\lim_{\theta \to 0} \Delta^{[\text{CMmod1}]}(F_k, F^{(\beta,\theta,\sigma_u)}) \\
&\quad = \lim_{\theta \to 0} \mathbb{E}\left(\left(F_k(X_{\beta,\theta,\sigma_u}) - F^{(\beta,\theta,\sigma_u)}(X_{\beta,\theta,\sigma_u})\right)^2\right) \\
&\quad = \lim_{\theta \to 0} \mathbb{E}\left(\left(F_k(\theta^{-1/\beta}\sigma_u X_\beta) - (1-\theta) - \theta F_*^{(\beta,\theta,\sigma_u)}(X_{\beta,\theta,\sigma_u})\right)^2\right) \\
&\quad = \mathbb{E}\left(\left(F_k(\lim_{\theta \to 0} \theta^{-1/\beta}\sigma_u X_\beta) - \lim_{\theta \to 0}((1-\theta) + \theta\, U)\right)^2\right) \\
&\quad = 0.
\end{aligned}$$

Hereby, we can interchange the limit and the expected value because of dominated convergence.

In the case of $l = \lceil k(1-\theta) \rceil = k$, which is the case if $\theta < \frac{1}{k}$, the distance function can be abbreviated in the following form. It is then only dependent on the largest IET $T_{(k)}(u)$.

$$
\begin{aligned}
&\Delta^{[\mathrm{CMmod2}]}\big(F_k, F^{(\beta,\theta,\sigma_u)}\big) \\
&\quad= \frac{1}{3}\frac{k^3-(k(1-\theta))^3}{k^3\theta^3} - \frac{k^2-(k(1-\theta))^2}{k^2\theta^3}F^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right) \\
&\qquad+ \frac{k-k(1-\theta)}{k\theta^3}F^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right)^2 \\
&\quad= \frac{1}{3}\frac{\theta^3-3\theta^2+3\theta}{\theta^3} + \frac{\theta^2-2\theta}{\theta^3}\left(1-\theta+\theta F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right)\right) \\
&\qquad+ \frac{\theta}{\theta^3}\left(1-\theta+\theta F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right)\right)^2 \\
&\quad= \frac{1}{3} + \frac{-\theta^2+\theta+\theta^2-2\theta-\theta^3+2\theta^2+\theta-2\theta^2+\theta^3}{\theta^3} \\
&\qquad+ \frac{\theta^3-2\theta^2}{\theta^3}F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right) + \frac{\theta^3}{\theta^3}F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right)^2 \\
&\qquad+ \frac{2\theta^2-2\theta^3}{\theta^3}F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right) \\
&\quad= \frac{1}{3} + \frac{-\theta^2+\theta+\theta^2-2\theta-\theta^3+2\theta^2+\theta-2\theta^2+\theta^3}{\theta^3} \\
&\qquad- \frac{\theta^3}{\theta^3}F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right) + \frac{\theta^3}{\theta^3}F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right)^2 \\
&\quad= \frac{1}{3} - F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right) + F_*^{(\beta,\theta,\sigma_u)}\left(T_{(k)}(u)\right)^2
\end{aligned}
$$

C Appendix to Chapter 5

C.1 Weakly consistency of CMmod2

In Chapter 5 we only considered the CMmod1 estimator. Similar results and theorems also hold for the CMmod2 estimator and would have been repetitiv for the reader therefore for completeness the results for CMmod2 are shown here in the appendix.

The second modified Cramér-von Mises distance between $F^{(\phi,\theta)}$ and another c.d.f. G of a positive valued distribution is defined as

$$\check{\Delta}(G, F^{(\phi,\theta)}) = \frac{1}{\theta^2} \int_0^\infty \left(\max\{G(x), 1-\theta\} - F^{(\phi,\theta)}(x)\right)^2 \cdot f_*^{(\phi)}(x)\, \mathrm{d}x \tag{C.1}$$

The CMmod2 estimator is then defined as

$$\widetilde{(\phi,\theta)}(F_{k_n}) = \underset{(\phi,\theta)\in\Phi\times[a,1]}{\arg\min}\ \check{\Delta}(F_{k_n}, F^{(\phi,\theta)}) \tag{C.2}$$

$$= \underset{(\phi,\theta)\in\Phi\times[a,1]}{\arg\min}\ \frac{1}{\theta^2}\Delta(\max\{F_{k_n}, 1-\theta\}, F^{(\phi,\theta)}). \tag{C.3}$$

The reasoning is very similar to the reasoning in Section 5.4. Additionally we use the following facts:

Lemma C.1.

- *For all* $\theta \in [a,1]$*:*

$$\frac{1}{\theta} \leqslant \frac{1}{a}. \tag{C.4}$$

- *For two monotone increasing functions* $G : \mathbb{R} \to \mathbb{R}$ *and* $H : \mathbb{R} \to \mathbb{R}$

$$|\max\{G(x), b\} - \max\{H(x), b\}| \leqslant |G(x) - H(x)| \quad \textit{for all } x, \tag{C.5}$$

i.e.,

– $|\check{F}_{k_n}^{(\theta)}(x) - \check{F}^{(n,\theta)}(x)| \leqslant |F_{k_n}(x) - F^{(n)}(x)|$,

– $|\check{F}_{k_n}^{(\theta)}(x) - F^{(\phi,\theta)}(x)| \leqslant |F_{k_n}(x) - F^{(\phi,\theta)}(x)|$,

– $|\check{F}^{(n,\theta)}(x) - F^{(\phi,\theta)}(x)| \leqslant |F^{(n)}(x) - F^{(\phi,\theta)}(x)|$ *and*

– $|\check{F}^{(\tilde{\phi},\tilde{\theta},\theta)}(x) - F^{(\phi,\theta)}(x)| \leqslant |F^{(\tilde{\phi},\tilde{\theta})}(x) - F^{(\phi,\theta)}(x)|$,

where $\check{F}_{k_n}^{(\theta)}(x) = \max\{F_{k_n}(x), 1-\theta\} \cdot \mathbb{1}_{[0,\infty)}(x)$, $\check{F}^{(n,\theta)}(x) = \max\{F^{(n)}(x), 1-\theta\} \cdot \mathbb{1}_{[0,\infty)}(x)$ *and* $\check{F}^{(\tilde{\phi},\tilde{\theta},\theta)}(x) = \max\{F^{(\tilde{\phi},\tilde{\theta})}(x), 1-\theta\} \cdot \mathbb{1}_{[0,\infty)}(x)$.

Theorem C.1. *Let* $F^{(n)}(x) \underset{n\to\infty}{\longrightarrow} F^{(\phi,\theta)}(x)$, $(\phi,\theta) \in \Phi \times [a,1]$, *for all* $x > 0$ *and let* $(\check{\phi}_{k_n}, \check{\theta}_{k_n}) = \widecheck{(\phi,\theta)}(F_{k_n})$.

a) Then,

$$\check{\Delta}\left(F^{(n)}, F^{(\check{\phi}_{k_n},\check{\theta}_{k_n})}\right) \underset{n\to\infty}{\longrightarrow} 0, \quad \textit{in probability.} \tag{C.6}$$

The distance converges almost surely if the Condition 3 for the sequence $(k_n)_{n\in\mathbb{N}}$ *is fulfilled.*

b) The distance between the asymptotic and the fitted distribution converges to 0. More precisely,

$$\check{\Delta}\left(F^{(\phi,\theta)}, F^{(\check{\phi}_{k_n},\check{\theta}_{k_n})}\right) \underset{n\to\infty}{\longrightarrow} 0, \quad \textit{in probability.} \tag{C.7}$$

Proof. For simplicity we denote $k = k_n$, $F_k = F_{k_n}$.

Part a) The proof is analogously to the proof of Theorem 3 using that

$$\begin{aligned} 0 < \check{\Delta}\left(F_k, F^{(\phi,\theta)}\right) &= \frac{1}{\theta^2}\int_0^\infty \left(\max\{F_k(x), 1-\theta\} - F^{(\phi,\theta)}\right)^2 f_*^{(\phi)}(x)\,\mathrm{d}x \\ &\leqslant \frac{1}{\theta^2}\int_0^\infty \left(F_k(x) - F^{(\phi,\theta)}\right)^2 f_*^{(\phi)}(x)\,\mathrm{d}x \\ &= \frac{1}{\theta^2}\Delta\left(F_k, F^{(\phi,\theta)}\right) \to 0. \end{aligned}$$

$$
\begin{aligned}
&\check{\Delta}\left(F^{(n)}, F^{(\check{\phi}_k,\check{\theta}_k)}\right) = \frac{1}{\check{\theta}_k^2}\int_0^\infty \left(\check{F}^{(n,\check{\theta}_k)}(x) - F^{(\check{\phi}_k,\check{\theta}_k)}(x)\right)^2 \cdot f_*^{(\check{\phi}_k)}(x)\,\mathrm{d}x \\
&= \check{\Delta}\left(F_k, F^{(\check{\phi}_k,\check{\theta}_k)}\right) + \frac{1}{\check{\theta}_k^2}\int_0^\infty \left(\check{F}^{(n,\check{\theta}_k)}(x) - \check{F}_k^{(\check{\theta}_k)}(x)\right)^2 \cdot f_*^{(\check{\phi}_k)}(x)\,\mathrm{d}x \\
&\quad + \frac{2}{\check{\theta}_k^2}\cdot\int_0^\infty \left(\check{F}_k^{(\check{\theta}_k)}(x) - F^{(\check{\phi}_k,\check{\theta}_k)}(x)\right)\left(\check{F}^{(n,\check{\theta}_k)}(x) - \check{F}_k^{(\check{\theta}_k)}(x)\right)\cdot f_*^{(\check{\phi}_k)}(x)\mathrm{d}x \\
&\leqslant \check{\Delta}\left(F_k, F^{(\phi,\theta)}\right) + \frac{1}{a^2}\int_0^\infty \left(F^{(n)}(x) - F_k(x)\right)^2 \cdot f_*^{(\check{\phi}_k)}(x)\,\mathrm{d}x \\
&\quad + \frac{2}{a^2}\cdot\int_0^\infty \left|F^{(n)}(x) - F_k(x)\right|\cdot f_*^{(\check{\phi}_k)}(x)\mathrm{d}x \\
&\leqslant \check{\Delta}\left(F_k, F^{(\phi,\theta)}\right) + \frac{1}{a^3}\cdot\|F^{(n)} - F_k\|_\infty \\
&\longrightarrow 0,
\end{aligned}
$$

where $\|g\|_\infty := \sup_x |g(x)|$.

Part b) This part can also be done analogously as for Theorem 3.

□

Theorem C.2. *Under the same conditions as in Theorem C.1 b), it holds*

$$\widetilde{(\phi,\theta)}(F_{k_n}) \to (\phi,\theta) \quad \textit{in probability.} \tag{C.8}$$

Proof of Theorem C.2

Proof. This proof can be analogously done as for Theorem 4. We define $\check{M}(\tilde{\phi},\tilde{\theta})$ as

$$\check{M}(\tilde{\phi},\tilde{\theta}) := \check{\Delta}(F^{(\phi,\theta)}, F^{(\tilde{\phi},\tilde{\theta})}). \tag{C.9}$$

We will show that

$$\forall\varepsilon > 0 : \inf_{(\tilde{\phi},\tilde{\theta})\in\Theta_\varepsilon} \check{M}(\tilde{\phi},\tilde{\theta}) > \check{M}(\phi,\theta) = 0, \tag{C.10}$$

where $\Theta_\varepsilon := \{(\tilde{\phi},\tilde{\theta}) \in \Phi\times[a,1] \mid \|(\tilde{\phi}-\phi,\tilde{\theta}-\theta)\| \geqslant \varepsilon\}$. From there the consistency follows as in Theorem 4.

To prove (C.10), we again show the following two conditions are fulfilled:

1. $\check{M}(\tilde{\phi},\tilde{\theta}) > 0 \quad \forall\,(\tilde{\phi},\tilde{\theta}) \neq (\phi,\theta)$
2. $\check{M}(\cdot,\cdot)$ is continuous.

The second condition can be shown analogously as in the proof for Theorem 4. For the first condition we consider two cases. Let be $(\tilde{\phi}, \tilde{\theta}) \neq (\phi, \theta)$. If $\tilde{\theta} \leqslant \theta$, then $\check{M}(\tilde{\phi}, \tilde{\theta}) = M(\tilde{\phi}, \tilde{\theta})$ ($M(\cdot, \cdot)$ defined as in the proof of Theorem 4) for which we have already shown the desired result. For $\tilde{\theta} > \theta$ it holds that $x_b > 0$ and $b > 0$ exist such that $\check{F}^{(\phi,\theta,\tilde{\theta})}(x_b) = 1-\tilde{\theta}$ and $F^{(\tilde{\phi},\tilde{\theta})}(x_b) = 1-\tilde{\theta}+b$, i.e. $|\check{F}^{(\phi,\theta,\tilde{\theta})}(x_b)-F^{(\tilde{\phi},\tilde{\theta})}(x_b)| = b > 0$, since $F^{(\phi,\theta)}$ and $F^{(\tilde{\phi},\tilde{\theta})}$ are continuous and strictly monotone increasing for all $x \geqslant 0$. Therefore $\exists c \in (0, b], \quad \exists B \in \mathcal{B}_{[0,\infty)}$ with $x_b \in B$ and $\mu_{\tilde{\phi}}(B) > 0$:

$$\inf_{x \in B} |\check{F}^{(\phi,\theta,\tilde{\theta})}(x) - F^{(\tilde{\phi},\tilde{\theta})}(x)| = c \tag{C.11}$$

This leads to

$$\check{M}(\tilde{\phi}, \tilde{\theta}) \geqslant c^2 \cdot \mu_{\tilde{\phi}}(B) > 0. \tag{C.12}$$

□

C.2 Comparison of the e.c.d.f. computed by independent or dependent interexceedance times

In Chapter 5, we proved several consistency theorems under a simplifying assumption of independent interexceedance times. Therefore we now illustrate briefly that this has no strong impact since the e.c.d.f. of dependent interexceedance times seems to approximate the theoretical c.d.f. similarly well as the e.c.d.f. of independent interexceedance times.

We did a small simulation study with the following scenarios. The magnitudes were generated from a max-autoregressive process and the waiting times are Mittag-Leffler distributed. We varied the parameters $(\beta, \theta) \in \{0.2, 0.5, 0.8, 1\}^2$ and in addition we set the threshold $u \in \{q_{0.9}, q_{0.95}, q_{0.99}, q_{0.999}\}$, where q_α is the α-quantile of the unit Fréchet distribution.

Since we do not know the true distribution of the interexceedance times $T(u)$, we approximated it using the e.c.d.f. of 1 000 000 independently generated interexceedance times and compared it with the e.c.d.f.s of 500 independently and dependently generated interexceedance times. We measured the distance with the original Cramér-von Mises distance. We repeated the procedure 500 times.

Figure C.1 (a) shows the mean distance of the dependently and independently generated e.c.d.f.s, where each point symbolises one of the 64 scenarios. If a point lies above the diagonal, the mean distance of the dependently generated e.c.d.f.s is smaller than the mean distance of the independently generated e.c.d.f.s., otherwise the

mean distance is larger. It can be seen that the points lie symmetrically around the diagonal. Thus, the distance of the e.c.d.f. generated by dependent observations does not appear to be systematically larger than the distance of the e.c.d.f. generated by independent observations. This finding also applies to each of the different scenarios summarised here. To confirm the result, we performed a one-sided two-sample t-test with $\alpha = 0.05$ for each scenario and used the Bonferroni-Holm adjustment to avoid the problem of multiple testing. No significant differences were found with the t-tests, too. In Figure C.1 (b) we see that the p-values lie relatively uniformly on the interval $[0, 1]$. Under the null hypothesis we would expect $64 \cdot 0.05 = 3.2$ p-value smaller than 0.05. In our sample there is only 1 unadjusted p-values lower than the α-level. Thus, the tests also did not reveal any evidence of large differences.

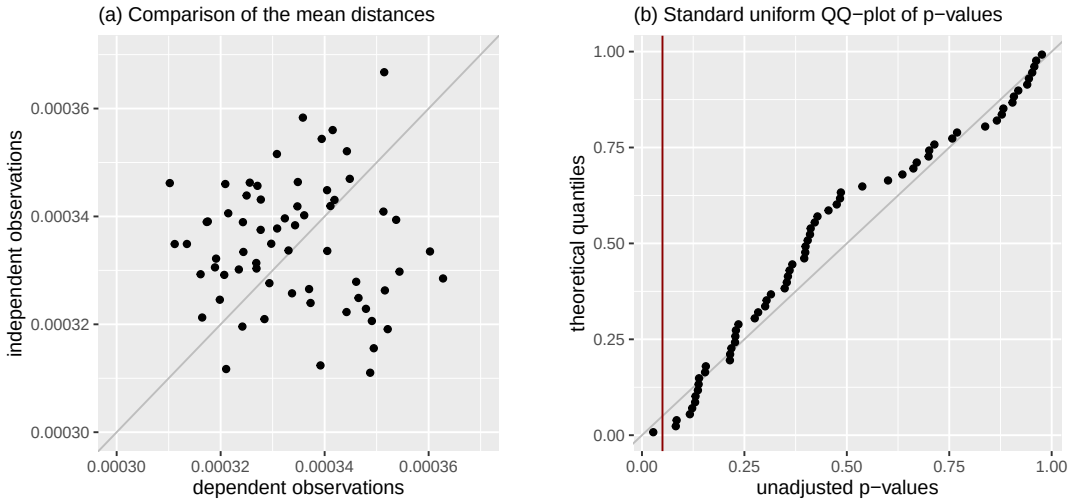

Figure C.1: (a) Comparison of the distance of the e.c.d.f. generated by 1 000 000 independent observations and the e.c.d.f. generated by 500 independent or dependent observations, respectively. Each point symbolises the mean of one scenario. (b) QQ-plot of the unadjusted t-test p-values against the quantiles of the standard uniform distribution. The red line marks $\alpha = 0.05$.

D Appendix to Chapter 6

Figures related to Subsection 6.2.2

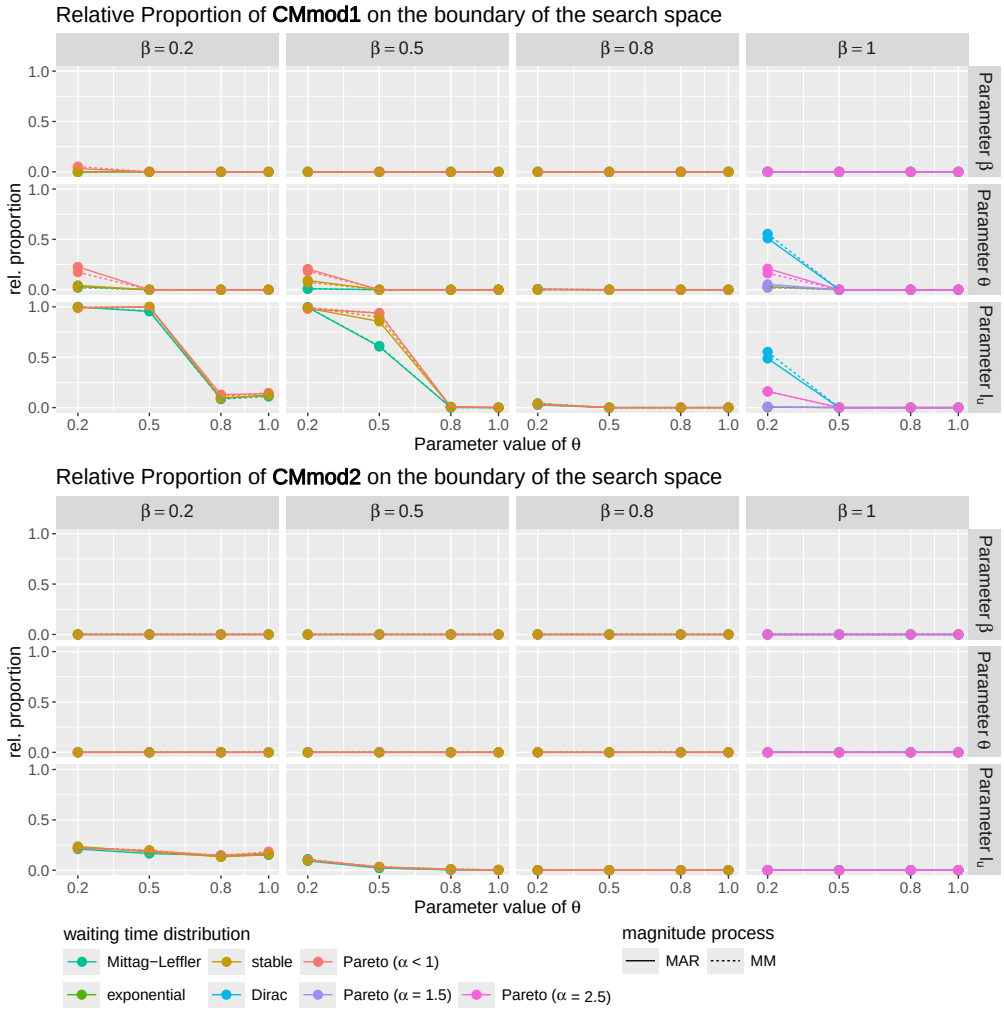

Figure D.1: Relative frequency of simulation runs in which the estimate lies at the (lower) boundary of the search space for CMmod1 (top) and CMmod2 (bottom) for different parameter values, for the MAR and MM process and a varity of waiting time distributions using sample size $n = 10\,000$ and number of exceedances $k = 200$.

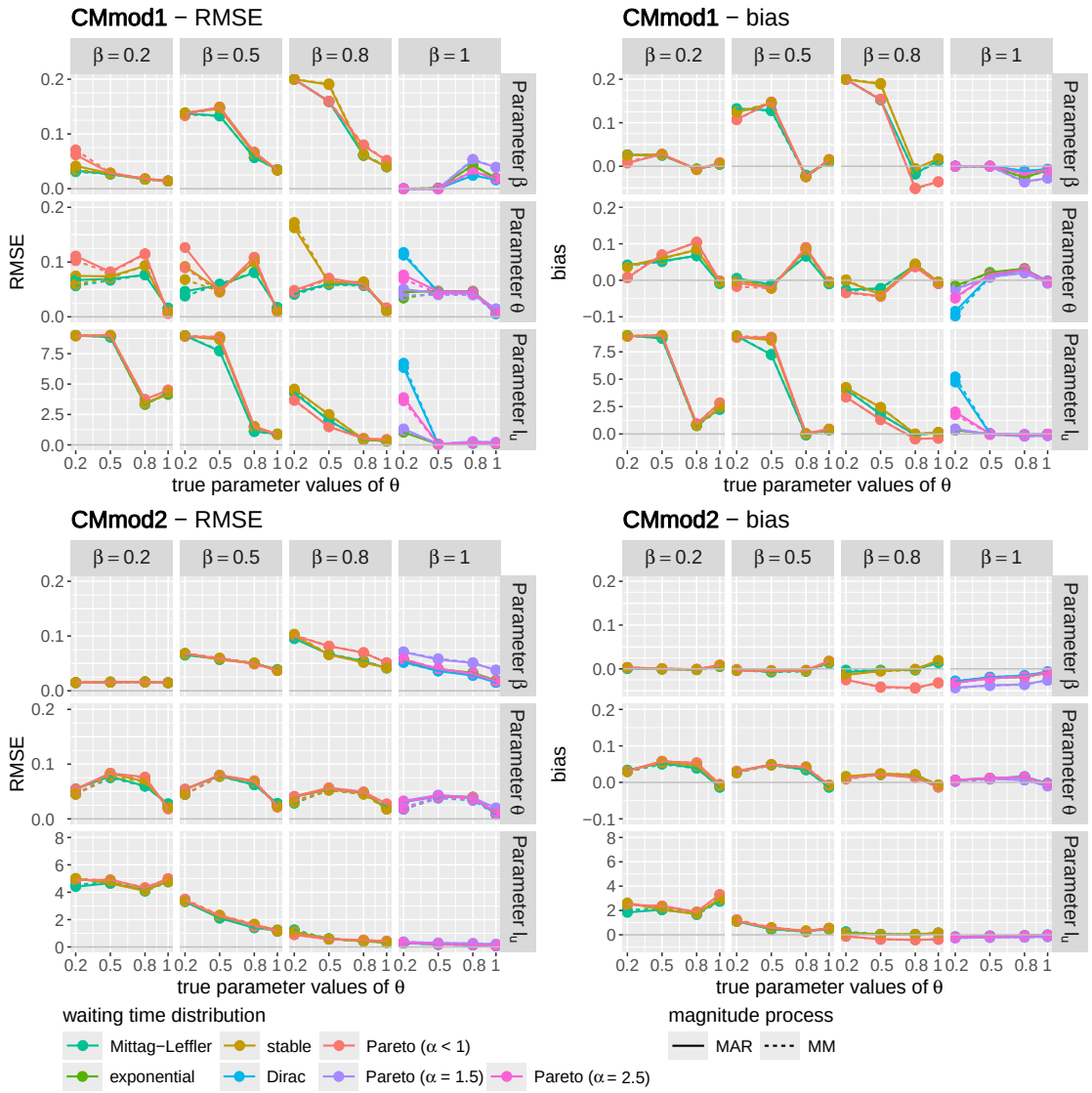

Figure D.2: Comparison of the RMSE (left) and bias (right) of the CMmod1 (top) and CMmod2 (bottom) estimators for the parameters β, θ and l_u for different parameter values, for the MAR and MM process and a varity of waiting time distributions using sample size $n = 10\,000$ and number of exceedances $k = 200$.

Figures related to Subsection 6.3.1

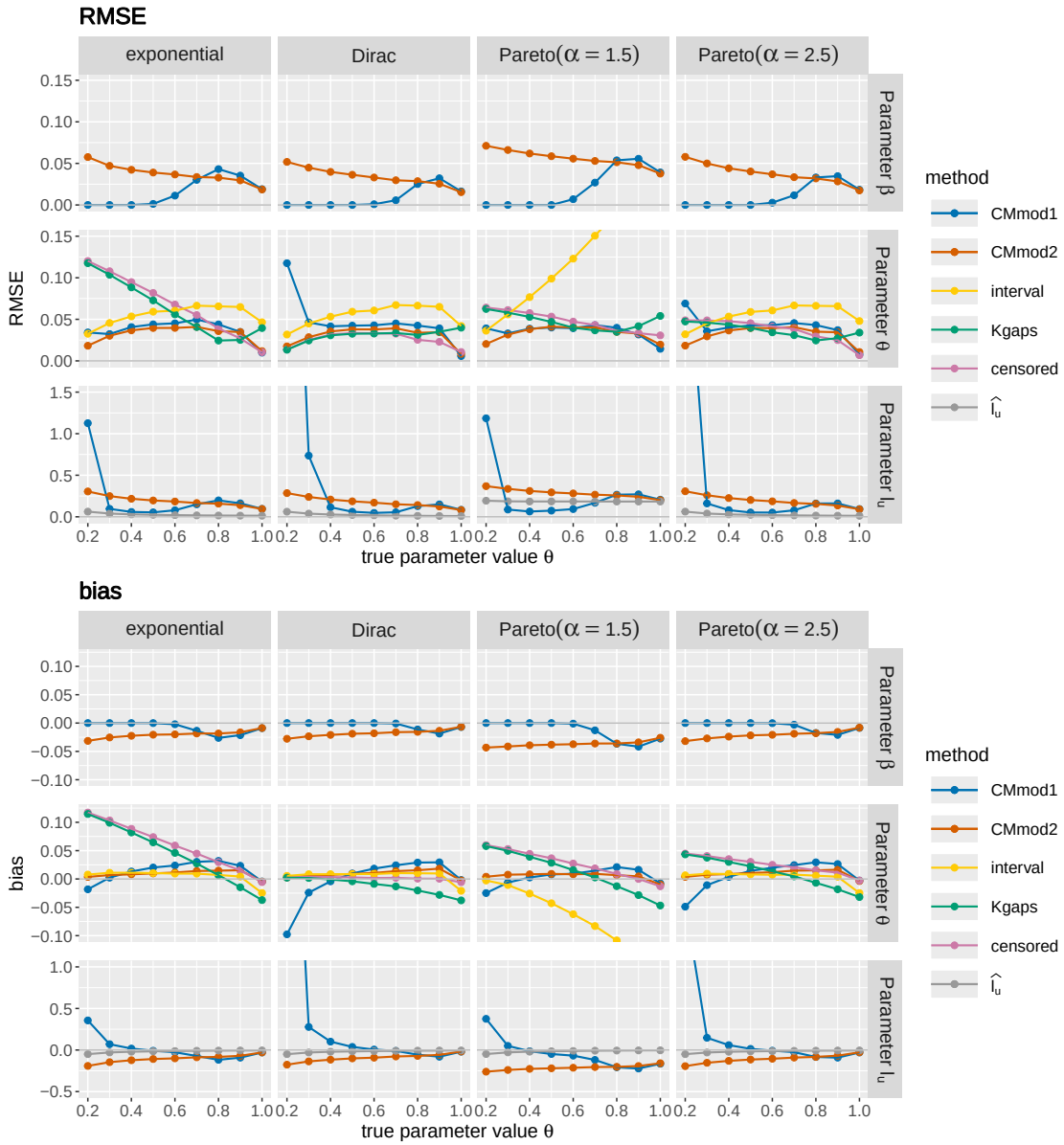

Figure D.3: Comparison of the RMSE (top) and bias (bottom) of CMmod1 and CMmod2 with different established estimators for the parameters (β,) θ and l_u for $\beta = 1$ and different parameter values of θ, and for the MM process and a varity of waiting time distributions using sample size $n = 10\,000$ and number of exceedances $k = 200$.

E Appendix to Chapter 7

Here are given some extra visualisations regarding the case study in Chapter 7.

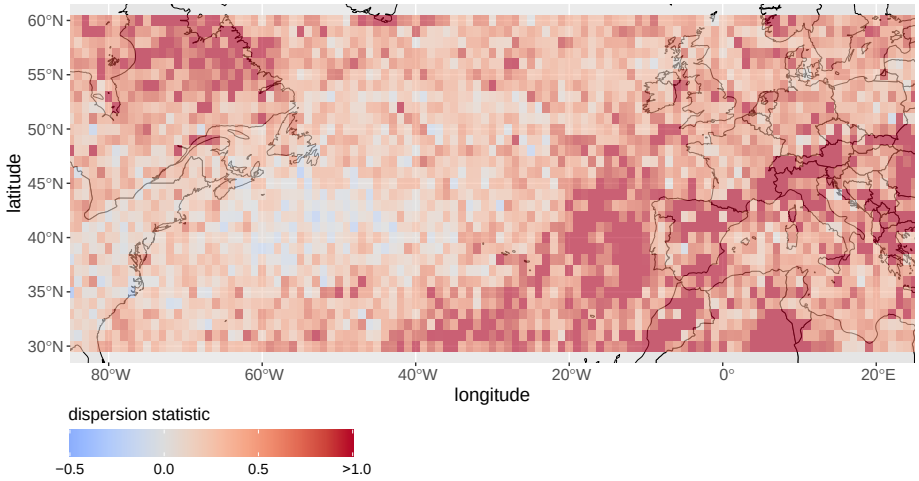

Figure E.1: Empirical dispersion statistics calculated with the number of exceedances within one winter.

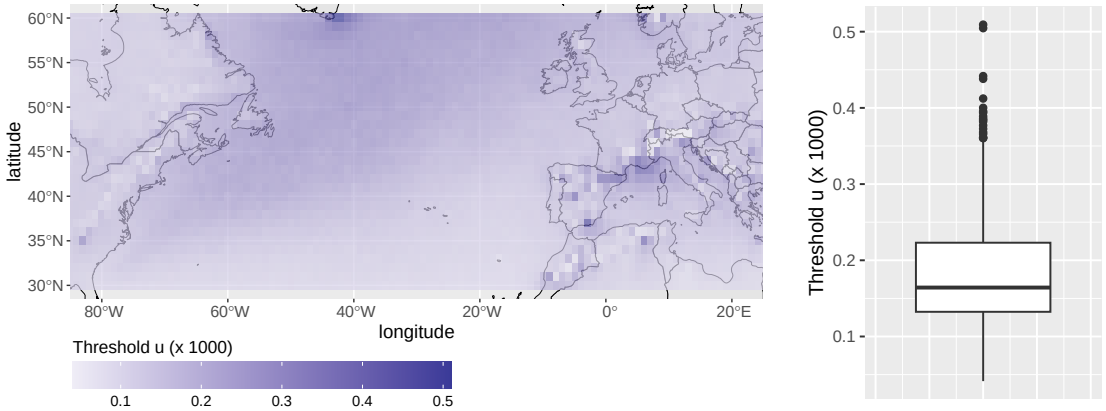

Figure E.2: Heatmap (left) and boxplot (right) of the used thresholds regarding the 99% quantile.

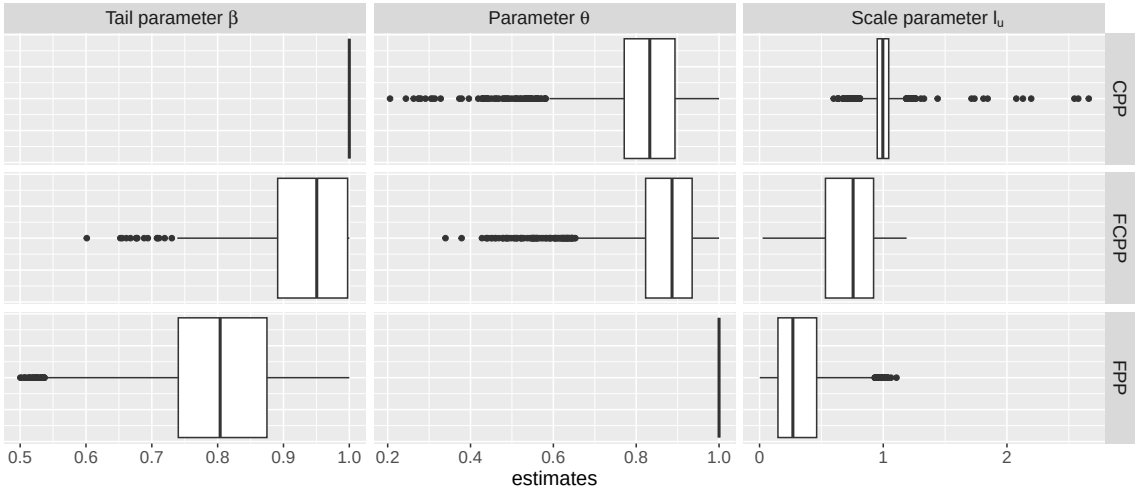

Figure E.3: Boxplots of the tail parameter β (left), extremal index θ (middle) and scale parameter l_u (right) in case of the CPP (top), the FCPP (middle) and FPP (bottom). The tail parameter is equal to 1 in the CPP, while this applies to θ in the FPP.