MODELLING OF MULTIVARIATE EXTREME DATA

by

ANDRéHETTE NEL

THESIS

Submitted in fulfilment of the requirements for the degree

PHILOSOPHIAE DOCTOR

IN

THE FACULTY OF NATURAL AND AGRICULTURAL SCIENCE

DEPARTMENT OF MATHEMATICAL STATISTICS

UNIVERSITY OF THE FREE STATE

BLOEMFONTEIN

NOVEMBER 2007

PROMOTER: PROF. D.J. DE WAAL

ACKNOWLEDGEMENTS

I give God the glory for giving me the ability, strength and endurance to complete this thesis. I thank Him for the opportunity to reach my goals.

A special word of gratitude goes to my promoter, Professor Danie de Waal, for his guidance, patience, encouragement and availability throughout the completion of this thesis.

Thank you to my colleagues, especially Sean, for their input.

To my family, thank you for all the support and motivation throughout my studies.

A special thank you goes to Jaco for bearing with me and encouraging me throughout the completion of this thesis.

TABLE OF CONTENTS

Chapter 1

Introduction

The main focus of this study is to investigate multivariate extreme value distributions and applications. As an introduction to multivariate extremes, univariate extreme value theory and distributions are first discussed. The following basic extreme value theory follows the book by Beirlant et al. (2004).

1.1 Univariate Extreme Value Theory

1.1.1 Introduction

Throughout this study the focus is on the behaviour of the extreme values of a data set, therefore the upper tail of the underlying distribution is of importance. Assume that the data, considered in this study, are realizations of a sample $X_1, X_2, ..., X_n$ of *n* independent and identically distributed (i.i.d.) random variables. The ordered data is then denoted by $X_{1,n} \le ... \le X_{n,n}$, therefore $X_{n,n} = \max\{X_1, X_2, ..., X_n\}$.

1.1.2 The probabilistic side of extreme value theory

A possible limiting distribution must be found for $X_{n,n}$. Extreme value theory concentrates on the search for distributions of X for which there exist a sequence of numbers $b_n, n \ge 1$, and a sequence of positive numbers $a_n, n \ge 1$ such that for all real values *x*

$$P\left(\frac{X_{n,n} - b_n}{a_n} \le x\right) \to G(x), n \to \infty, \qquad (1.1)$$

for some non-degenerate distribution function *G*.

All extreme value distributions given by the equation

$$G_\gamma(x) = \exp\left(-(1+\gamma x)^{-\frac{1}{\gamma}}\right), \gamma \in \square \tag{1.2}$$

can occur as limits in (1.1). This was shown by Fisher and Tippet (1928), Gnedenko (1943) and de Haan (1970). They are also the only limits of (1.1). For all bounded and continuous functions z over the domain of F

$$E\left(z\left(a_n^{-1}\left(X_{n,n} - b_n\right)\right)\right) \to \int z(\upsilon)\partial G_\gamma(\upsilon), n \to \infty. \tag{1.3}$$

Then, because $P\left(X_{n,n} \le x\right) = F^n(x)$,

$$E\left(z\left(a_n^{-1}\left(X_{n,n} - b_n\right)\right)\right) = n\int_{-\infty}^{\infty} z\left(\frac{x-b_n}{a_n}\right)F^{n-1}(x)\partial F(x). \tag{1.4}$$

Let $F(x) = 1 - \frac{v}{n}$, then

$$E\left(z\left(a_n^{-1}\left(X_{n,n} - b_n\right)\right)\right) = \int_0^n z\left(\frac{Q\left(1-\frac{v}{n}\right) - b_n}{a_n}\right)\left(1-\frac{v}{n}\right)^{n-1}\partial v \tag{1.5}$$

where $Q\left(1-\frac{v}{n}\right)$ represents the tail quantile function.

Let $U(x) = Q\left(1-\frac{1}{x}\right)$ and $b_n = Q\left(1-\frac{1}{n}\right)$. Following from the fact that $\left(1-\frac{v}{n}\right)^{n-1} \to e^{-v}$ as $n \to \infty$ a limit for $E\left(z\left(a_n^{-1}\left(X_{n,n} - b_n\right)\right)\right)$ can be obtained for some function a when

$$\left\{U(xu) - U(x)\right\}/a(x) \text{ converges for } u > 0 \text{ as } x \to \infty. \tag{1.6}$$

The possible limits in (1.6) are then given by

$$h_\gamma(u) = \int_1^u \upsilon^{\gamma-1}\partial\upsilon = \frac{u^\gamma - 1}{\gamma}. \tag{1.7}$$

Therefore under (1.6)

$$E\left(z\left(a_n^{-1}\left(X_{n,n} - b_n\right)\right)\right) = \int_0^\infty z\left(h_\gamma\left(\frac{1}{\upsilon}\right)\right)e^{-\upsilon}\partial\upsilon \text{ as } n \to \infty. \tag{1.8}$$

Three types of standard extremal type appear:

(i) The Fréchet-Pareto-type class of distributions, $\gamma > 0$

$$E\left(z\left(a_n^{-1}\left(X_{n,n} - b_n\right)\right)\right) = \int_{-\gamma^{-1}}^\infty z(u)\partial\left(\exp\left(-(1+\gamma u)^{\frac{-1}{\gamma}}\right)\right); \tag{1.9}$$

(ii) The Gumbel class of distributions, $\gamma = 0$

$$E\left(z\left(a_n^{-1}\left(X_{n,n} - b_n\right)\right)\right) = \int_{-\infty}^\infty z(u)\partial\exp\left(-e^{-u}\right); \tag{1.10}$$

(iii) The Extremal Weibull Bounded Pareto class of distributions, $\gamma < 0$

$$E\left(z\left(a_n^{-1}\left(X_{n,n} - b_n\right)\right)\right) = \int_{-\infty}^{-\gamma^{-1}} z(u)\partial\left(\exp\left(-(1+\gamma u)^{\frac{-1}{\gamma}}\right)\right) \tag{1.11}$$

where z is any bounded and continuous function over the domain of the distribution function F. γ is called the extreme value index and gives an indication on the heaviness of the tail of the underlying distribution. Pareto type distributions have the heaviest tails and are discussed in more detail in the following section.

1.1.3 Pareto-type distributions

The Pareto-type distributions are considered here because Pareto-type refers to the tail of the distribution which means that as $x \to \infty$ the survival function $1 - F(x)$ tends to zero at a polynomial speed. The Burr, Generalized Pareto, Loggamma and Fréchet distributions are examples of Pareto-type distributions.

The Pareto-type class of distributions is a broad class of distributions where $U(x) = x^{\gamma} l(x)$. $l(x)$ is called the slowly varying function, where l satisfies

$$\lim_{x \to \infty} \frac{l(xt)}{l(x)} = 1 \text{ for all } t > 0. \qquad (1.12)$$

The following shows that (1.6), from Section 1.1.2, is satisfied when choosing $a(x) = \gamma x^{\gamma} l(x) = \gamma U(x)$:

$$\begin{aligned} \{U(xu) - U(x)\} / a(x) &= \left((xu)^{\gamma} l(xu) - x^{\gamma} l(x) \right) / a(x) \\ &= \frac{l(x) x^{\gamma}}{a(x)} \left(\frac{l(xu)}{l(x)} u^{\gamma} - 1 \right) \\ &\to_{x \to \infty} \frac{\left(u^{\gamma} - 1\right)}{\gamma}. \end{aligned}$$

1.1.3.1 The Pareto Quantile Plot

A Pareto quantile plot can be used to determine whether the underlying distribution is a Pareto-type distribution. The underlying distribution is a Pareto-type distribution when the plot shows an ultimate linear effect. This is proven in the Theorem 1.1.

Theorem 1.1

In the case of an underlying Pareto-type distribution, the Pareto quantile plot has a linear effect.

Proof:

In the case of the Pareto-type distribution the tail quantile function U is given by the following equation:

$$U(x) = x^{\gamma} l(x) \tag{1.13}$$

where l, as mentioned before, is a slowly varying function, i.e. $\frac{l(xt)}{l(x)} \to 1$ for all $t > 0$ as $x \to \infty$. Thus

$$Q(1-p) = p^{-\gamma} l\left(\frac{1}{p}\right), \; p \in (0,1) \tag{1.14}$$

and

$$\log Q(1-p) = -\gamma \log p + \log l\left(\frac{1}{p}\right) \tag{1.15}$$

and for every slowly varying function l

$$\frac{\log l\left(\frac{1}{p}\right)}{\log p} \to 0 \text{ as } p \to 0. \tag{1.16}$$

Therefore

$$\frac{\log Q(1-p)}{\log p} \to \gamma \text{ as } p \to 0 . \quad (1.17)$$

#

One can therefore conclude that a linear effect is obtained in a Pareto-quantile plot. Beirlant *et al*. (2004, p. 64) also shows that for the Pareto-type distributions, the follow plot: $\left(\log x_{i,n}, -\log\left(1-p_{i,n}\right)\right), i = 1,...,n$, called the Pareto probability plot, will eventually be linear with a slope equal to $\frac{1}{\gamma}$.

1.1.3.2 Mean excess function

In actuarial practice, especially in reinsurance, the conditioning of a random variable X on some event *i.e.* $(X > t)$ is of great importance. If an actuary wants to decide on what the value of t (the priority level) should be, it is important that he calculates the expected amount that will be paid out per client when a given level t is chosen. Therefore the actuary should calculate the mean excess function, which is given by the following equation:

$$e(t) = E(X - t \mid X > t) \quad (1.18)$$

where $E(X) < \infty$.

The estimates of the mean excesses are given by the following equation:

$$e_{k,n} := \hat{e}_n\left(x_{n-k,n}\right) = \frac{1}{k}\sum_{j=1}^{k} x_{n-j+1,n} - x_{n-k,n}, k = 1,...,n-1 . \quad (1.19)$$

(1.19) is basically the mean of the largest *k* observations that exceed the

(k + 1)th observation. It is also an estimate of $\frac{1}{\lambda}$ if the data comes from the exponential class with $\bar{F}(x) = e^{-\lambda x} l(x)$. $\frac{1}{\lambda}$ is the ultimate slope of the exponential quantile plot.

1.1.3.3 The Hill estimator

The Pareto-type distribution is given by the following survival or equivalent quantile function, (1.20) and (1.21) respectively:

$$\bar{F}(x) = x^{\frac{-1}{\gamma}} l_F(x) \tag{1.20}$$

or

$$Q\left(1-\frac{1}{x}\right) = U(x) = x^{\gamma} l_U(x) \tag{1.21}$$

where l_F and l_U are related slowly varying functions. In this section the Hill estimator is discussed as an estimator of γ, the extreme value index (EVI), specifically when heavy tailed Pareto-type distributions are considered. According to Beirlant *et al*. (2004, p. 99): In practice the slowly varying function l in general is unknown.

Assume a sample of i.i.d. values $\{X_i, 1 \le i \le n\}$ from a Pareto-type tail, $1-F$.

There are different approaches to define and introduce the Hill estimator, the first approach is the quantile method. In Theorem 1.1 it was shown that

$$\frac{\log l\left(\frac{1}{p}\right)}{\log p} \to 0 \text{ as } p \to 0 \tag{1.22}$$

and that the Pareto quantile plot is ultimately linear with slope γ near the largest observations, and in Section 1.1.3.2 the mean excess values, $E_{k,n}$ were discussed. Combining (1.22) and the mean excess values results in the Hill estimator, as shown in the following equation:

$$H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n} \,. \tag{1.23}$$

Thus, taking the log of *X* in (1.19) gives the Hill estimator. Mason (1982) showed that $H_{k,n}$ is a consistent estimator for γ, as $k \to \infty, n \to \infty$ and $\frac{k}{n} \to 0$, whatever the slowly varying function is. The problem of the Hill estimator is that it is biased for different k values and it should be adjusted to reduce the bias. This is shown in the following section. Some guidelines are necessary for selecting *k*.

1.1.3.3.1 Properties of the Hill estimator

The Hill estimator given in (1.23) can also be rewritten in terms of log spacings. This is explained by Rényi's exponential representation. First random variables

$$Z_j := j\left(\log X_{n-j+1,n} - \log X_{n-j,n}\right) =: jT_j \tag{1.24}$$

are introduced. Through partial summation it follows that:

$$\sum_{j=1}^{k} Z_j = \sum_{j=1}^{k} jT_j = \sum_{j=1}^{k}\sum_{i=1}^{j} T_j = \sum_{i=1}^{k}\sum_{j=1}^{k} T_j \tag{1.25}$$

this leads to

$$H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} Z_j = \bar{Z}_k \,. \tag{1.26}$$

Therefore

$$H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} j\left(\log X_{n-j+1,n} - \log X_{n-j,n}\right). \tag{1.27}$$

When elaborations are done by these spacings the following models can be derived to reduce the bias of the Hill:

$$Z_j \approx \left(\gamma + A\left(\frac{n+1}{j+1}\right)\right) f_j, \tag{1.28}$$

$$Z_j \approx \left(\gamma + b_{n,k}\left(\frac{j}{k+1}\right)^{\beta}\right) f_j, \tag{1.29}$$

$$Z_j \approx \gamma + b_{n,k}\left(\frac{j}{k+1}\right)^{\beta} + \gamma\varepsilon_j, 1 \le j \le k, \tag{1.30}$$

where $A \in \mathbb{R}$ and $b_{n,k} = A\left(\frac{k+1}{n+1}\right)^{\beta}$ are constants, $f_1, f_2, \ldots$ are independent $Exp(1)$ distributed random variables and $\varepsilon_j = f_j - 1$. (1.30) is a generalized regression model, with Z_j approximately exponentially distributed with a mean value given by $\gamma + b_{n,k}\left(\frac{j}{n+1}\right)^{\beta}$. If $b_{n,k} > 0$ the mean increases with values of *j*.

The parameters $\gamma, b_{n,k}$ and β in (1.28) and (1.29) can be obtained through maximum likelihood estimation or least squares estimation. For β = 1 the least squares estimate becomes

$$\hat{\gamma}_{LS} = H_{k,n} - \frac{1}{2}\hat{b}^{+}_{LS} \tag{1.31}$$

where

$$\hat{b}_{LS}^{+} = \frac{12}{k}\sum_{j=1}^{k}\left(\left(\frac{j}{k+1}\right)-\frac{1}{2}\right)Z_j \,. \tag{1.32}$$

The asymptotic bias of the Hill estimator is given by the following equation:

$$\begin{aligned} ABias\left(H_{k,n}\right) &\sim b_{n,k}\,\frac{1}{k}\sum_{j=1}^{k}\left(\frac{j}{k+1}\right)^{\beta} \\ &\sim \frac{b_{n,k}}{1+\beta}. \end{aligned} \tag{1.33}$$

The bias will be small if $b_{n,k}$ is small; this means that k should also be small.

The asymptotic variance of the Hill estimator is given by

$$AVar\left(H_{k,n}\right) \sim V\,\mathrm{ar}\left(\frac{\gamma}{k}\sum_{j=1}^{k}E_j\right) \sim \frac{\gamma^2}{k}\,. \tag{1.34}$$

The variance will be small if k is large.

1.1.4 Tail estimation

In Chapter 4 tail estimation is discussed when the underlying distribution is Pareto-type, while in this section tail estimation is discussed in general when $\bar{F}$ falls in any class, whether γ is positive, negative or equal to zero. Three sets of methods are considered:

1.1.4.1 Method of block maxima

At the beginning of this chapter it was mentioned that the extreme value distributions are the only possible limits for a normalized maximum of a random sample when a non-degenerate limit exists. Therefore the EVI can be

estimated by fitting the generalized extreme value distribution (GEV) given by the following equation:

$$G(x;\sigma,\gamma,\mu)=\begin{cases}\exp\left(-\left(1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\gamma}}\right), 1+\gamma\frac{(x-\mu)}{\sigma}>0, \gamma\neq 0\\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), x\in\mathbb{R}, \gamma=0\end{cases} \quad (1.35)$$

where $\sigma>0$ and $\mu\in\mathbb{R}$. μ is the location parameter, σ is the scale parameter and γ,the EVI, is the shape parameter.

This method consists of grouping the data into blocks of equal length. For example, a block size of one year can be chosen. When only modelling block maxima other useful extreme data may be lost, what if there are several extreme values in a block? Therefore, the other two sets of methods may be more relevant. In these methods thresholds are taken into consideration.

1.1.4.2 Methods of extreme order statistics

Three methods of estimating the EVI, based on extreme order statistics, are discussed.

(i) Pickands' estimator

From (1.6), at the beginning of this chapter, the following can be derived:

$$\begin{aligned}
&\frac{U\left(\frac{x}{2}\right)-U(x)}{a(x)} \to \frac{2^{-\gamma}-1}{\gamma}, \\
&\frac{U\left(\frac{x}{4}\right)-U\left(\frac{x}{2}\right)}{a\left(\frac{x}{2}\right)} \to \frac{2^{-\gamma}-1}{\gamma}, \\
&\frac{a(x)}{a\left(\frac{x}{2}\right)} \to 2^{\gamma}
\end{aligned} \tag{1.36}$$

where $U(x)$ is the quantile function $Q\left(1-\frac{1}{x}\right)$, therefore

$$\frac{U\left(\frac{x}{2}\right)-U(x)}{U\left(\frac{x}{4}\right)-U\left(\frac{x}{2}\right)} \to 2^{\gamma}. \tag{1.37}$$

And as $x \to \infty$

$$\frac{1}{\log 2}\log\left(\frac{U\left(\frac{x}{2}\right)-U(x)}{U\left(\frac{x}{4}\right)-U\left(\frac{x}{2}\right)}\right) \to \gamma. \tag{1.38}$$

When choosing $x=\frac{n+1}{k}$ so that $\hat{U}(x)=X_{n-k+1,n}, \hat{U}\left(\frac{x}{2}\right)=X_{n-2k+1,n}$ and $\hat{U}\left(\frac{x}{4}\right)=X_{n-4k+1,n}$, the Pickands estimator is obtained, and given by the following equation:

$$\hat{\gamma}_{P,k}=\frac{1}{\log 2}\log\left(\frac{X_{n-\left[\frac{k}{4}\right]+1,n}-X_{n-\left[\frac{k}{2}\right]+1,n}}{X_{n-\left[\frac{k}{2}\right]+1,n}-X_{n-k+1,n}}\right), \tag{1.39}$$

$k = 1,...,n,$ Pickands (1975).

(ii) Moment estimator

This is a direct generalization of the Hill estimator and was introduced by Dekkers *et al*. (1989). The moment estimator is derived and explained by Beirlant *et al*. (2004, p. 142-143) and defined as

$$M_{k,n} = H_{k,n} + 1 - \frac{1}{2}\left(1 - \frac{H_{k,n}^2}{H_{k,n}^{(2)}}\right)^{-1} \tag{1.40}$$

where

$$H_{k,n}^{(2)} = \frac{1}{k}\sum_{j=1}^{k}\left(\log X_{n-j+1,n} - \log X_{n-k,n}\right)^2 . \tag{1.41}$$

(iii) Generalized Pareto quantile-plot

The Generalized Pareto quantile plot is ultimately linear and the slope of the plot is an estimate for γ. The Generalized Pareto quantile plot is given by

$$\left(\log\left(\frac{n+1}{k+1}\right), \log\left(X_{n-k,n}H_{k,n}\right)\right), k = 1,...,n-1 . \tag{1.42}$$

1.1.4.3 Peaks over threshold methods

The following equation gives the conditional survival function of the exceedances (peaks or excesses) , $Y = X - t$, over a threshold t :

$$\bar{F}_t(y) \sim \left(1 + \frac{y\gamma}{b(t)}\right)^{\frac{-1}{\gamma}} \tag{1.43}$$

where F_t is a conditional distribution (Balkema and de Haan 1974; Pickands 1975).

If $b(t)$ is regarded as a scale parameter σ, it leads to a Generalized Pareto distribution (GPD), given by the following equations:

$$\begin{cases} 1-\left(1+\dfrac{y\gamma}{\sigma}\right)^{\frac{-1}{\gamma}}, y \in (0,\infty) \text{ if } \gamma > 0 \\ 1-\exp\left(\dfrac{-y}{\sigma}\right), y \in (0,\infty) \text{ if } \gamma = 0 \\ 1-\left(1+\dfrac{y\gamma}{\sigma}\right)^{\frac{-1}{\gamma}}, y \in \left(0,\dfrac{-\sigma}{\gamma}\right) \text{ if } \gamma < 0 \end{cases} \quad (1.44)$$

which are fitted to exceedances over a high threshold value.

If the value of t is known and the number of observations from a sample $X_1,...,X_n$, that exceeds the threshold t, is given by N_t, the parameters γ and σ can be estimated. The estimation of the parameters can be done by using the maximum likelihood method, the method of moments or the Bayesian method where Monte Carlo Markov Chain (MCMC) procedures are used. The estimation of these parameters is discussed in Chapter 4.

1.2 Multivariate Extreme Value Theory

Multivariate Extreme Value Theory is a rapidly growing field because many problems involving extreme events are necessarily multivariate. A question to ask is how does extremes in one variable relate to extremes in another variable? In other words, what is the dependence structure between variables and how can one estimate this dependence?

The study of multivariate extremes can be split into two phases. Firstly the marginal distribution phase, where each margin is modelled using the

techniques developed in the univariate case; and secondly the dependence structure phase, where an inter-relationship is studied.
The problem with multivariate extremes is that the possible limiting dependence structure cannot be captured in a finite-dimensional parametric family.

1.2.1 Multivariate domain of attraction

There are various ways to order multivariate observations, here the marginal ordering is considered.

Consider a sample of *d*-dimensional observations, $\mathbf{X}_i = \left(X_{i,1},...,X_{i,d}\right)$, $i = 1,...,n$. $\mathbf{M}_n$ is defined as the sample maxima which is a vector of component-wise maxima, where the components of $\mathbf{M}_n = \max\left(X_{ij}\right), i = 1,...,n;\ j = 1,...,d$.

The distribution function of the component-wise maximum, $\mathbf{M}_n$, from a distribution function F is given by

$$P\left[\mathbf{M}_n \le \mathbf{x}\right] = P\left[\mathbf{X}_1 \le \mathbf{x},...,\mathbf{X}_n \le \mathbf{x}\right] = F^n\left(\mathbf{x}\right), \mathbf{x} \in \mathbb{R}^d. \qquad (1.45)$$

Again a non-trivial limit distribution must be found. Therefore, sequences of vectors, $\left(\mathbf{a}_n\right)_n$ and $\left(\mathbf{b}_n\right)_n$, where $\mathbf{a}_n > \mathbf{0} = \left(0,...,0\right)$ such that $\mathbf{a}_n^{-1}\left(\mathbf{M}_n - \mathbf{b}_n\right)$ converges to a non-degenerate limit, must be found. These sequences must be chosen so that there exists a *d*-variate distribution function G with non-degenerate margins for which

$$F^n\left(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n\right) \xrightarrow{D} G\left(\mathbf{x}\right), n \to \infty. \qquad (1.46)$$

G is called a multivariate extreme value distribution function.

Let F_j and G_j denote the j^{th} marginal distribution functions of F and G respectively and assume G_j is non-degenerate. A sequence of random vectors can only converge in distribution if the corresponding marginal sequences do. Therefore, for $j = 1,...,d$

$$F_j^n\left(a_{n,j}x_j + b_{n,j}\right) \xrightarrow{D} G_j\left(x_j\right), n \to \infty. \tag{1.47}$$

Each G_j is by itself a univariate extreme value distribution function with F_j in its domain of attraction.

Next, the dependence aspects of the marginal distribution must be isolated. This follows from Kotz and Nadarajah (2000, p. 96-97). The components of F and G are now transformed to standard marginal distributions. For convenience the standard marginal distribution is the unit Fréchet distribution with distribution function $\exp\left(-y^{-1}\right), y > 0$. This is also denoted by $\Phi_1\left(y\right)$ where y denotes the Fréchet random variables. The transformation is shown in (1.48). If G is a multivariate distribution function with continuous marginals then the transformation is given by

$$G_*\left(y_1,...,y_d\right) = G\left(\left(\frac{1}{\left(-\log G_1\right)}\right)^{\leftarrow}\left(y_1\right),...,\left(\frac{1}{\left(-\log G_d\right)}\right)^{\leftarrow}\left(y_d\right)\right), y_1 \geq 0,..., y_d \geq 0, \tag{1.48}$$

where $\leftarrow$ indicates the inverse of the function in brackets. G_* has marginal distributions $G_{*j}\left(y\right) = \Phi_1\left(y\right)$. G is a multivariate extreme value distribution if and only if G_* is also a multivariate extreme value distribution and $F \in D\left(G\right)$ if and only if $F_* \in D\left(G_*\right)$.

1.2.2 Characterizations of the Domain of attraction

First the point process characterization by de Haan (1985) is discussed. This discussion follows from Kotz and Nadarajah (2000, p. 97-98).

Let $\left(X_{i,1},...,X_{i,d}\right), i=1,...,n$ be i.i.d. *d*-variate random vectors with common joint distribution function $F \in D(G)$. Define *T* as follows:

$$T\left(y_1,...,y_d\right)=\left(\sum_{j=1}^{d} y_j, \frac{y_1}{\sum_{j=1}^{d} y_j},...,\frac{y_{d-1}}{\sum_{j=1}^{d} y_j}\right) \tag{1.49}$$

where *y* denotes Fréchet variables.

Let

$$S_d=\left\{\left(\omega_1,...,\omega_{d-1}\right): \sum_{j=1}^{d-1}\omega_j \le 1, \omega_j \ge 0,\ j=1,...,d-1\right\} \tag{1.50}$$

be the $(d-1)$-dimensional unit simplex. The point-process is then given by the following equation:

$$P_n=\left\{\left(\left(Y_{i,1}\right)/n,...,\left(Y_{i,d}\right)/n\right), i=1,...,n\right\} \to P \tag{1.51}$$

as $n \to \infty$. P is a non-homogenous Poisson process with intensity measure μ_* where μ_* satisfies

$$\mu_* \circ T^{\leftarrow}\left(\partial r, \partial \boldsymbol{\omega}\right)=r^{-2}\partial r H_*\left(\partial \boldsymbol{\omega}\right), r>0, \boldsymbol{\omega} \in S_d. \tag{1.52}$$

H_* is a non-negative measure on S_d with

$$H_*(S_d) = d \text{ and } \int_{S_d} \omega_j H_*(d\boldsymbol{\omega}) = 1,\ j = 1,...,p-1. \tag{1.53}$$

From the preceding the following equation can be derived:

$$G_*(y_1,...,y_d) = \exp\left(-V(y_1,...,y_d)\right) \tag{1.54}$$

where

$$\begin{aligned} V(y_1,...,y_d) &= \mu_*\left(\left([0,y_1]\times...\times[0,y_d]\right)^c\right) \\ &= \int_{S_d} \max\left(\frac{\omega_1}{y_1},...,\frac{1-\sum_{j=1}^{d-1}\omega_j}{y_d}\right) H_*(d\boldsymbol{\omega}). \end{aligned} \tag{1.55}$$

V is called the exponent measure function (Kotz and Nadarajah 2000, p. 97-98).

Another point process characterization, discussed by Kotz and Nadarajah (2000, p. 102), is when the marginal variables are linearly ordered. Suppose that the joint distribution function F of (X_1, X_2) belongs to the domain of attraction G. Assume now the following order $X_1 \le X_2 \le mX_1, m > 1$, then H_* in (1.52), defined on $S_2 = [0,1]$ is concentrated in the interval

$$\left[\liminf_{y\to\infty}\left\{\frac{l(y)}{y+l(y)}\right\}, \limsup_{y\to\infty}\left\{\frac{y}{y+g(y)}\right\}\right] \tag{1.56}$$

with

$$\liminf_{y\to\infty}\left\{\frac{g(y)}{y}\right\} \le 1 \le \limsup_{y\to\infty}\left\{\frac{y}{l(y)}\right\}, g(y) = U_2\left\{U_1^{\leftarrow}(y)\right\} \text{ and } l(y) = U_1\left\{m^{-1}U_2^{\leftarrow}(y)\right\}$$

where $U_j\left(X_j\right) = -\dfrac{1}{\log F_j\left(X_j\right)} = Y_j$, $j = 1,...,d$.

Linear ordering between the marginal variables reduce the domain of H_* to $[a,b]$, where $a \le \dfrac{1}{2}$ and $b \ge \dfrac{1}{2}$ (Kotz and Nadarajah 2000, p. 102).

Nadarajah *et al*. (1998) and Kotz and Nadarajah (2000 p. 102) show that if $H_*^\dagger$ is an absolutely continuous positive measure on $[0,1]$ that satisfies (1.53), and $h^\dagger$ is the density of $H_*^\dagger$, then the subinterval $[a,b]$ of $[0,1]$ with $a \le \dfrac{1}{2} \le b$ defines a measure H_* on $[a,b]$ as follows:

Let H_* have atoms of mass

$$\begin{aligned} &H_*\left(\{a\}\right) = \gamma_1, \\ &H_*\left(\{b\}\right) = \gamma_2, \text{ at } a \text{ and } b \text{ respectively} \end{aligned} \tag{1.57}$$

where

$$\begin{aligned} &0 \le \gamma_1 \le \frac{2b-1}{b-a}, \\ &0 \le \gamma_2 \le \frac{1-2a}{b-a}, \end{aligned} \tag{1.58}$$

and let H_* be continuous in the interior $[a,b]$, then the density of $h(\omega)$ is

$$h(\omega) = \frac{(b-a)(\alpha\beta)^2}{\left\{\alpha(\omega-a)+\beta(b-\omega)\right\}^3} h^\dagger\left\{\frac{\alpha(\omega-a)}{\alpha(\omega-a)+\beta(b-\omega)}\right\}, \omega \in (a,b) \tag{1.59}$$

where $\alpha = 2b-1+\gamma_1(a-b)$ and $\beta = 1-2a-\gamma_2(a-b)$.

1.2.3 Characterizations of Multivariate Extreme Value Distributions

Tiago de Oliveira (1958) introduced the following characterization:

$$G_*(y_1, y_2) = \left\{\Phi_1(y_1)\Phi_1(y_2)\right\}^{\nu(\log y_2 - \log y_1)} \tag{1.60}$$

where ν is the dependence function.

It is shown by Obretenov (1991) that ν is related to H_* by the following equation:

$$\nu\left(\log\frac{y}{1-y}\right) = \int_{[0,1]} \max\left\{\omega(1-y), (1-\omega)y\right\} H_*(\partial\omega). \tag{1.61}$$

Nadarajah *et al*. (2000, p. 106) shows that, when more than two variables are considered, (1.60) and (1.61) can be generalized to

$$G_*(y_1, ..., y_d) = \left\{\Phi_1(y_1)...\Phi_1(y_d)\right\}^{\nu(\log y_2 - \log y_1, .., \log y_d - \log y_1)} \tag{1.62}$$

with

$$\nu\left(\log\frac{y_1}{y_2}, ..., \log\frac{y_1}{y_d}\right) = \int_{S_d} \max\left(\frac{\omega_1 y_1}{\sum_{j=1}^{d} y_j}, ..., \frac{\omega_d y_d}{\sum_{j=1}^{d} y_j}\right) H_*(\partial\boldsymbol{\omega}) \tag{1.63}$$

where S_d is the $(d-1)$-dimensional unit simplex.

Pickands (1981) gives as alternative way of writing the point process in (1.54): For $d = 2$,

$$G_*(y_1, y_2) = \exp\left\{-\left(\frac{1}{y_1}+\frac{1}{y_2}\right)A\left(\frac{y_1}{y_1+y_2}\right)\right\}, \qquad (1.64)$$

A is the dependence function. A is related to H_* by the following equation:

$$A(\omega) = \int_{[0,1]} \max\left\{\omega(1-q),(1-\omega)q\right\} H_*(\partial q). \qquad (1.65)$$

Section 3.4 and 3.5 of Extreme Value Distributions, Theory and Applications by Kotz and Nadarajah (2000), discusses various parametric families for bivariate and multivariate extreme value distributions.

In Chapter 5 the work of Boldi & Davison (2006) is discussed, where a mixture of Dirichlet distributions are used to model H_*.

The rest of this thesis is laid out as follows: In Chapter 2 the Multivariate Generalized Burr-Gamma distribution is discussed as a model for modelling multivariate extreme values. In Chapter 3 multivariate extreme values are modelled through multivariate regression. Chapter 4 discusses different methods on how a threshold value *t* can be chosen when the interest is only in modelling the data above a threshold. In Chapter 5 multivariate extreme values are modelled through copula functions and mixtures of Dirichlet distributions. Chapter 6 discusses the modelling of ordered multivariate extreme values, and Chapter 7 concludes the thesis.

Chapter 2

The Multivariate Generalized Burr-Gamma distribution

2.1 Introduction

In this chapter the Multivariate Generalized Burr-Gamma (MGBG) distribution is discussed with some applications.

In Section 2.2 the model of the Univariate Generalized Burr-Gamma (GBG) distribution is defined and some special cases of the GBG distribution are discussed.

In Section 2.3 the model of the Multivariate Generalized Burr-Gamma (MGBG) distribution with some properties is defined and again, some special cases of this distribution are discussed.

In Section 2.4 the MGBG distribution is applied to a simulated data set and different approaches are followed to estimate the parameter values. This section also considers the goodness of fit for the different estimation approaches with a comparison between the different estimation approaches.

In Section 2.5 the MGBG distribution is applied to a real data set: the maximum monthly inflow of water into the Gariep Dam. Again the estimation of the parameters is discussed together with the goodness of fit. Predictions, of maximum monthly inflow values, are also discussed.

2.2 The Generalized Burr-Gamma (GBG) distribution

The GBG distribution is now defined. The GBG distribution models all the data.

Definition 2.1 (Beirlant *et al*. 2002, p. 115): A random variable *X* is GBG(k, μ, σ, ξ) distributed when the distribution function is given by

$$F(x) = P(X \le x) = \frac{1}{\Gamma(k)} \int_0^{\upsilon_\xi(x)} e^{-u} u^{k-1} du \qquad (2.1)$$

where

$$\upsilon_\xi(x) = \frac{1}{\xi} \log\left(1 + \xi \nu(x)\right) > 0,$$

$$\nu(x) = \left(x e^{\alpha}\right)^{\frac{1}{\beta}}$$

or

$$= e^{\left\{\psi(k) + \frac{\log x + \mu}{\sigma}\sqrt{\psi'(k)}\right\}}.$$

$\psi(k) = \frac{\partial}{\partial k} \log \Gamma(k)$ and $\psi'(k) = \frac{\partial}{\partial k}\psi(k)$ represent the digamma and trigamma functions respectively.

The parameterizations (α, β, ξ) and (μ, σ, ξ) are linked through the following: $\mu = \alpha - \beta\psi(k)$ and $\sigma = \beta\sqrt{\psi'(k)}$.

The parameter space is defined as $\Omega = \{-\infty < \mu < \infty, \sigma > 0, k > 0, -\infty < \xi < \infty\}$.

If $V_\xi = \frac{1}{\xi}\log(1 + \xi V)$ and $V = \left(X e^{\alpha}\right)^{\frac{1}{\beta}}$ then it follows from (2.1) that $V_\xi \sim \text{GAM}(k,1)$ (Beirlant *et al*. 2002, p. 115).

2.2.1 Special cases

In Beirlant *et al*. (2002, p. 115-117) it is shown that the following special cases can occur with the GBG distribution:

(i) If $\xi = 0$, then $V_\xi = V \sim \text{GAM}(k,1)$ and X is generalized gamma distributed, $X \sim \text{GGAM}(k, \mu, \sigma)$.

(ii) If k = 1 and $\sigma = \sqrt{\psi'(1)} = \frac{\pi}{\sqrt{6}}$ or if k = 1 and $\beta = 1$, then Y is Generalized Pareto distributed (GPD). Y represents the excesses of the observations X exceeding some threshold u, thus $Y = X-u$. The distribution function of Y is given by

$$H(y) = P(X < y+u \mid X > u) = 1 - \left\{1 + \frac{\xi y}{\hat{\sigma}}\right\}^{\frac{-1}{\xi}}, y > 0 \quad (2.2)$$

where ξ is the main parameter, called the extreme value index, and $\hat{\sigma}$ is dependent on u and describes the variation.

(iii) If k = 1 and $\beta \neq 1$, X follows a Burr distribution given by the following equation:

$$P(X > x) = P(V_\xi > \nu_\xi(x)) = \exp(-\nu_\xi(x)) = (1+\xi\nu(x))^{\frac{-1}{\xi}} \quad (2.3)$$

where

$$\nu_\xi(x) = \frac{1}{\xi}\log(1+\xi\nu(x)) \text{ and } \nu(x) = x^{\frac{1}{\beta}} e^{\frac{\alpha}{\beta}}.$$

2.3 The Multivariate Generalized Burr-Gamma distribution

The Multivariate Generalized Burr-Gamma distribution (MGBG) is an extension of the univariate generalized Burr-Gamma family of distributions and it allows the modelling of multivariate data that includes extreme values.

Definition 2.2 (Beirlant *et al*. 2000, p. 113): A random vector $\underline{X} = \left(X_1,...X_p\right)'$ is MGBG($\underline{k},\underline{\mu},\sum,\underline{\xi}$) distributed if the joint density function is given by the following equation:

$$f(\underline{x}) = \frac{1}{|\sum|^{\frac{1}{2}}} \prod_{i=1}^{p} \frac{\sqrt{\psi'(k_i)}}{\Gamma(k_i)x_i} \exp\left(-\nu_{\xi_i}(\underline{x})\right) \nu_{\xi_i}(\underline{x})^{k_i-1} \left(1+\xi_i \nu_i(\underline{x})\right)^{-1} \nu_i(\underline{x}),\ 1+\xi_i\nu_i(\underline{x}) > 0,$$
$$i = 1,..., p \qquad (2.4)$$

where

$$V_{\xi_i} = \frac{1}{\xi_i}\log\left(1+\xi_i V_i\right) \sim \text{GAM}(k_i,1),\ \text{independent for } i = 1,..., p$$
$$V_i = \exp\left\{\psi(k_i) - \sqrt{\psi'(k_i)}\,\sum_{(i)}^{\frac{-1}{2}}\left(\underline{Y}-\underline{\mu}\right)\right\},$$
$$\underline{Y} = \left(Y_1,...,Y_p\right)',\ Y_i = -\log\left(X_i\right),$$
$$\psi(k_i) = \frac{\partial}{\partial k_i}\log\Gamma(k_i) \text{ and } \psi'(k_i) = \frac{\partial}{\partial k_i}\psi(k_i).$$

$\psi(k_i)$ and $\psi'(k_i)$ represent the digamma and trigamma functions respectively. For dimension p, $\sum_{(i)}^{\frac{-1}{2}}$ is the i^{th} row of the symmetric square root matrix $\sum^{-1}$, where $\sum$ is a symmetric positive definite $p \times p$ matrix. $\underline{k} = \left(k_1,...,k_p\right)$ denotes a positive vector of shape parameters, $\underline{\mu} = \left(\mu_1,...,\mu_p\right)$ denotes a vector of location parameters and $\underline{\xi} = \left(\xi_1,...,\xi_p\right)$ denotes a vector of extreme value indices. The parameter space is defined as

$\Omega = \{k_i > 0, -\infty < \mu_i < \infty, \sum > 0, -\infty < \xi_i < \infty\}$, $i = 1,...p$.

2.3.1 Properties of the MGBG

(i) The expected value of $\underline{X}$ is stated and proven in Approximation 2.1.

(ii) The variance of $\underline{X}$ is stated and proven in Approximation 2.2.

Approximation 2.1

The expected value of $X_i, i = 1,..., p$, where $V_{\xi_i} = \frac{1}{\xi_i}\log(1+\xi_i V_i)$~GAM$(k_i,1)$ is given by the approximation

$$E(X_i) \approx \exp(-\mu_{Y_i}) + \frac{1}{2}\sigma_{Y_i}^2 \exp(-\mu_{Y_i}). \qquad (2.5)$$

Proof:

From Definition 2.2 $V_i = \frac{e^{\xi_i V_{\xi_i}} - 1}{\xi_i}$, therefore $E\left[\log\left(\frac{e^{\xi_i V_{\xi_i}} - 1}{\xi_i}\right)\right] = E\left[\log(V_i)\right]$.

To obtain the $E\left[\log(V_i)\right]$ the delta method given by Rice (1995, p. 149) is used. This method involves the expansion of the Taylor series to the second order to improve the approximation. Therefore

$$E(\log(V_i)) \approx g(\mu_{V_{\xi_i}}) + \frac{1}{2}\sigma_{V_{\xi_i}}^2 g''(\mu_{V_{\xi_i}}),$$

where

$$\mu_{V_{\xi_i}} = k_i,$$

$$\sigma_{V_{\xi_i}}^2 = k_i,$$

$$g(\mu_{V_{\xi_i}}) = \log\left(\frac{e^{\xi_i V_{\xi_i}} - 1}{\xi_i}\right),$$

$$g'\left(\mu_{V_{\xi_i}}\right)=\frac{\xi_i e^{\xi_i V_{\xi_i}}}{e^{\xi_i V_{\xi_i}}-1}$$

and

$$g''\left(\mu_{V_{\xi_i}}\right)=\frac{-\xi_i^2 e^{\xi_i V_{\xi_i}}}{\left(e^{\xi_i V_{\xi_i}}-1\right)^2}.$$

Therefore

$$E\left(\log\left(V_i\right)\right)\approx\log\left(\frac{e^{\xi_i k}-1}{\xi_i}\right)-\frac{k_i\xi_i^2 e^{\xi_i k_i}}{2\left(e^{\xi_i k_i}-1\right)^2}. \tag{2.6}$$

The delta method also indicates that $Var\left(\log V\right)\approx\sigma^2_{V_{\xi_i}}\left(g'\left(\mu_{V_{\xi_i}}\right)\right)^2$, therefore

$$Var\left(\log V_i\right)\approx k_i\left(\frac{\xi_i e^{\xi_i k_i}}{e^{\xi_i k_i}-1}\right)^2. \tag{2.7}$$

From Definition 2.2 $Y_i=-\log\left(X_i\right)$, therefore $X_i=g\left(Y_i\right)=\exp\left(-Y_i\right)$, where

$$\begin{aligned} Y_i &= -\sum_i^{\frac{1}{2}} D_\psi^{-1}\begin{pmatrix}\log\left(\dfrac{e^{\xi_1 V_{\xi_1}}-1}{\xi_1}\right)-\psi\left(k_1\right)\\ .\\ .\\ .\\ \log\left(\dfrac{e^{\xi_p V_{\xi p}}-1}{\xi_p}\right)-\psi\left(k_p\right)\end{pmatrix}+\mu_i \\ &= -\sum_i^{\frac{1}{2}} D_\psi^{-1}\begin{pmatrix}\log\left(V_1\right)-\psi\left(k_1\right)\\ .\\ .\\ .\\ \log\left(V_p\right)-\psi\left(k_p\right)\end{pmatrix}+\mu_i \end{aligned} \tag{2.8}$$

and

$$D_{\psi} = \text{diag}\left(\sqrt{\psi'(k_1)}, \sqrt{\psi'(k_2)}, ..., \sqrt{\psi'(k_p)}\right). \tag{2.9}$$

Again the delta method is used to obtain an approximation for $E(X_i)$ (Rice 1995, p. 149). Thus, when following the delta method, the following is obtained:

$$\begin{aligned} &E(X_i) = g(\mu_{Y_i}) + \frac{1}{2}\sigma_{Y_i}^2 g''(\mu_{Y_i}), \\ &g'(Y_i) = -\exp(-Y_i) \text{ and } g''(Y_i) = \exp(-Y_i). \end{aligned} \tag{2.10}$$

Therefore

$$E(X_i) \approx \exp(\mu_{Y_i}) + \frac{1}{2}\sigma_{Y_i}^2 \exp(\mu_{Y_i}) \tag{2.11}$$

and

$$\mu_{Y_i} = E(Y_i) = -\sum_{(i)}^{\frac{1}{2}} D_{\psi}^{-1} E\begin{bmatrix} \log V_1 \\ . \\ . \\ . \\ \log V_p \end{bmatrix} + \sum_{(i)}^{\frac{1}{2}} D_{\psi}^{-1} \begin{bmatrix} \psi(k_1) \\ . \\ . \\ . \\ \psi(k_p) \end{bmatrix} + \mu_i \tag{2.12}$$

where

$$E(\log(V_i)) \approx \log\left(\frac{e^{\xi_i k} - 1}{\xi_i}\right) - \frac{k_i \xi_i^2 e^{\xi_i k_i}}{2\left(e^{\xi_i k_i} - 1\right)^2}. \tag{2.13}$$

$\sigma_{Y_i}^2$ is the diagonal of

$$Cov\left(Y_i, Y_i'\right) = \sum_i^{\frac{1}{2}} D_\psi^{-1} Cov\left(\log V_i, \log V_i'\right) D_\psi^{-1} \sum_i^{\frac{1}{2}}$$

$$= \sum_i^{\frac{1}{2}} D_\psi^{-1} H D_\psi^{-1} \sum_i^{\frac{1}{2}}. \qquad (2.14)$$

$Var\left(\log V_i\right) \approx k_i \left(\frac{\xi_i e^{\xi_i k_i}}{e^{\xi_i k i} - 1} \right)^2$ and $Cov\left(\log V_i, \log V_j\right), i = 1,..., p; j = 1,..., p, i \neq j$ must be known to calculate *H*. Because V_{ξ_i} and V_{ξ_j} are independent, V_i and $V_j, i \neq j$ are also independent and the $Cov\left(\log V_i, \log V_j\right) = h_{ij} = 0$. $\sigma_{Y_i}^2$ is now a diagonal matrix with $Var\left(\log V_i\right)$ given on the diagonal.

Therefore $E\left(X_i\right)$ is equal to

$$\left[\exp\left(\sum_I^{\frac{1}{2}} D_\psi^{-1} \begin{bmatrix} \log\left(\frac{e^{\xi_1 k_1} - 1}{\xi_1}\right) - \frac{k_1 \xi_1^2 e^{\xi_1 k_1}}{2\left(e^{\xi_1 k_1} - 1\right)^2} \\ . \\ . \\ . \\ \log\left(\frac{e^{\xi_p k_p} - 1}{\xi_p}\right) - \frac{k_p \xi_p^2 e^{\xi_p k_p}}{2\left(e^{\xi_p k_p} - 1\right)^2} \end{bmatrix} \right) - \sum_I^{\frac{1}{2}} D_\psi^{-1} \begin{bmatrix} \psi_1 \\ . \\ . \\ . \\ \psi_p \end{bmatrix} - \mu_i \right] + \frac{1}{2} \sigma_{Y_i}^2 \left[\exp\left(\sum_I^{\frac{1}{2}} D_\psi^{-1} \begin{bmatrix} \log\left(\frac{e^{\xi_1 k_1} - 1}{\xi_1}\right) - \frac{k_1 \xi_1^2 e^{\xi_1 k_1}}{2\left(e^{\xi_1 k_1} - 1\right)^2} \\ . \\ . \\ . \\ \log\left(\frac{e^{\xi_p k_p} - 1}{\xi_p}\right) - \frac{k_p \xi_p^2 e^{\xi_p k_p}}{2\left(e^{\xi_p k_p} - 1\right)^2} \end{bmatrix} \right) - \sum_I^{\frac{1}{2}} D_\psi^{-1} \begin{bmatrix} \psi_1 \\ . \\ . \\ . \\ \psi_p \end{bmatrix} - \mu_i \right]$$

(2.15)

#

Approximation 2.2

The variance of $X_i, i = 1,..., p$, where $V_{\xi_i} = \frac{1}{\xi_i} \log\left(1 + \xi_i V_i\right)$~GAM$\left(k_i, 1\right)$ is given by the approximation

$$Var\left(X_i\right) \approx \sum_{j=1}^{p} k_j \left[\exp\left(\sum\nolimits_i^{\frac{1}{2}} D_\psi^{-1} \begin{pmatrix} \log\left(\frac{e^{\xi_1 V_{\xi_1}} - 1}{\xi_1}\right) - \psi\left(k_1\right) \\ . \\ . \\ . \\ \log\left(\frac{e^{\xi_p V_{\xi_p}} - 1}{\xi_p}\right) - \psi\left(k_p\right) \end{pmatrix} + \mu_i \right) \left(\left[\sum\nolimits_i^{\frac{1}{2}} D_\psi^{-1} \left(\frac{\xi_j^{\,2}}{e^{\xi_j k_j} - 1} \right) e^{\xi_j k_j} \right] \right) \right]^2 \tag{2.16}$$

<u>Proof:</u>

From Approximation 2.1 Y_i is defined as $Y_i = -\sum_i^{\frac{1}{2}} D_\psi^{-1} \begin{pmatrix} \log\left(V_1\right) - \psi\left(k_1\right) \\ . \\ . \\ . \\ \log\left(V_p\right) - \psi\left(k_p\right) \end{pmatrix} + \mu_i$.

Let $a_i = \sum_i^{\frac{1}{2}} D_\psi^{-1}$ and $b = \begin{pmatrix} \log\left(V_1\right) - \psi\left(k_1\right) \\ . \\ . \\ . \\ \log\left(V_p\right) - \psi\left(k_p\right) \end{pmatrix}$, then $Y_i = -a_i b + \mu_i$.

In Approximation 2.1 it was shown that $X_i = \exp\left(-Y_i\right), i = 1,..., p$, therefore

$$\begin{aligned} X_i &= \exp\left(\sum_{j=1}^{p} a_{ij} b_j - \mu_i \right) \\ &= g\left(V_{\xi_1},...,V_{\xi_p}\right). \end{aligned} \tag{2.17}$$

When applying the delta method

$$X_i \approx g\left(\mu_{V_\xi}\right) + \sum_{j=1}^{p}\left(V_{\xi_j} - \mu_{V_{\xi_j}}\right)\frac{\partial g\left(\mu_{V_\xi}\right)}{\partial V_{\xi_j}}. \qquad (2.18)$$

Now it follows that:

$$Var\left(X_i\right) \approx \sum_{j=1}^{p} Var\left(V_{\xi_j}\right)\left(\frac{\partial g\left(\mu_{V_\xi}\right)}{\partial V_{\xi_j}}\right)^2 + Cov\left(V_{\xi_i}, V_{\xi j}\right), \text{ where } i \neq j. \qquad (2.19)$$

$Cov\left(V_{\xi_i}, V_{\xi j}\right) = 0$, because V_{ξ_i} and $V_{\xi j}$ are independently distributed, and $E\left(V_{\xi_i}\right) = \mu_{V_\xi} = k_i$, because $V_{\xi_i} = \frac{1}{\xi_i}\log\left(1+\xi_i V_i\right)$~GAM$\left(k_i, 1\right)$.

Therefore $Var\left(X_i\right)$ is approximately

$$\sum_{j=1}^{p} k_j \left[\exp\left(\sum_i^{\frac{1}{2}} D_\psi^{-1}\begin{pmatrix} \log\left(\frac{e^{\xi_1 k_1}-1}{\xi_1}\right) - \psi\left(k_1\right) \\ . \\ . \\ . \\ \log\left(\frac{e^{\xi_p k_p}-1}{\xi_p}\right) - \psi\left(k_p\right) \end{pmatrix} + \mu_i\right)\left(\left[\sum_i^{\frac{1}{2}} D_\psi^{-1}\left(\frac{\xi_j^{\,2}}{e^{\xi_j k_j}-1}\right)e^{\xi_j k_j}\right]\right)\right]^2 \qquad (2.20)$$

#

2.3.2 Special cases

In Beirlant *et al*. (2002, p. 114 -116) it is shown that the following special cases can occur with the MGBG distribution:

(i) If $\underline{\xi} = \underline{0}$, $V_{\xi_i} = V_i \sim \text{GAM}(k_i, 1)$ are independent variables. $\underline{X}$ is then multivariate generalized gamma distributed, MGG($\underline{k}, \underline{\mu}, \sum$), and $\underline{Y}$ = (-logX_1,…,-logX_p)' is multivariate generalized extreme value distributed (MGEV). The reasoning behind this is because for p = 1 the type I extreme value distribution is obtained, EV(α, β), where $\alpha = \gamma\beta - \mu$ and $\beta = \frac{\sqrt{6}\sigma}{\pi}$. γ denotes Euler's constant.

(ii) If $\underline{\xi} = \underline{0}$ and $\underline{k} = (1,...,1)$, $\underline{X}$ is multivariate Weibull distributed, MWEI($\underline{\mu}, \sum$). For p = 1, $\xi = 0$ and k = 1 the MGBG becomes the Weibull distribution, WEI(α, β), where $\alpha = -\gamma\beta - \mu$ and $\beta = \frac{\sqrt{6}\sigma}{\pi}$.

(iii) If $\underline{\xi} > \underline{0}$ and $\underline{k} = (1,...,1)$, $\underline{X}$ is multivariate Burr distributed and the marginals are univariate Burr distributions. For the event $V_i > t_i$ for some threshold t_i, the joint survival function of the excesses $Z_i = V_i - t_i$, as $t_i \to \infty$, becomes

$$\begin{aligned} \bar{H}(\underline{z} \mid \underline{t}) &= P(V_1 > \nu_1, ..., V_p > \nu_p \mid V_i > t_i) \\ &= \prod_{i=1}^{p} \left\{ 1 + \frac{\xi_i z_i}{\tilde{\sigma}_i} \right\}^{\frac{-1}{\xi_i}} \end{aligned} \qquad (2.21)$$

where

$$\tilde{\sigma}_i = \xi_i t_i, \; i = 1, ..., p.$$

For p = 1, $\bar{H}$ becomes the survival function of the Generalized Pareto distribution, discussed in Section 2.2.1.

2.4 A simulated example

The following process shows how to simulate an observation $\underline{x} = (x_1, ..., x_p)$ from a MGBG($\underline{k}, \underline{\mu}, \sum, \underline{\xi}$):

(i) Select $\nu_{\xi_i} \in \text{GAM}(k_i, 1), i = 1, ..., p.$

(ii) Compute $\nu_i = \frac{1}{\xi}\left\{\exp\left(\xi_i \nu_{\xi_i}\right) - 1\right\}, i = 1, ..., p$.

(iii) Compute y_i where

$$y_i = -\frac{1}{\sqrt{\psi'(\underline{k})}} \sum_{(i)}^{\frac{1}{2}} \left\{\log \underline{\nu} - \psi(\underline{k})\right\} + \mu_i, i = 1, ..., p$$

where $\log \underline{\nu} = (\log \nu_i, ..., \log \nu_p)'$

and $\psi(\underline{k}) = \left(\psi(k_1), ..., \psi(k_p)\right)'$.

(iv) Compute $x_i = \exp(-y_i), i = 1, ..., p.$

Using the above process an observation $\underline{x} = (x_1, ..., x_p)$, with p = 3, $\underline{k} = (3; 2,1; 2,6)$, $\underline{\mu} = (-3,9; -3,8; -3,6)$, $\underline{\xi} = (0,9; 0,1; 0,5)$ and $\sum = \begin{bmatrix} 0,014 & 0,004 & 0,007 \\ 0,004 & 0,007 & 0,004 \\ 0,007 & 0,004 & 0,013 \end{bmatrix}$, is simulated from a MGBG($\underline{k}, \underline{\mu}, \sum, \underline{\xi}$). This is repeated n = 700 times.

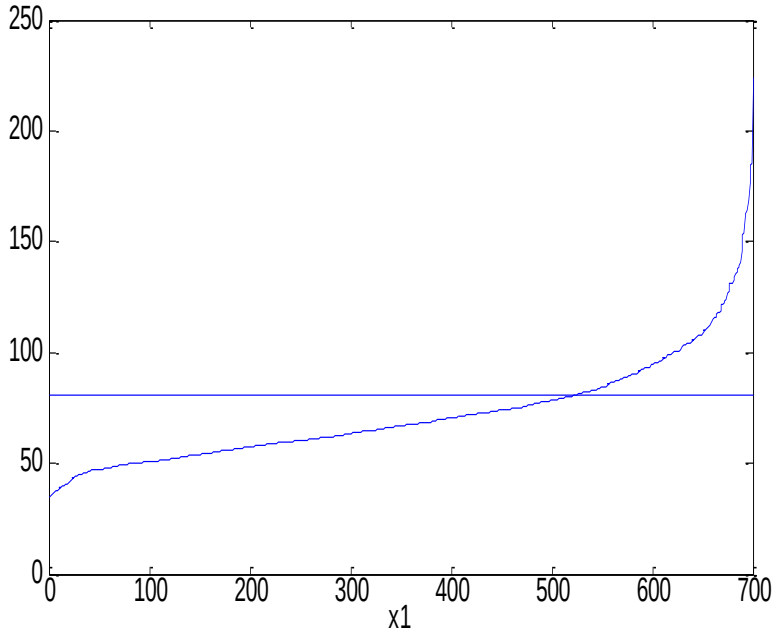
2.4.1 Estimating the parameters

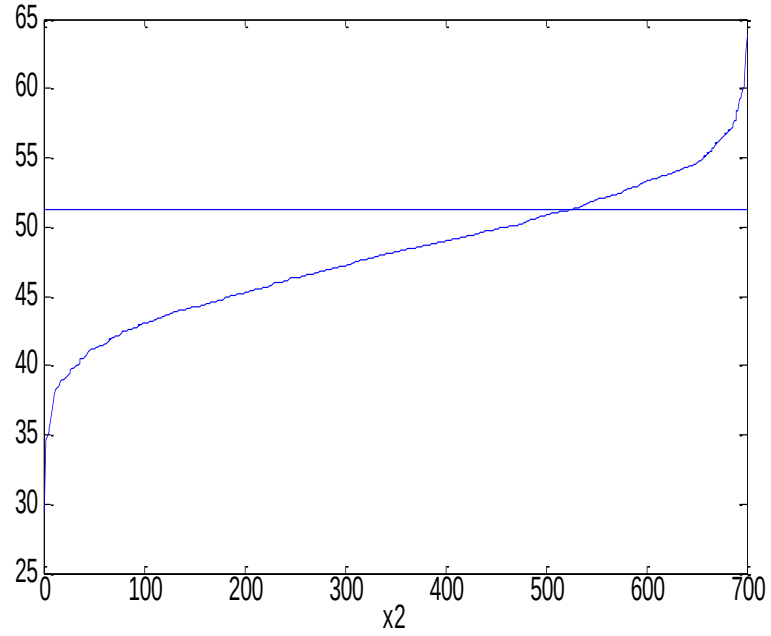
In the following sections three different methods for estimating the parameters are discussed.

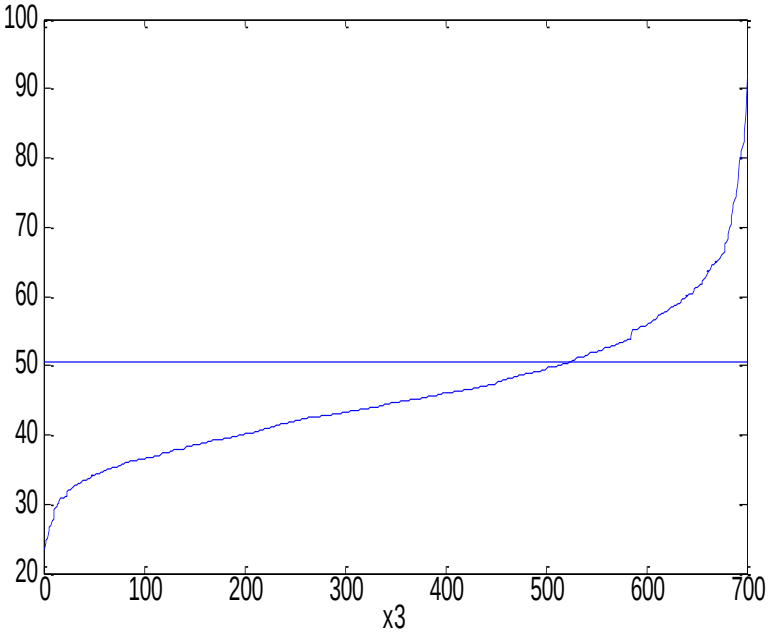
2.4.1.1 Method of Moments

The four parameters from the random sample X_{ij}, i = (1,..,p), j = (1,…,n) simulated from a MGBG distribution are now estimated. First it is assumed that $\xi_i = 0, i = 1,..., p$, thus the extreme are ignored for now. In this case a trimmed dataset is considered, thus only data below a threshold u_i is considered. Let $U = \left\{ x_{ij} : x_{1j} < u_1, x_{2j} < u_2, ..., x_{pj} < u_p, i = 1,..., p; j = 1,..., n \right\}$ be the trimmed dataset and n_u the sample size of U. The values of $u_i, i = 1,..., p$ are obtained by calculating the upper 75% quantile of each variable $X_{ij}, i = 1,..., p; j = 1,..., n$. The following gives the obtained $u_i, i = 1,..., p$ values: $\underline{u}$ = (80,7; 51,243; 50,5565). Figure 2.1 shows the ordered simulated values $x_{ij}, i = 1,..., p; j = 1,..., n$ with the upper quantiles $u_i, i = 1,..., p$ chosen as thresholds.

Fig. 2.1 Upper quantiles of the ordered simulated values







In Section 2.3.2 it was mentioned that $\xi_i = 0, i = 1,..., p$ is a special case of the MGBG distribution, thus the distribution of *U* is approximately MGG($\underline{k}, \underline{\mu}, \sum$) distributed. An easy method for estimating $\underline{\mu}$ and $\sum$ is using the method of moments, shown in the following equations, where only data under the threshold $u_i, i = 1,..., p$ are considered:

$$\hat{\underline{\mu}} = \underline{\bar{y}} = \sum_{j=1}^{n_u} \left(y_{1j}, ..., y_{pj} \right)' / n_u \qquad (2.22)$$

and

$$\hat{\Sigma} = \left\{ \sum_{j=1}^{n_u} \left(y_{ij}, ..., y_{pj} \right)' \left(y_{ij}, .., y_{pj} \right) - \hat{\underline{\mu}} \hat{\underline{\mu}}' \right\} / n_u, i = 1, ..., p, \; j = 1, ..., n_u . \qquad (2.23)$$

The following estimates are obtained when using the method of moments:

$$\hat{\underline{\mu}} = \left(-4,0606; -3,8153; -3,6912 \right) \text{ and } \hat{\Sigma} = \begin{bmatrix} 0,034 & 0,0049 & 0,0096 \\ 0,0049 & 0,0069 & 0,0096 \\ 0,0096 & 0,0043 & 0,0206 \end{bmatrix}.$$

The actual parameter values from the beginning of section 2.4 are as follows:

$$\underline{\mu} = \left(-3,9; -3,8; -3,6 \right) \text{ and } \Sigma = \begin{bmatrix} 0,014 & 0,004 & 0,007 \\ 0,004 & 0,007 & 0,004 \\ 0,007 & 0,004 & 0,013 \end{bmatrix}.$$

It would furthermore be interesting to repeat this simulation in order to consider the distribution of the estimates.

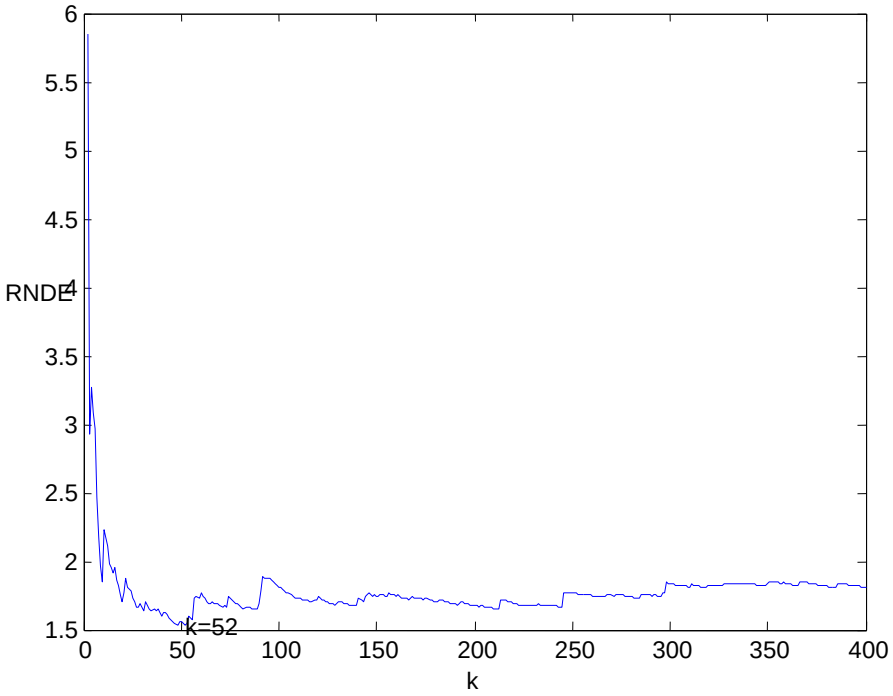
Choosing thresholds by using the upper quantiles is a rough way of selecting thresholds. A more thorough way of selecting thresholds is by making use of the Dirichlet process. In Chapter 4 this method, and other methods on how to select a threshold, are discussed in detail.

Basically, the Dirichlet process suggests that a threshold *t* is selected by considering the entropy of the Dirichlet distribution (De Waal *et al*. 2007, p. 625). Refer to Chapter 4, Section 4.3.4 for a discussion on the entropy of the Dirichlet approach. For the simulated data the Dirichlet process results in the threshold being chosen at *k* = 52, thus the threshold is the 52nd ordered sample value. Therefore, in the case of the simulated example a new threshold *t* = 130,1441 is selected. The threshold is selected based on the

distance measure; refer to Chapter 4, p. 80. Therefore only one threshold is selected.

Figure 2.2 shows that the negative differential entropy of the Dirichlet process reaches a minimum at k = 52.

Fig. 2.2 The negative differential entropy



μ and $\sum$ are now re-estimated by using the method of moments given in the following equations but with a new trimmed data set where the threshold is chosen by the Dirichlet process as t = 130,1441:

$$\hat{\underline{\mu}} = \overline{\underline{y}} = \sum_{j=1}^{n_t} \left(y_{1j}, ..., y_{pj} \right)' / n_t \qquad (2.24)$$

and

$$\hat{\Sigma} = \left\{ \sum_{j=1}^{n_t} \left(y_{ij}, ..., y_{pj} \right)' \left(y_{ij}, .., y_{pj} \right) - \hat{\underline{\mu}} \hat{\underline{\mu}}' \right\} / n_t, i = 1, ..., p, \; j = 1, ..., n_t \qquad (2.25)$$

where n_t denotes the number of observations below the threshold *t*.

The following new estimates are obtained: $\hat{\underline{\mu}} = (-4,2297; -3,8656; -3,8074)$ and

$$\hat{\Sigma} = \begin{bmatrix} 0,0875 & 0,0144 & 0,0324 \\ 0,0144 & 0,0101 & 0,0099 \\ 0,0324 & 0,0099 & 0,0415 \end{bmatrix}.$$

We note especially a large difference, up to 300%, in the estimation of Σ in comparison with p. 36.

After obtaining $\hat{\underline{\mu}}$ and $\hat{\Sigma}$ the other two parameters $\underline{k}$ and $\underline{\xi}$ are estimated, this is discussed in Sections 2.4.1.2 and 2.4.1.3.

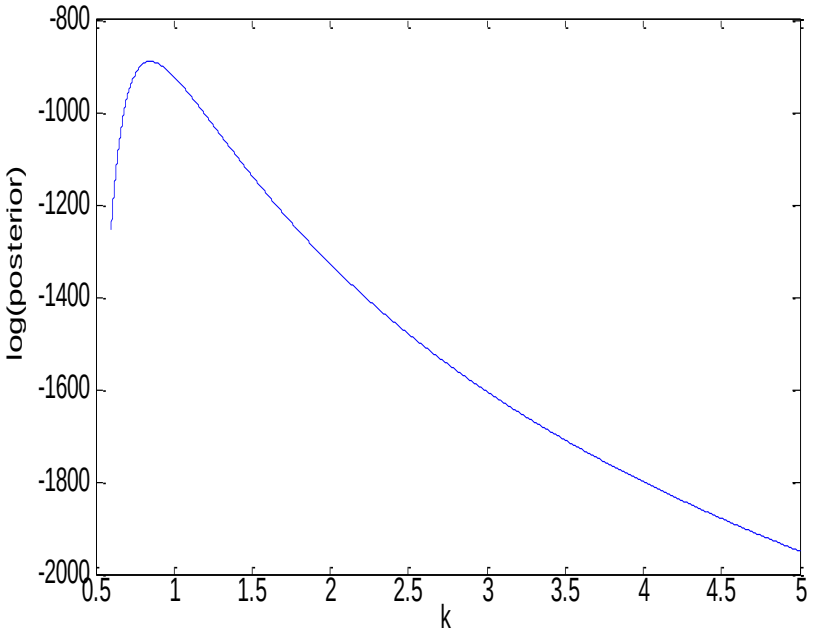
2.4.1.2 Bayesian-Moments method

A method for estimating $\underline{k}$ and $\underline{\xi}$ consecutively is to first estimate $k_i, i = 1,..., p$ by using a Bayesian approach, taking $\xi_i = 0, i = 1,..., p$. Again only the data below the threshold *t* is considered. An estimate for $k_i, i = 1,..., p$ is where the posterior distribution researches a maximum for different values of $k_i, i = 1,..., p$. A uniform prior is assumed. The following equation gives the posterior distribution of $k_i, i = 1,..., p$:

$$\pi\left(k_i \mid V_{ij}, j = 1,..., n_t\right) = n_t \log\left(\frac{\sqrt{\psi'(k_i)}}{\Gamma(k_i)}\right) - \sum_{j=1}^{n_t} V_{ij} + (k-1)\sum_{j=1}^{n_t} \log V_{ij}, i = 1,..., p. \tag{2.26}$$

Figure 2.3 shows that the posterior distribution of $k_i, i = 1,..., p$ reaches a maximum at k_1 = 0,851 for the first dimension. After considering each dimensions the estimates of $\underline{k}$ is as follows: $\hat{\underline{k}} = [0,851; 0,35; 0,693]$.

Fig. 2.3 The joint posterior distribution for $p=1$



After obtaining $\hat{k}_i, i=1,...,p$, $\xi_i, i=1,...,p$ is estimated by using the method of moments approach. All the data are now considered, not only the data below the threshold. To estimate ξ_i we make use of the relationship (2.27) in Theorem 2.1. The estimator of $E(V_i)$ is given by $\bar{V}_i = \sum \frac{V_{ij}}{n_t}$, which can be set equal to the expression $\frac{1}{\hat{\xi}_i}\left\{\left(1-\hat{\xi}_i\right)^{-\hat{k}_i}-1\right\}$ in order to solve for the estimator $\hat{\xi}_i$.

The method considered here is illustrated in Figure 2.4. For a specific $\hat{k}_i$ different values of ξ_i are substituted into (2.34). An estimate for ξ_i is the value of ξ_i for which $E(V_i)=\bar{V}_i$. This may be a crude method of estimation but it is a possible solution to the problem. The expected value of V_i is derived in Theorem 2.1.

Theorem 2.1

Given the assumption $V_{\xi_i} \sim \text{GAM}(k_i,1), i=1,...,p$, the expected value of $V_i, i=1,...,p$ is calculated as follows:

$$E(V_i) = E\left\{\frac{\exp\left(\xi_i V_{\xi_i}\right)-1}{\xi_i}\right\} = \frac{1}{\xi_i}\left\{\left(1-\xi_i\right)^{-k_i} - 1\right\}, i = 1,..., p\,. \tag{2.27}$$

Proof:

From Definition 2.2

$$V_i = \frac{\exp\left(\xi_i V_{\xi_i}\right)-1}{\xi_i}, \tag{2.28}$$

and therefore

$$\begin{aligned} E(V_i) &= \frac{1}{\Gamma(k_i)}\int_0^\infty e^{-V_{\xi_i}} V_{\xi_i}^{k_i-1}\left(\frac{e^{\xi_i V_{\xi_i}} - 1}{\xi_i}\right) dV_{\xi_i}, \\ &= \frac{1}{\xi_i \Gamma(k_i)}\left[\int_0^\infty e^{V_{\xi_i}(\xi_i - 1)} V_{\xi_i}^{k_i-1}\partial V_{\xi_i} - \int_0^\infty e^{-V_{\xi_i}} V_{\xi_i}^{k_i-1}\partial V_{\xi_i}\right]. \end{aligned} \tag{2.29}$$

But because $V_{\xi_i} \sim GAM(k_i 1)$,

$$\int_0^\infty e^{-V_{\xi_i}} V_{\xi_i}^{k_i-1}\partial V_{\xi_i} = \Gamma(k_i). \tag{2.30}$$

Let $s_i = 1-\xi_i$, then

$$E(V_i) = \frac{1}{\xi_i \Gamma(k_i)}\left[\int_0^\infty e^{-V_{\xi_i} s_i} V_{\xi_i}^{k_i-1}\partial V_{\xi_i} - \Gamma(k_i)\right]. \tag{2.31}$$

Consider $\int_0^\infty e^{-V_{\xi_i} s_i} V_{\xi_i}^{k_i-1}\partial V_{\xi_i}$ with substitution. Let $u_i = V_{\xi_i} s_i$, then

$$V_{\xi_i} = \frac{u_i}{s_i} \text{ and } \partial V_{\xi_i} = \frac{1}{s_i}\partial u_i\,. \tag{2.32}$$

Thus

$$\int_0^\infty e^{V_{\xi_i} s_i} V_{\xi_i}^{k_i-1} \partial V_{\xi_i} = \int_0^\infty e^{-u_i} \left(\frac{u_i}{s_i}\right)^{k_i-1} \left(\frac{1}{s_i}\right) \partial u_i$$
$$= \frac{1}{s_i^{k_i}} \int_0^\infty e^{-u_i} u_i^{k_i-1} \partial u_i \qquad (2.33)$$
$$= \frac{1}{s_i^{k_i}} \Gamma(k_i).$$

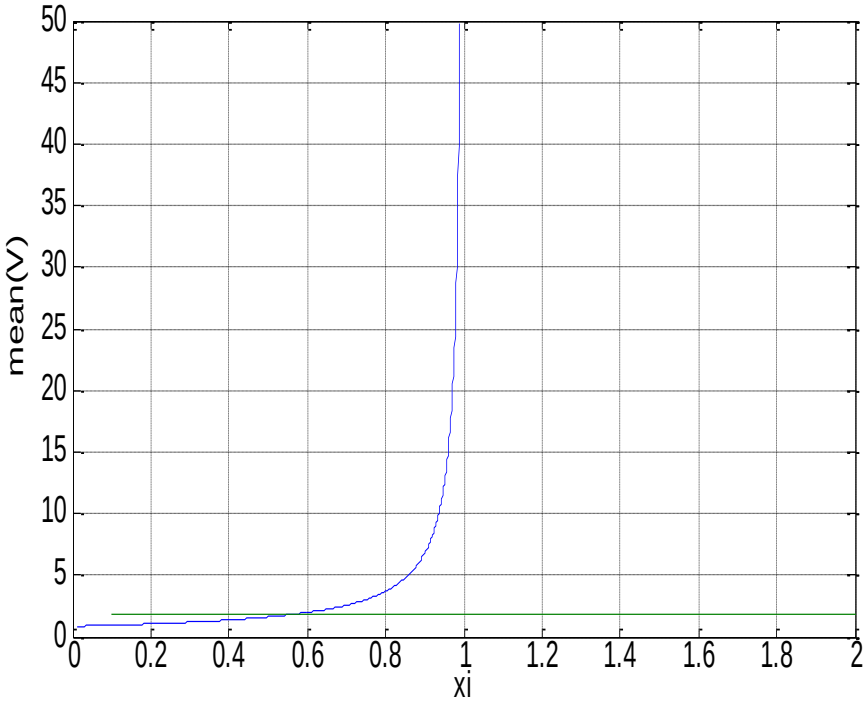
Therefore

$$E(V_i) = \frac{1}{\xi_i \Gamma(k_i)} \left[\frac{\Gamma(k_i)}{s_i^{k_i}} - \Gamma(k_i) \right]$$
$$= \frac{\Gamma(k_i)}{\xi_i \Gamma(k_i)} \left[\frac{1}{(1-\xi_i)^{k_i}} - 1 \right] \qquad (2.34)$$
$$= \frac{1}{\xi_i} \left[(1-\xi_i)^{-k_i} - 1 \right].$$

#

The disadvantage of this method is that it can only be used when $0 < \xi_i < 1, i = 1, ..., p$. Figure 2.4 shows that, for *p* = 1, $E(V_1) = \bar{V}_1$ when ξ_1 = 0,548, thus $\hat{\xi}_1 = 0,548$.

Fig. 2.4 Estimation of ξ_1



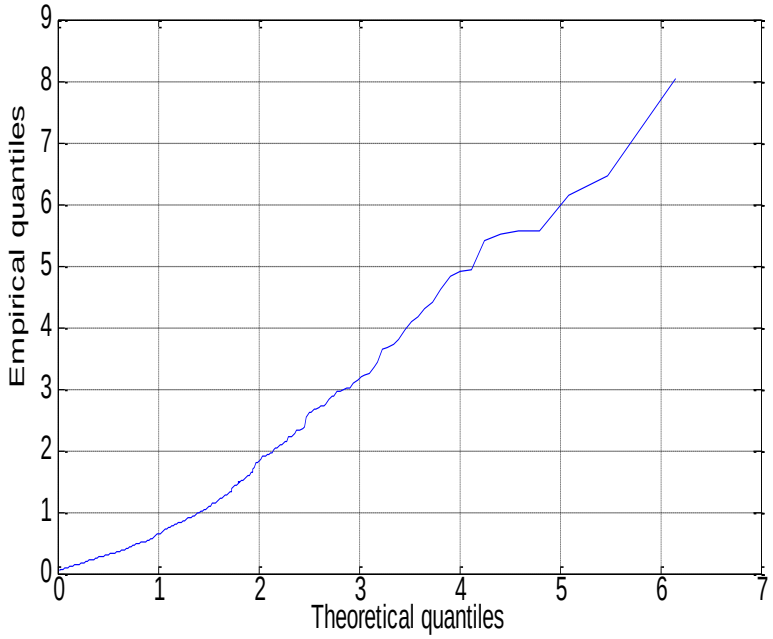
After considering all dimensions, the estimate of $\underline{\xi}$ is $\hat{\underline{\xi}} = [0,548; 0,901; 0,652]$.

2.4.1.2.1 Goodness of fit

As a goodness of fit test the correlation coefficient on a QQ-plot is calculated for the estimated $\underline{k}$ and $\underline{\xi}$. The correlation coefficient, r_Q, can be used to measure the fit of the model to the empirical data by judging the value of r_Q. If r_Q = 1 all the data points lie on a perfectly straight line. The closer r_Q is to 1 the better the fit (Beirlant *et al*. 2004, p. 9). In this simulated case, where k_1 = 0,851 and $\hat{\xi}_1 = 0,548$, r_{Q_1} = 0,9745, this value is close to 1, therefore the model seems to be a good fit to the empirical data. A quantile-quantile plot (QQ-plot) is also a measure of goodness of fit, the theoretical quantiles are plotted against the empirical quantiles, when this plot follows a straight line the underlying model is a good fit to the data (Beirlant *et al*. 2004, p. 4). In Chapter 4, Section 4.4.3, the QQ-plot is discussed in more detail. Figure 2.5 shows the QQ-plot, this plot is fairly close to a straight line, therefore the

MGBG model seems to be a good fit to the data when using the Bayesian-Moments method for estimating the parameters.

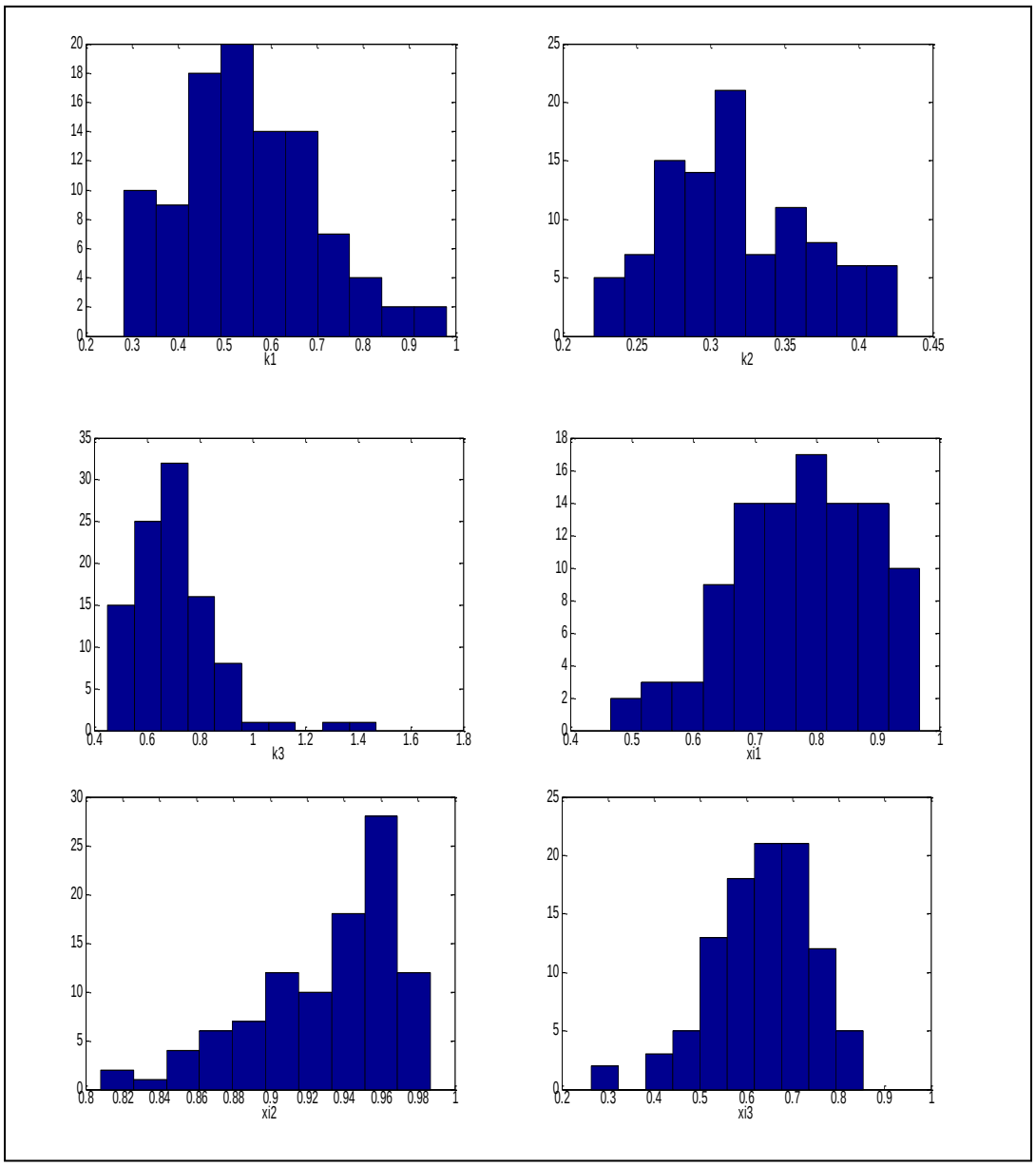
Fig. 2.5 Quantile-Quantile plot for $p = 1$



The Bayesian-Moments method is now repeated a hundred times to ensure that the method is consistent. Each repetition consists of simulating a dataset as discussed at the beginning of Section 2.4. A threshold is then chosen through the entropy of the Dirichlet process, and $\underline{\hat{\mu}}$ and $\hat{\Sigma}$ are obtained by the method of moments. Then $\hat{k}_{ij}, i = 1,..., p, j = 1,...,100$ is taken as the value of *k* where the posterior reaches a maximum. For this $\hat{k}_{ij}, i = 1,..., p, j = 1,...,100$, $\hat{\xi}_{ij}, i = 1,..., p, j = 1,...,100$ is estimated as the value of ξ where $E\left(V_{ij}\right) = \overline{V}_{ij}, i = 1,..., p, j = 1,...,100$. Then $r_{Q_{ij}}, i = 1,..., p, j = 1,...,100$ is calculated. The values $\hat{k}_{ij}$, $\hat{\xi}_{ij}$ and $r_{Q_{ij}}, i = 1,..., p, j = 1,...,100$ are kept for each repetition. After a hundred repetitions the mean of the vector of $\hat{k}_{ij}, i = 1,..., p; j = 1,...,100$ and $\hat{\xi}_{ij}, i = 1,..., p' j = 1,...,100$ gives an accurate estimate for $k_i, i = 1,..., p$ and $\xi_i, i = 1,..., p$. The following estimates are obtained: $\underline{\hat{k}}$ =

$[0,5489; 0,3181; 0,7023]$, $\hat{\underline{\xi}}$ = $[0,7736; 0,9304; 0,6362]$, and r_Q lies between $[[0,9432; 0,9965]; [0,9832; 0,9978]; [0,9209; 0,9935]]$. Figure 2.6 shows the histograms of the simulated $\hat{\underline{k}}$'s and $\hat{\underline{\xi}}$'s.

Fig. 2.6 Simulated histograms



2.4.1.3 A pure Bayesian method

Another method for estimating $\underline{k}$ and $\underline{\xi}$ simultaneously is by using the Bayesian approach with the Metropolis Hastings algorithm and a type of Gibbs sampler. Now there is no restriction on $\underline{\xi}$. For the estimation of $k_i, i = 1,..., p$ all the data is taken into consideration, not only the data below a threshold. A value for $\xi_i, i = 1,..., p$ is chosen, *ex*. ξ_1 = 1,9. For this value of $\xi_i, i = 1,..., p$, $k_i, i = 1,..., p$ can be estimated as the value of $k_i, i = 1,..., p$ where the joint posterior distribution reaches a maximum. Theorem 2.2 and Approximation 2.3 gives the joint likelihood and maximal data information (MDI) prior that are use to construct the joint posterior distribution.

Theorem 2.2

Given the assumption $V_{\xi_i} \sim \text{GAM}(k_i, 1), i = 1,..., p$, the following equation is derived as the joint likelihood function of $k_i, i = 1,..., p$ and $\xi_i, i = 1,..., p$:

$$f\left(y_{ij} \mid k_i, \xi_i\right) = n \log\left(\frac{\sqrt{\psi'(k_i)}}{\Gamma(k_i)}\right) - (1+\xi_i)\sum_{j=1}^{n} V_{\xi_{ij}} + (k_i - 1)\sum_{j=1}^{n} \log V_{\xi_{ij}} + \sum_{j=1}^{n} \log V_{ij},$$
$$i = 1,..., p,\ j = 1,...n. \tag{2.35}$$

Proof:

Because $V_{\xi_i} \sim GAM\left(k_i, 1\right), i = 1,..., p,$

$$f\left(V_{\xi_i}\right) = \frac{1}{\Gamma(k_i)} e^{-V_{\xi_i}} V_{\xi_i}^{k_i - 1}, \tag{2.36}$$

and from Definition 2.2

$$V_{\xi_i} = \frac{1}{\xi_i}\log\left(1+\xi_i V_i\right). \tag{2.37}$$

Next the Jacobean is found and it follows that:

$$\begin{aligned} f\left(V_i\right) &= f\left(w\left(V_i\right)\right)\left|\frac{\partial}{\partial V_i} w\left(V_i\right)\right|, \text{ where } w\left(V_i\right) = V_{\xi_i} \\ &= \frac{1}{\Gamma\left(k_i\right)}\left(1+\xi_i V_i\right)^{-\frac{1}{\xi_i}-1}\left(\frac{1}{\xi_i}\log\left(1+\xi_i V_i\right)\right). \end{aligned} \tag{2.38}$$

From Definition 2.2

$$V_i = \exp\left(\psi\left(k_i\right) - \frac{\underline{Y}-\underline{\mu}}{\sigma}\sqrt{\psi'\left(k_i\right)}\right). \tag{2.39}$$

Now for $w\left(V_i\right) = V_{\xi_i}$

$$\begin{aligned} f\left(y_i\right) &= f\left(w\left(y_i\right)\right)\left|\frac{\partial}{\partial y_i} w\left(y_i\right)\right| \\ &= \frac{\sqrt{\psi'\left(k_i\right)}}{\sigma\Gamma\left(k_i\right)}\left(1+\xi_i V_i\right)^{\frac{-1}{\xi_i}-1}\left[\frac{1}{\xi_i}\log\left(1+\xi_i V_i\right)\right]^{k_i-1} V_i. \end{aligned} \tag{2.40}$$

Therefore

$$\log f\left(y_i\right) \propto \log\left(\frac{\sqrt{\psi'\left(k_i\right)}}{\Gamma\left(k_i\right)}\right) - \left(1+\xi_i\right)V_{\xi_i} + \left(k_i-1\right)\log V_{\xi_i} + \log V_i\,. \tag{2.41}$$

The joint likelihood function of k_i and ξ_i is therefore given by

$$f\left(y_{ij} \mid k_i, \xi_i\right) = n\log\left(\frac{\sqrt{\psi'\left(k_i\right)}}{\Gamma\left(k_i\right)}\right) - \left(1+\xi_i\right)\sum_{j=1}^{n} V_{\xi_{ij}} + \left(k_i - 1\right)\sum_{j=1}^{n}\log V_{\xi_{ij}} + \sum_{j=1}^{n}\log V_{ij},$$

$$i = 1,..., p,\ j = 1,...n. \tag{2.42}$$

#

Approximation 2.3

Given the assumption $V_{\xi_i} \sim \text{GAM}\left(k_i, 1\right), i = 1,..., p$ and that $\underline{Y} = -\log \underline{X}$ the MDI prior is derived as shown in the following equation:

$$\pi\left(k_i, \xi_i\right) \propto \exp\left\{\log\frac{\sqrt{\psi'\left(k_i\right)}}{\Gamma\left(k_i\right)} - \log\sigma - \left(1+\xi\ \right)k_i + \left(k_i - 1\right)\psi\left(k_i\right) + \log\left[\frac{1}{\xi_i}\left(e^{\xi_i k_i} - 1\right)\right] - \frac{\xi_i^{\,2} k_i e^{\xi_i k_i}}{2\left(e^{\xi_i k_i} - 1\right)^2}\right\}$$

$$i = 1,..., p. \tag{2.43}$$

Proof:

The MDI prior is defined as $\pi\left(k_i, \xi_i\right) \propto \exp E\left\{\log f\left(\underline{Y} \mid k_i, \xi_i\right)\right\}$ (Beirlant *et al.* 2004, p. 432) and the density of $\underline{Y} = -\log \underline{X}$ is defined as

$$f\left(y_i\right) = \frac{\sqrt{\psi'\left(k_i\right)}}{\sigma\Gamma\left(k_i\right)}\left(1+\xi_i v_i\right)^{-\frac{1}{\xi_i}-1} v_{\xi_i}^{k_i - 1} v_i\ . \tag{2.44}$$

We know that $V_{\xi_i} \sim GAM\left(k_i, 1\right)$ and $V_{\xi_i} = \frac{1}{\xi_i}\log\left(1+\xi_i V_i\right)$, where

$$V_i = \exp\left\{\psi\left(k_i\right) - \frac{\left(\underline{Y} - \underline{\mu}\right)}{\sigma}\sqrt{\psi'\left(k_i\right)}\right\}. \tag{2.45}$$

Therefore

$$\log f\left(y_i \mid k_i, \xi_i\right) = \log \frac{\sqrt{\psi'\left(k_i\right)}}{\Gamma\left(k_i\right)} - \log \sigma - \left(1+\xi_i\right) V_{\xi_i} + \left(k_i - 1\right) \log V_{\xi_i} + \log V_i \tag{2.46}$$

and

$$E \log f\left(y_i \mid k_i, \xi_i\right) = \log \frac{\sqrt{\psi'\left(k_i\right)}}{\Gamma\left(k_i\right)} - \log \sigma - \left(1+\xi_i\right) k_i + \left(k_i - 1\right) \psi\left(k_i\right) + E \log V_i \,. \tag{2.47}$$

The $E \log\left(V_i\right)$ was approximated by the delta method (see Approximation 2.1) as $E\left(\log\left(V_i\right)\right) \approx \log\left(\frac{e^{\xi_i k} - 1}{\xi_i}\right) - \frac{k_i \xi_i^2 e^{\xi_i k_i}}{2\left(e^{\xi_i k_i} - 1\right)^2}$.

Therefore, the MDI prior is

$$\pi\left(k_i, \xi_i\right) \propto \exp\left\{\log \frac{\sqrt{\psi'\left(k_i\right)}}{\Gamma\left(k_i\right)} - \log \sigma - \left(1+\xi\right) k_i + \left(k_i - 1\right) \psi\left(k_i\right) + \log\left[\frac{1}{\xi_i}\left(e^{\xi_i k_i} - 1\right)\right] - \frac{\xi_i^2 k_i e^{\xi_i k_i}}{2\left(e^{\xi_i k_i} - 1\right)^2}\right\}$$

$i = 1, \ldots, p.$

(2.48)

#

The joint posterior distribution is now defined as follows:

$$\pi\left(k_i, \xi_i \mid y_{ij}\right) \propto f\left(y_{ij} \mid k_i, \xi_i\right) \pi\left(k_i, \xi_i\right), i = 1, \ldots, p; j = 1, \ldots, n \,. \tag{2.49}$$

For the specific $\hat{k}_i, i = 1, \ldots, p$, where the posterior distribution reaches a maximum, $V_{ij} = \frac{1}{\xi_i}\left\{\exp\left(\xi_i V_{\xi_i}\right)\right\}, i = 1, \ldots, p; j = 1, \ldots, n$ is obtained. A new threshold t is now chosen on $V_{ij}, i = 1, \ldots, p; j = 1, \ldots, n$. Again the threshold is chosen by making use of the entropy of the Dirichlet process. $\xi_i, i = 1, \ldots, p$ is estimated by fitting a Generalized Pareto distribution (GPD) on the

$V_i, i = 1,..., p$ values above the chosen threshold *t*. The GPD is given by the equation

$$P\left(V_i - t > v_{ij} \mid V_i > t\right) = \left(1 + \frac{\xi_i V_i}{\sigma_t}\right)^{\frac{-1}{\xi_i}},\ v_{ij} > 0; \sigma_t > 0, i = 1,..., p;\ j = 1,..., N_t \tag{2.50}$$

where N_t is the number of observations above the threshold *t*.

Theorem 2.3 shows that the distribution of the
$V_{ij} > t, i = 1,..., p;\ j = N_t - (N_t - k),..., N_t$ values, above the threshold *t,* is GPD.

Theorem 2.3

For large values of *t*, $V_{ij} > t, i = 1,..., p;\ j = N_t - (N_t - k),..., N_t$ follows a GPD.

Proof:

$$P\left(V_i - t > v_{ij} \mid V_i > t\right) = P\left(\frac{e^{\xi_i V_{\xi_i}} - 1}{\xi_i} > t + v_{ij} \mid \frac{e^{\xi_i V_{\xi_i}} - 1}{\xi_i} > t\right), \text{ where } V_i = \frac{e^{\xi_i V_{\xi_i}} - 1}{\xi_i}$$

$$= P\left(V_{\xi_i} > \frac{\log\left(\xi_i\left(t + v_{ij}\right) + 1\right)}{\xi_i}\right) \div P\left(V_{\xi_i} > \frac{\log\left(\xi_i t + 1\right)}{\xi_i}\right).$$

(2.51)

Let $\frac{\log\left(\xi_i\left(t + v_{ij}\right) + 1\right)}{\xi_i} = a$ and $V_{\xi_i} > \frac{\log\left(\xi_i t + 1\right)}{\xi_i} = b,$ then

$$P\left(V_i - t > v_{ij} \mid V_i > t\right) = \frac{\Gamma\left(k, a\right)}{\Gamma\left(k, b\right)}. \tag{2.52}$$

The incomplete Gamma function is given by the integral

$$\Gamma(k,a)=\int_{a}^{\infty}h^{k-1}e^{-h}dh\,. \tag{2.53}$$

The following equation, shown by Amore (2005), gives an approximate formula for the incomplete gamma function in the region of large values of *a*:

$$\Gamma(k,a)\approx e^{-a}(1+a)^{k-1}\,. \tag{2.54}$$

This approximation can be used for arbitrary values of *a* and *k* (Amore 2005).

Thus

$$\begin{aligned}\frac{\Gamma(k,a)}{\Gamma(k,b)}&=\frac{(1+\xi_i t+\xi_i v_{ij})^{\frac{-1}{\xi_i}}}{(1+\xi_i t)^{\frac{-1}{\xi_i}}}\frac{\left\{1+\frac{1}{\xi_i}\left[\ln(1+\xi_i t+\xi_i v_{ij})\right]\right\}^{k_i-1}}{\left\{1+\frac{1}{\xi_i}\left[\ln(1+\xi_i t)\right]\right\}^{k_i-1}}\\&=\left(1+\frac{\xi_i v_{ij}}{1+\xi_i t}\right)^{-\frac{1}{\xi_i}}\left(\frac{1+\frac{1}{\xi_i}\left[\ln(1+\xi_i t+\xi_i v_{ij})\right]}{1+\frac{1}{\xi_i}\left[\ln(1+\xi_i t)\right]}\right)^{k_i-1}.\end{aligned} \tag{2.55}$$

Then, by making use of L'Hospital's rule (Salas *et al*. 1999), $\left(\frac{1+\frac{1}{\xi_i}\left[\ln(1+\xi_i t+\xi_i v_{ij})\right]}{1+\frac{1}{\xi_i}\left[\ln(1+\xi_i t)\right]}\right)^{k_i-1}\to 1$, as $t\to\infty$. It can be shown that for large *t,*

$\frac{\Gamma(k,a)}{\Gamma(k,b)}=\left(1+\frac{\xi_i v_{ij}}{\sigma_t}\right)^{-\frac{1}{\xi_i}}$. This is the distribution function of the GPD$(\xi_i,\sigma_t), i=1,...,p$ on the exceedances above *t*, where $\sigma_t=1+\xi_i t$.

#

The following equations show the MDI prior and the posterior distribution of the GPD respectively:

$$\pi\left(\sigma_t, \xi_i\right) \propto \frac{1}{\sigma_t} e^{-\xi_i}, i = 1, ..., p \tag{2.56}$$

(Beirlant *et al*. 2004, p. 447) and

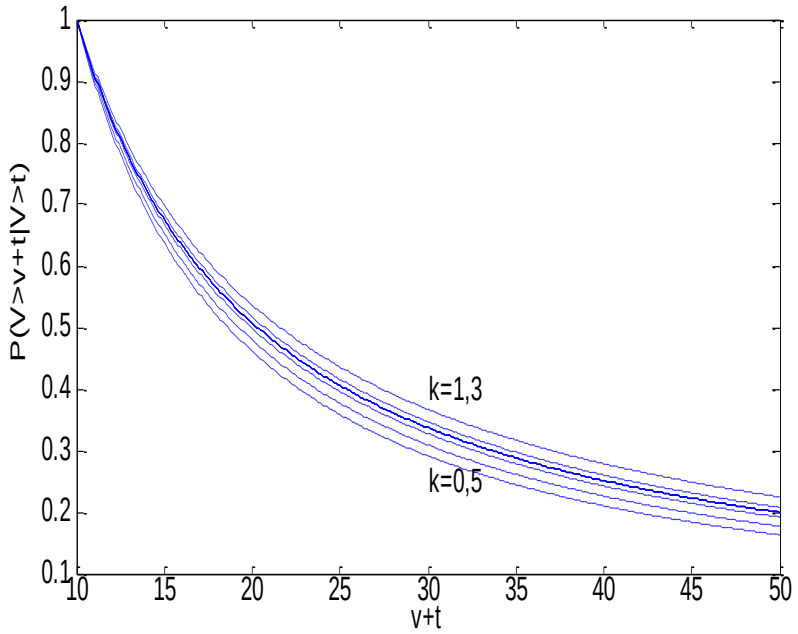
$$\pi\left(k_i, \xi_i \mid v_{ij}\right) = -\left(\frac{1}{\xi_i} + 1\right) \sum_{j=1}^{N_t} \log\left(1 + \frac{\xi_i\left(v_{ij} - t\right)}{\sigma_t}\right) - \left(N_t + 1\right) \log\left(\sigma_t\right) - \xi_i, i = 1, ..., p. \tag{2.57}$$

Next the GPD is compared with the incomplete gamma distribution. As mentioned in Definition 2.2 V_ξ follows a gamma distribution and Theorem 2.3 shows that the gamma distribution is approximated by the GPD, given by the equation

$$F\left(v + t \mid t\right) = \left\{1 + \frac{\xi\left(v + t\right)}{1 + \xi t}\right\}^{\frac{-1}{\xi}}. \tag{2.58}$$

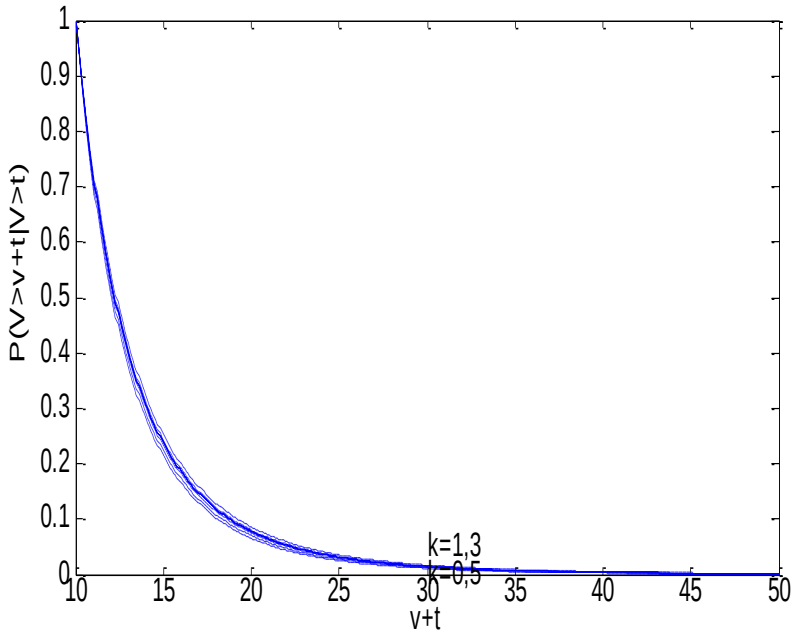
Figures 2.7, 2.8 and 2.9 show how the approximated GPD fits the data above the threshold when compared to the actual incomplete gamma distribution. In Figure 2.7 a threshold is chosen at *t* = 10, ξ is chosen as 0,95 and *k* takes on values between 0,5 and 1,3.

Fig. 2.7 **Comparison between the approximated GPD and the incomplete gamma distribution**



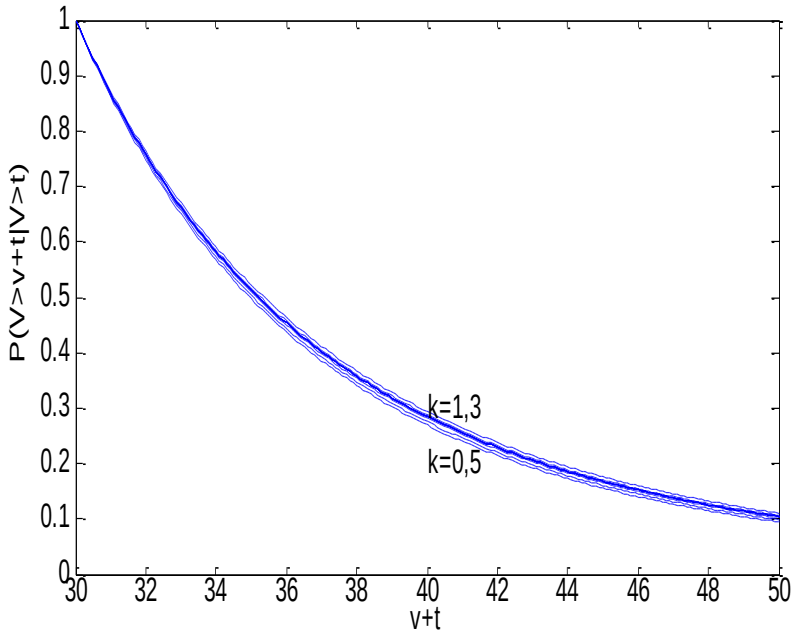
In Figure 2.8 a threshold is again chosen at t = 10, now ξ is chosen as $\xi = 0,2$ and k is chosen as different values between 0,5 and 1,3.

Fig. 2.8 **Comparison between the approximated GPD and the incomplete gamma distribution with a smaller value of ξ**



In Figure 2.9 a threshold is chosen at a larger value *t* = 30, ξ is chosen as $\xi = 0,2$ and *k* is chosen as different values between 0,5 and 1,3.

Fig. 2.9 Comparison between the approximated GPD and the incomplete gamma integral with a larger value of *t*



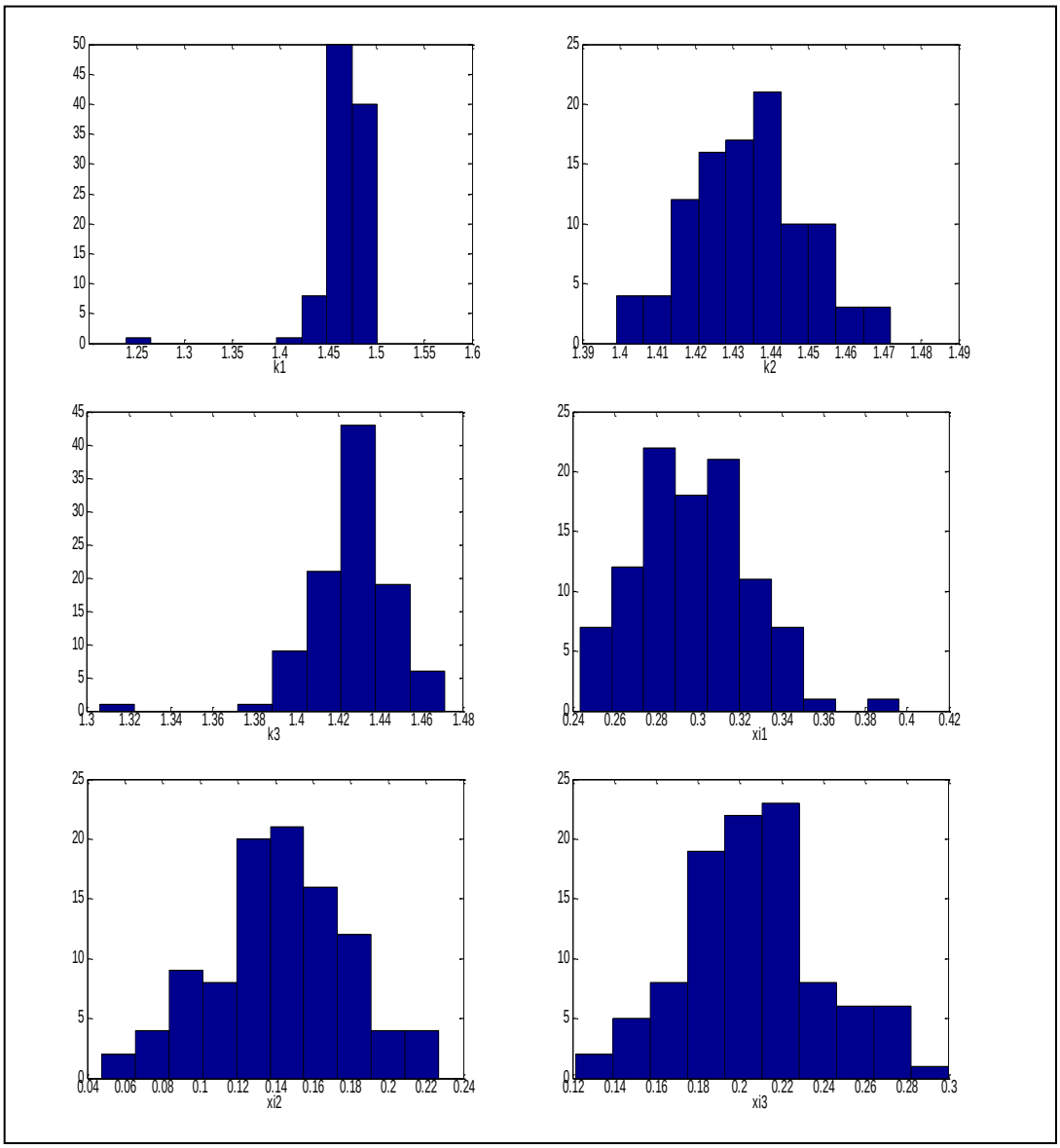
Figures 2.7, 2.8 and 2.9 show that, for small values of ξ**,** the approximated GPD follows the incomplete gamma distribution more closely.

When a GPD is fitted to the data above the threshold an estimate of $\xi_i, i = 1,..., p$ is obtained by making use of the Metropolis Hastings algorithm with 5000 iterations. When ξ_i converges to a certain value, an estimate of $\xi_i, i = 1,..., p$ is the mean of the values of $\xi_i, i = 1,..., p$ from the 5000 iterations. This $\hat{\xi}_i, i = 1,..., p$ can now be used to estimate a new $k_i, i = 1,..., p$, by taking the value where the joint posterior distribution of $k_i, i = 1,..., p$ and $\xi_i, i = 1,..., p$ reaches a maximum. The new $\hat{k}_i, i = 1,..., p$ is used to obtain a new $\hat{\xi}_i, i = 1,..., p$. This process is repeated a hundred times. To estimate the final

$k_i, i = 1,..., p$ and $\xi_i, i = 1,..., p$ the mean of $\hat{\underline{k}}$ and $\hat{\underline{\xi}}$ after the hundred repetitions are taken. The following estimates are obtained: $\hat{\underline{k}} = [1,4673; 1,434; 1,4268]$ and $\hat{\underline{\xi}} = [0,2974; 0,1433; 0,2062]$.

Figure 2.10 shows the histograms of the simulated $\hat{k}_i$'s and $\hat{\xi}_i$'s, $i = 1,..., p$:

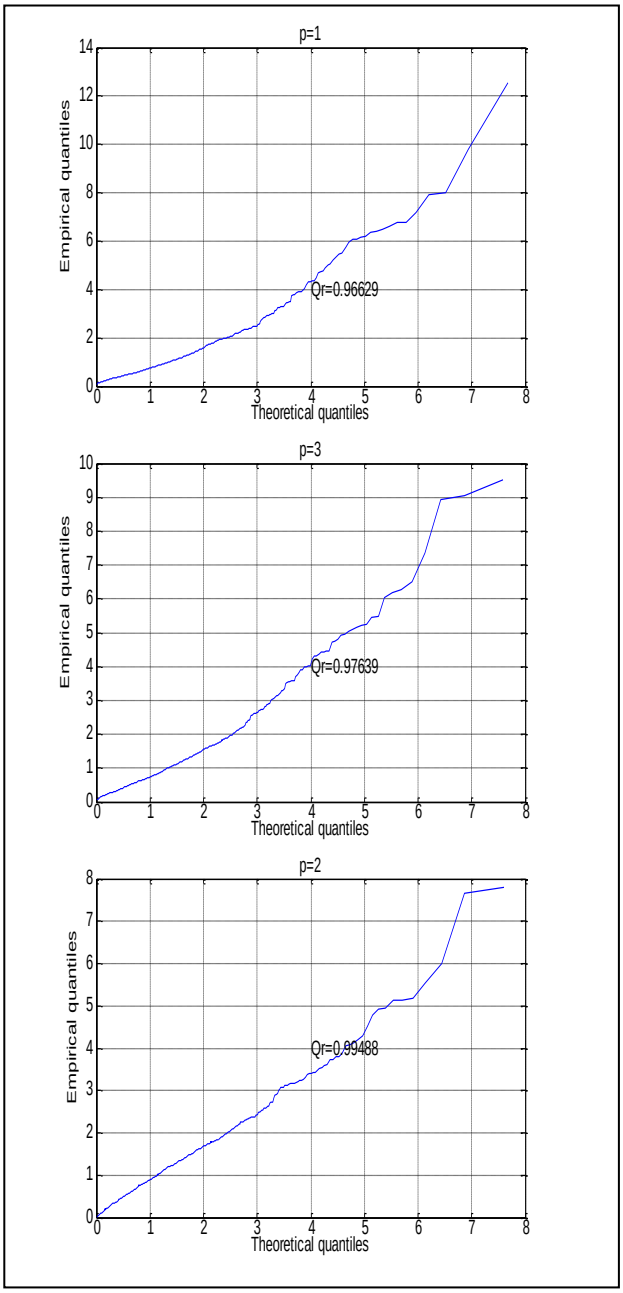
Fig. 2.10 Histograms of the simulated k_i's and ξ_i's, $i = 1,..., p$



2.4.1.3.1 Goodness of fit

As a goodness of fit test the quantile-quantile plots are constructed for each dimension. These plots are shown in Figure 2.11. The plots in all three cases are very close to a straight line indicating a reasonably good fit. The r_Q value indicated on each graph lies close to 1, indicating a strong correlation and a good fit.

Fig. 2.11 Quantile-Quantile plots for $i = 1, \ldots, p$



Instead of assuming the MDI prior one can assume a uniform prior. The estimated parameter values when using an MDI and uniform prior respectively are very close to each other, as shown in Table 2.1. The QQ-plots also indicate a reasonable good fit when a uniform prior is used.

Table 2.1 Comparing the estimates of k_i and $\xi_i, i = 1,..., p$ using an MDI and uniform prior respectively

Using an MDI prior			
	$i=1$	$i=2$	$i=3$
$\hat{k}_i$	1,4673	1,4340	1,4268
$\hat{\xi}_i$	0,2974	0,1433	0,2062
Using a uniform prior			
	$i=1$	$i=2$	$i=3$
$\hat{k}_i$	1,3806	1,0732	1,0953
$\hat{\xi}_i$	0,2868	0,1660	0,1535

2.4.1.4 Comparing different estimation approaches

The same procedures are followed when different data sets are simulated. Two other data sets are simulated and Table 2.2 indicates the different estimated parameter values. The method of moments is used on the data below the threshold to estimate $\underline{\mu}$ and $\sum$ then $\underline{k}$ and $\underline{\xi}$ are estimated by using the Bayesian-Moments method and the pure Bayesian method (with an MDI prior) respectively.

For the second data set observations $\underline{x} = (x_1,..., x_p)$ are simulated from a MGBG($\underline{k}, \underline{\mu}, \sum, \underline{\xi}$) with a sample size *n* = 700, *p*=3, $\underline{k} = (0,8; 0,5; 0,7)$,

$\underline{\mu} = (-3,9; -3,8; -3,6)$, $\underline{\xi} = (0,5; 0,8; 0,3)$ and

$$\Sigma = \begin{bmatrix} 0,0225 & 0,0045 & 0,0119 \\ 0,0045 & 0,0173 & 0,0041 \\ 0,0119 & 0,0041 & 0,0274 \end{bmatrix}.$$

For the third data set observations $\underline{x} = (x_1, ..., x_p)$ are simulated from a MGBG($\underline{k}, \underline{\mu}, \Sigma, \underline{\xi}$) with a sample size *n* = 700, *p*=3, $\underline{k} = (5,5; 6,2; 7,3)$, $\underline{\mu} = (-2,1; -0,5; -1,9)$, $\underline{\xi} = (0,2; 0,3; 0,4)$ and

$$\Sigma = \begin{bmatrix} 0,0225 & 0,0045 & 0,0119 \\ 0,0045 & 0,0173 & 0,0041 \\ 0,0119 & 0,0041 & 0,0274 \end{bmatrix}.$$

Table 2.2 Comparing estimated parameter values for different data sets and different estimation methods

Data set 1					
μ_1 = -4,2297		μ_2 = -3,8656		μ_3 = -3,8074	
$\hat{\Sigma} = \begin{bmatrix} 0,0875 & 0,0144 & 0,0324 \\ 0,0144 & 0,0101 & 0,0099 \\ 0,0324 & 0,0099 & 0,0415 \end{bmatrix}$					
Bayesian-Moments method			**Pure Bayesian method**		
$\hat{k}_1$ =0,5489	$\hat{k}_2$ =0,3181	$\hat{k}_3$ =0,7023	$\hat{k}_1$ =1,4673	$\hat{k}_2$ =1,4340	$\hat{k}_3$ =1,4268
$\hat{\xi}_1$ =0,7736	$\hat{\xi}_2$ =0,9304	$\hat{\xi}_3$ =0,6362	$\hat{\xi}_1$ =0,2974	$\hat{\xi}_2$ =0,1433	$\hat{\xi}_3$ =0,2062
Data set 2					
μ_1 = -3,9280		μ_2 = -3,8051		μ_3 = -3,6135	
$\hat{\Sigma} = \begin{bmatrix} 0,0265 & 0,0048 & 0,0135 \\ 0,0048 & 0,0226 & 0,0048 \\ 0,0135 & 0,0048 & 0,0305 \end{bmatrix}$					
Bayesian-Moments method			**Pure Bayesian method**		
$\hat{k}_1$ = 0,2967	$\hat{k}_2$ = 0,3037	$\hat{k}_3$ = 0,2239	$\hat{k}_1$ = 1,4296	$\hat{k}_2$ = 1,4055	$\hat{k}_3$ = 1,4203
$\hat{\xi}_1$ = 0,9359	$\hat{\xi}_2$ = 0,9179	$\hat{\xi}_3$ = 0,9755	$\hat{\xi}_1$ = 0,1564	$\hat{\xi}_2$ = 0,1523	$\hat{\xi}_3$ = 0,1236

Data set 3					
$\mu_1 =$ -2,4924		$\mu_2 =$ -0,9213		$\mu_3 =$ -2,7480	
$\hat{\Sigma} = \begin{bmatrix} 0,0664 & 0,0226 & 0,0631 \\ 0,0226 & 0,0757 & 0,0243 \\ 0,0631 & 0,0243 & 0,2394 \end{bmatrix}$					
Bayesian-Moments method			**Pure Bayesian method**		
$\hat{k}_1 =$ 0,6010	$\hat{k}_2 =$ 0,7808	$\hat{k}_3 =$ 0,4889	$\hat{k}_1 =$ 1,4397	$\hat{k}_2 =$ 1,4632	$\hat{k}_3 =$ 1,4430
$\hat{\xi}_1 =$ 0,7098	$\hat{\xi}_2 =$ 0,5881	$\hat{\xi}_3 =$ 0,8159	$\hat{\xi}_1 =$ 0,2452	$\hat{\xi}_2 =$ 0,2787	$\hat{\xi}_3 =$ 0,2595

The estimation of k_i is problematic, therefore in the Bayesian-Moment method k_i is first estimated from its posterior for $\xi_i = 0$ and then for a given $\hat{k}_i$ an estimate of ξ_i is obtained through a moments method. Both parameters have an effect on the tail of the distribution and therefore our judgment suggests that the parameters should be estimated simultaneously. Thus the Pure Bayesian method is suggested to be a more appropriate method for estimating k_i and ξ_i jointly.

2.5 Fitting the MGBG to the Gariep inflow data

The estimation of parameters is now done for a real data set. The data set is the maximum monthly inflow of water into the Gariep Dam. This data set consists of the maximum inflow of water for each month over a period of 29 years, from 1976 up to 2006, excluding 1980 and 1982, due to large amounts of missing data. The dimension is *p* = 12 (12 months). A threshold *t* is chosen by using the Dirichlet process, the threshold is chosen at *t* = 5374,9 m^3/s. The parameters $\underline{\mu}$ and $\sum$ are estimated from the data below the threshold using the method of moments, $\underline{\mu}$ and $\sum$ are estimated as

$$\hat{\underline{\mu}} = [-6,5378; -6,6496; -6,5717; -5,9564; -5,1574; -4,4494; -3,8307; -4,1064; -5,0404; -5,7812; -6,2471; -6,3737]$$

and $\hat{\sum}$ =

1.4787	1.0199	0.9875	0.4355	0.6837	0.8893	0.2403	0.5067	0.4804	-0.3263	-0.1273	-0.1017
1.0199	1.8190	1.3972	0.6580	0.5275	1.1619	0.4746	0.0759	0.3457	-0.0996	0.1599	0.0420
0.9875	1.3972	1.6046	0.7662	0.7402	0.7701	0.4404	-0.1492	0.4780	0.2128	0.2256	0.2263
0.4355	0.6580	0.7662	1.2378	1.1826	0.7480	0.2891	0.0953	0.6125	-0.0939	-0.1007	0.2343
0.6837	0.5275	0.7402	1.1826	1.8114	1.1471	0.7054	0.7334	0.9955	-0.5115	-0.2965	0.0559
0.8893	1.1619	0.7701	0.7480	1.1471	2.3765	1.3109	0.9366	0.6296	-0.6636	-0.1367	0.0969
0.2403	0.4746	0.4404	0.2891	0.7054	1.3109	1.5761	0.8621	0.5305	-0.2873	-0.2767	0.2882
0.5067	0.0759	-0.1492	0.0953	0.7334	0.9366	0.8621	2.7845	1.6242	-0.1564	-0.2836	0.1032
0.4804	0.3457	0.4780	0.6125	0.9955	0.6296	0.5305	1.6242	3.5719	1.3096	0.5810	0.5213
-0.3263	-0.0996	0.2128	-0.0939	-0.5115	-0.6636	-0.2873	-0.1564	1.3096	2.6819	1.5615	0.9351
-0.1273	0.1599	0.2256	-0.1007	-0.2965	-0.1367	-0.2767	-0.2836	0.5810	1.5615	1.6886	0.6328
-0.1017	0.0420	0.2263	0.2343	0.0559	0.0969	0.2882	0.1032	0.5213	0.9351	0.6328	1.4292

The parameters $\underline{\xi}$ and $\underline{k}$ are estimated by using the Pure Bayesian method because it is not known whether $0 < \underline{\xi} < 1$. $\underline{k}$ is estimated as

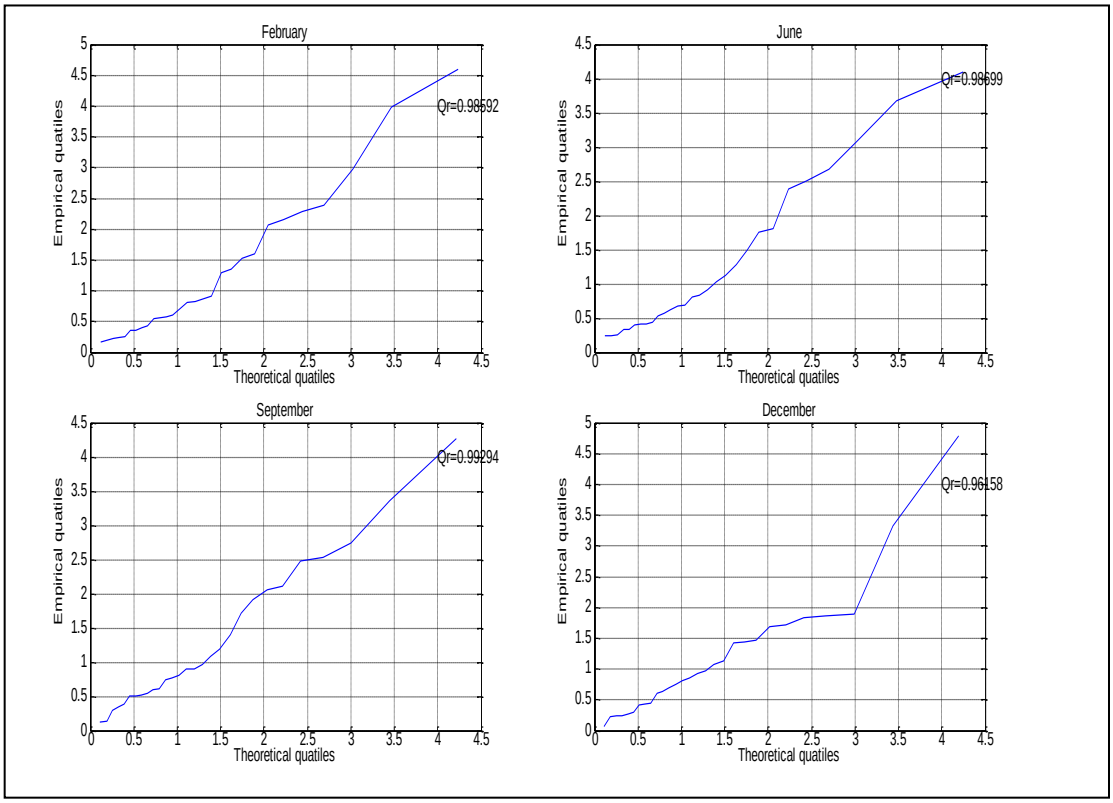
$$\hat{\underline{k}} = [1,3925; 1,4329; 1,4870; 1,4302; 1,4639; 1,4385; 1,4749; 1,4756; 1,4212; 1,4642; 1,5002; 1,4131]$$

and

$$\hat{\underline{\xi}} = [0,1427; 0,1544; 0,1551; 0,1505; 0,1495; 0,1491; 0,1562; 0,1542; 0,1439; 0,1662; 0,1639; 0,1446].$$

As a goodness of fit test the QQ-plots are constructed for each dimension, the plots in all twelve cases are very close to a straight line, indicating a reasonable good fit. The r_Q values lie close to 1, indicating a strong correlation and a good fit. Figure 2.12 shows the QQ-plots for four months; February, June, September and December.

Fig. 2.12 Quantile-quantile plot for four months



Judging by the quantile-quantile plots the MGBG distribution seems to be a good fit to the data. Therefore, it is reasonable to assume that the MGBG can be used when modelling the maximum monthly inflows into the Gariep Dam.

2.5.1 Prediction of future values

After successfully estimating the parameters the aim is to predict future maximum monthly inflow values. The approach is to calculate $E(\underline{X})$, the expected maximum monthly inflow values. In Approximation 2.1 the following approximation was derived for the expected value of $X_i, i = 1,...p$:

$$E(X_i) \approx \exp\left(-\mu_{Y_i}\right) + \frac{1}{2}\sigma_{Y_i}^2 \exp\left(-\mu_{Y_i}\right). \qquad (2.59)$$

When making a prediction it is significant to know the variance and the standard deviation. In Approximation 2.2 the following approximation was obtained for the variance of $X_i, i = 1,..., p$:

$$Var\left(X_i\right) = \sum_{j=1}^{p} k_j \left[\exp\left(\sum_i^{\frac{1}{2}} D_\psi^{-1} \begin{pmatrix} \log\left(\frac{e^{\xi_1 V_{\xi_1}} - 1}{\xi_1} \right) - \psi\left(k_1\right) \\ . \\ . \\ . \\ \log\left(\frac{e^{\xi_p V_{\xi_p}} - 1}{\xi_p} \right) - \psi\left(k_p\right) \end{pmatrix} + \mu_i \right) \left(\left[\sum_i^{\frac{1}{2}} D_\psi^{-1} \left(\frac{{\xi_j}^2}{e^{\xi_j k_j} - 1} \right) e^{\xi_j k_j} \right] \right) \right]^2 .$$

(2.60)

Table 2.3 shows the mean maximum monthly inflow predictions for each month in the first column. The second column gives the standard deviation of the prediction and the third column gives the actual mean monthly maximum inflow from the data set over the time period 1976 to 2006, excluding 1980 and 1982. The maximum monthly inflows can get very extreme, for example, during February 1988 an extreme inflow of 11263m^3/s occurred, therefore the standard deviations will also be large. The standard deviation can be used to construct an interval to give an indication of how large the predicted maximum monthly inflow can become. The interval is $\left[0; \text{Prediction} + \left(2 \times \text{standard deviation}\right)\right]$. The lower bound is zero because inflow values can not be negative.

Table 2.3 Future maximum inflow predictions

	Prediction	**Standard deviation**	**Actual mean monthly maximum**
January	810,1379	2408,7	1081,3
February	910,8510	3882,1	1560,2
March	777,9235	3597	1349,3
April	423,5936	1237,3	621,6
May	205,4390	857,9	350,8
June	107,0454	617	233,0
July	54,8819	170,8	91,3
August	91,9005	335,1	245,2
September	221,0192	1834,8	528,2
October	505,4896	1461,8	838,8
November	676,2649	1584,1	1045,4
December	696,5503	1803,8	851,7

To obtain the predictive distribution would be more appropriate in a Bayesian framework, but due to the complicated nature predictions are made by estimating the mean (2.59).

2.6 Conclusion

Judging by the QQ-plots in Section 2.5 it seems that the MGBG distribution is a good fit to the data. If one considers the presence of extreme values in the actual data set the predicted monthly maximum inflows (from Table 2.3) are not too far from the actual maximum monthly inflows.

Chapter 3

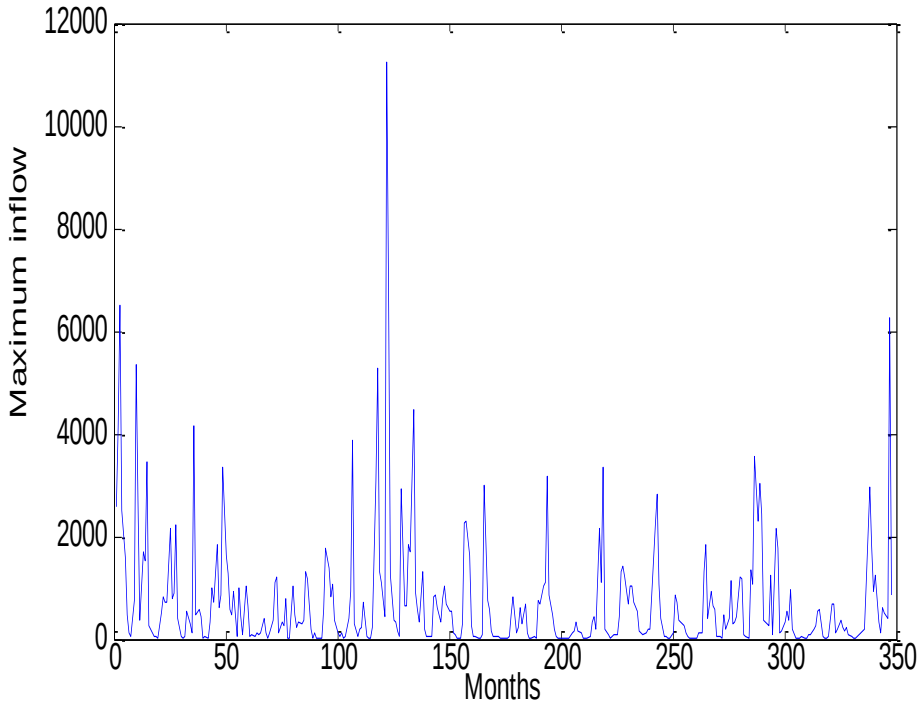
Multivariate regression

This chapter discusses the modelling of multivariate extreme values through multivariate regression.

3.1 Introduction

As mentioned before the maximum monthly inflow of water into the Gariep Dam is recorded from January 1976 until December 2006, excluding 1980 and 1982, due to a large number of missing data values. The monthly maximum inflow values are graphed in Figure 3.1.

Fig. 3.1 Monthly maximum inflow values into the Gariep Dam



In Chapter 2, Section 2.5, the MGBG distribution was used to model the Gariep data set. An upper threshold t = 5374,9 m^3/s was chosen and the following estimates were obtained for μ and $\sum$:

$$\hat{\underline{\mu}} = [-6,5378; -6,6496; -6,5717; -5,9564; -5,1574; -4,4494; -3,8307; -4,1064; -5,0404; -5,7812; -6,2471; -6,3737]$$

and

$$\hat{\Sigma} = \begin{bmatrix} 1.4787 & 1.0199 & 0.9875 & 0.4355 & 0.6837 & 0.8893 & 0.2403 & 0.5067 & 0.4804 & -0.3263 & -0.1273 & -0.1017 \\ 1.0199 & 1.8190 & 1.3972 & 0.6580 & 0.5275 & 1.1619 & 0.4746 & 0.0759 & 0.3457 & -0.0996 & 0.1599 & 0.0420 \\ 0.9875 & 1.3972 & 1.6046 & 0.7662 & 0.7402 & 0.7701 & 0.4404 & -0.1492 & 0.4780 & 0.2128 & 0.2256 & 0.2263 \\ 0.4355 & 0.6580 & 0.7662 & 1.2378 & 1.1826 & 0.7480 & 0.2891 & 0.0953 & 0.6125 & -0.0939 & -0.1007 & 0.2343 \\ 0.6837 & 0.5275 & 0.7402 & 1.1826 & 1.8114 & 1.1471 & 0.7054 & 0.7334 & 0.9955 & -0.5115 & -0.2965 & 0.0559 \\ 0.8893 & 1.1619 & 0.7701 & 0.7480 & 1.1471 & 2.3765 & 1.3109 & 0.9366 & 0.6296 & -0.6636 & -0.1367 & 0.0969 \\ 0.2403 & 0.4746 & 0.4404 & 0.2891 & 0.7054 & 1.3109 & 1.5761 & 0.8621 & 0.5305 & -0.2873 & -0.2767 & 0.2882 \\ 0.5067 & 0.0759 & -0.1492 & 0.0953 & 0.7334 & 0.9366 & 0.8621 & 2.7845 & 1.6242 & -0.1564 & -0.2836 & 0.1032 \\ 0.4804 & 0.3457 & 0.4780 & 0.6125 & 0.9955 & 0.6296 & 0.5305 & 1.6242 & 3.5719 & 1.3096 & 0.5810 & 0.5213 \\ -0.3263 & -0.0996 & 0.2128 & -0.0939 & -0.5115 & -0.6636 & -0.2873 & -0.1564 & 1.3096 & 2.6819 & 1.5615 & 0.9351 \\ -0.1273 & 0.1599 & 0.2256 & -0.1007 & -0.2965 & -0.1367 & -0.2767 & -0.2836 & 0.5810 & 1.5615 & 1.6886 & 0.6328 \\ -0.1017 & 0.0420 & 0.2263 & 0.2343 & 0.0559 & 0.0969 & 0.2882 & 0.1032 & 0.5213 & 0.9351 & 0.6328 & 1.4292 \end{bmatrix}$$

The Pure Bayesian method was used to estimate k and ξ and the following estimates were obtained:

$$\hat{\underline{k}} = [1,3925; 1,4329; 1,4870; 1,4302; 1,4639; 1,4385; 1,4749; 1,4756; 1,4212; 1,4642; 1,5002; 1,4131]$$

and

$$\hat{\underline{\xi}} = [0,1427; 0,1544; 0,1551; 0,1505; 0,1495; 0,1491; 0,1562; 0,1542; 0,1439; 0,1662; 0,1639; 0,1446].$$

In this chapter the maximum monthly inflows are modelled by using multivariate regression.

In Section 3.2 the theory around multivariate regression and estimation with regression are discussed.

In Section 3.3 the parameters of the data set, the maximum inflow of water into the Gariep Dam in each month, are estimated by using multivariate regression and the Pure Bayesian approach. SOI values are introduced as

covariates. SOI stands for the Southern Oscillation Index that represents the sea level pressure between Tahiti and Darwin. This data set was obtained from the website of the National Weather Service (2007).

In Section 3.4 future monthly maximum inflow values are predicted using the SOI values as covariates. Section 3.5 shows another application, the estimation of tail probabilities. A conclusion is given in Section 3.6.

3.2 Multivariate Regression

Assume that $\underline{Y} = \left(y_1, y_2, ..., y_p\right)$ is an $n \times p$ matrix of observations on *p* variables, and that $\underline{Y} = \left(y_1, y_2, ..., y_p\right)$ is generated by the multivariate model

$$E\left(\underline{Y}\right) = \underline{X}\,\underline{\beta} \tag{3.1}$$

where $\underline{X}$ is an $n \times \left(p+1\right)$ matrix, with rank *p* of given observations on *p* independent variables and $\underline{\beta} = \left(\beta_1, \beta_2, ..., \beta_p\right)$ is a $\left(p+1\right) \times p$ matrix of regression parameters (Zellner 1996, p. 224 - 225). Due to the large number of parameters in the model the Bayesian approach is not considered for estimation we considered simpler methods. μ and Σ are estimated using the method of moments approach and β is estimated using least squares.

3.2.1 Least squares estimation

The least squares estimate for $\underline{\beta}$ is given by

$$\underline{\hat{\beta}} = \left(\underline{X}'\underline{X}\right)^{-1}\underline{X}'\underline{Y} \tag{3.2}$$

and

$$\hat{S}_e = \frac{\left(\underline{Y} - \underline{X}\,\underline{\hat{\beta}}\right)' \left(\underline{Y} - \underline{X}\,\underline{\hat{\beta}}\right)}{n - p - 1} \qquad (3.3)$$

is an estimate of the covariance matrix $\sum$ (Rencher 2002, p. 338-340).

3.3 Estimating the parameters of the Gariep Dam data set

Multivariate regression is now applied to the maximum monthly inflows into the Gariep Dam to estimate $\underline{\mu}$ and $\sum$. In Chapter 2 $\underline{\mu}$ and $\sum$ were estimated by only considering the data values below a threshold. For the Gariep data set the threshold was chosen as *t* = 5374,9, by making use of the entropy of the Dirichlet process (this was discussed in Chapter 2, Section 2.5). From Chapter 2, Definition 2.2, it was also shown that the maximum monthly inflow data must be transformed as follows:

$$Y_i = -\log(X_i), i = 1, ..., p\,. \qquad (3.4)$$

After transforming the data below the threshold the multivariate regression approach in Section 3.2 is followed where $\underline{Y}$ is an $n \times p$ matrix consisting of the monthly maximum inflows below the threshold, *n* is the number of data points below the threshold and *p* is the number of variables. For this study *n* = 25 and *p* = 12 (*p* denotes the different months, January up to December). $\underline{X}$ is an $n \times (p+1)$ matrix, thus for this study $\underline{X}$ is an 25×13 matrix. $\underline{X}$ incorporates the SOI values as covariates. The SOI values influence the inflow values with a lag of three months (De Waal & Van Wyk, 1998). For example, the inflow of January will be influenced by the SOI value of four months earlier, which is September of the previous year.

For this example the following multivariate regression model is introduced when the covariates are taken into consideration:

$$\hat{\mu}_i = E(Y_i) = \beta_{0i} + \beta_{1i}\mathrm{SOI}_{i-4} + \beta_{2i}\mathrm{SOI}_{i-3} + ... + \beta_{pi}\mathrm{SOI}_{i-1}, i = 1, ..., p\,. \qquad (3.5)$$

By using the least squares estimation method, $\underline{\beta}$ and $\sum$ are estimated as

$$\underline{\hat{\beta}} = \begin{bmatrix} -5.9791 & 0.0150 & 0.3158 & 0.1672 & -0.3371 & -0.2802 & 0.1918 & 0.2105 & 0.2691 & -0.2258 & 0.4520 & 0.1761 & 0.1827 \\ -5.4108 & 0.4624 & 0.0210 & 0.0820 & 0.0165 & 0.1194 & 0.1989 & 0.6097 & 0.3703 & -0.2232 & -0.0019 & -0.0846 & 0.3392 \\ -6.2183 & 0.2013 & 0.0512 & 0.0262 & -0.2260 & 0.0325 & -0.0321 & 0.3479 & 0.1962 & -0.3285 & 0.1378 & 0.0679 & 0.3580 \\ -5.8383 & 0.2359 & 0.2084 & 0.0098 & -0.0165 & 0.0257 & 0.0937 & 0.0060 & 0.1313 & -0.0981 & 0.0949 & -0.0359 & 0.2516 \\ -4.9484 & 0.0577 & 0.1613 & 0.1072 & 0.0822 & -0.1130 & 0.3051 & -0.0105 & 0.1024 & -0.2362 & 0.2298 & -0.0932 & 0.3297 \\ -1.9330 & 0.5703 & 0.1966 & 0.3014 & 0.2746 & 0.2863 & 0.8656 & 0.8365 & 0.1721 & -0.0102 & 0.0528 & -0.1464 & 0.1683 \\ -3.6654 & 0.2071 & -0.1651 & 0.0378 & 0.1177 & 0.3191 & 0.3766 & 0.2562 & -0.0695 & -0.0631 & 0.2190 & -0.3161 & 0.1957 \\ -2.9253 & 0.2841 & -0.1281 & 0.2105 & 0.4002 & 0.2693 & 0.6459 & 0.2189 & 0.4099 & 0.0521 & -0.1142 & -0.3763 & -0.0536 \\ -5.0038 & 0.2477 & -0.1359 & -0.0638 & 0.4430 & 0.5262 & 0.4538 & -0.1225 & 0.4483 & -0.2376 & 0.2081 & 0.1055 & 0.3426 \\ -6.4068 & -0.1547 & 0.2856 & -0.1391 & 0.1359 & 0.6271 & -0.1058 & 0.0207 & -0.6237 & 0.1293 & -0.1699 & -0.2439 & -0.1831 \\ -5.2966 & 0.1063 & 0.4615 & 0.1133 & 0.2890 & 0.4135 & 0.3596 & 0.2302 & -0.3823 & 0.1938 & -0.4472 & -0.3687 & -0.2501 \\ -7.4690 & 0.4099 & 0.1040 & -0.0587 & 0.2046 & 0.9433 & 0.2295 & -0.3465 & -0.5383 & 0.1122 & -0.2938 & -0.5676 & -0.1680 \end{bmatrix}$$

and

$$\hat{\Sigma} = \begin{bmatrix} 1.0937 & 0.4777 & 0.3941 & -0.1337 & 0.0981 & 0.1094 & -0.3067 & 0.6498 & -0.1856 & -0.5922 & -0.4074 & -0.1829 \\ 0.4777 & 0.7950 & 0.3865 & -0.1380 & -0.3284 & -0.1220 & -0.3133 & -0.2072 & -0.6784 & -0.1929 & 0.1212 & 0.0187 \\ 0.3941 & 0.3865 & 0.9065 & 0.2578 & 0.2013 & -0.2345 & -0.1412 & -0.3267 & -0.2898 & 0.2286 & 0.3845 & 0.4196 \\ -0.1337 & -0.1380 & 0.2578 & 1.4270 & 1.4030 & 0.2559 & -0.0668 & 0.1424 & 0.3071 & -0.0816 & -0.2230 & 0.1959 \\ 0.0981 & -0.3284 & 0.2013 & 1.4030 & 2.0700 & 0.7418 & 0.4656 & 0.8569 & 0.5387 & -0.6077 & -0.6439 & -0.1395 \\ 0.1094 & -0.1220 & -0.2345 & 0.2559 & 0.7418 & 1.0533 & 0.8808 & 0.5355 & -0.6489 & -1.3438 & -1.1123 & -0.4160 \\ -0.3067 & -0.3133 & -0.1412 & -0.0668 & 0.4656 & 0.8808 & 1.6086 & 0.8159 & -0.3142 & -0.8328 & -0.7041 & -0.3041 \\ 0.6498 & -0.2072 & -0.3267 & 0.1424 & 0.8569 & 0.5355 & 0.8159 & 3.6574 & 1.7788 & -0.2180 & -1.0950 & -0.5644 \\ -0.1856 & -0.6784 & -0.2898 & 0.3071 & 0.5387 & -0.6489 & -0.3142 & 1.7788 & 3.4791 & 2.0590 & 0.8895 & 0.1275 \\ -0.5922 & -0.1929 & 0.2286 & -0.0816 & -0.6077 & -1.3438 & -0.8328 & -0.2180 & 2.0590 & 2.9543 & 1.9105 & 0.9066 \\ -0.4074 & 0.1212 & 0.3845 & -0.2230 & -0.6439 & -1.1123 & -0.7041 & -1.0950 & 0.8895 & 1.9105 & 1.9902 & 0.5173 \\ -0.1829 & 0.0187 & 0.4196 & 0.1959 & -0.1395 & -0.4160 & -0.3041 & -0.5644 & 0.1275 & 0.9066 & 0.5173 & 0.6295 \end{bmatrix}$$

Now the formula (3.1) given in Section 3.2 is used to estimate $\underline{\mu}$, $\underline{\hat{\mu}} = E(\underline{Y}) = \underline{X}\underline{\hat{\beta}}$.

The correlation coefficient between $\underline{Y}$ and $\underline{X}\underline{\hat{\beta}}$ is calculated as 0,8033, which can be considered as high. The coefficient of determination (R^2) is 0,6453, therefore 65% of variability under the threshold is explained by the SOI covariates.

After obtaining $\underline{\hat{\mu}}$ and $\hat{\Sigma}$, $\underline{k}$ and $\underline{\xi}$ are estimated by using the Pure Bayesian approach discussed in Chapter 2, Section 2.4.1.3. The same threshold as before is used and the following estimates are obtained:

$\hat{\underline{k}} = [2,0475; 1,6118; 1,4612; 1,388; 1,5916; 1,7424; 1,5043; 1,8411; 1,5672; 1,2729; 1,9771; 1,7957]$

and

$\hat{\underline{\xi}} = [0,2916; 0,1715; 0,1481; 0,1547; 0,1886; 0,1939; 0,1586; 0,2418; 0,178; 0,1423; 0,2619; 0,243]$.

3.4 Predicting future monthly maximum inflows

After successfully estimating the parameters, predictions for the maximum monthly inflows are made for each year, given specific SOI values. These predictions are compared to the actual observed maximum monthly inflows for the specific year. For example, Table 3.1 shows the maximum monthly inflows predicted for 2006, given specific SOI values, and the actual maximum monthly inflows for 2006.

Table 3.1 The predicted maximum inflow values for 2006 compared to the actual maximum inflow values of 2006

	Jan	**Feb**	**Mar**	**Apr**	**May**	**June**
Predict	267,7	199,5	207	89,4	63,9	150,9
SOI	0,7	1,9	-0,5	-0,3	2,9	-0,3
Actual	1102,3	2976,92	1791,32	933,12	1258,61	347,4

	July	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**
Predict	236	81,7	797,4	2665	736,2	1445,1
SOI	2,3	1,5	-1,4	-1,1	-1,3	-2,7
Actual	106,11	627,46	496,39	419,97	6273,48	873,01

The correlation coefficient between the actual observed maximum monthly inflows and the predicted maximum monthly inflows over a period of 29 years is calculated as 0,4293. The coefficient of determination (R^2) is 0,1843.

Therefore a small amount of 18% of variability above the threshold is explained by the SOI covariates.

The predictions in Table 3.1 differ from the predictions made in Chapter 2, therefore it can be concluded that the SOI values have an influence on the predictions. This is shown in Table 3.2. The second column represents the predictions of the expected maximum monthly inflow for different months and the third column represents the predictions of the maximum monthly inflow for the year 2006 with specific SOI values.

Table 3.2 Predicted maximum monthly inflow values, with and without SOI as a covariate

	Prediction of maximum monthly inflow without SOI	**Prediction of maximum monthly inflow with SOI**
January	810,1379	267,7
February	910,8510	199,5
March	777,9235	207
April	423,5936	89,4
May	205,4390	63,9
June	107,0454	105,9
July	54,8819	236
August	91,9005	81,7
September	221,0192	797,4
October	505,4896	2665
November	676,2469	736,2
December	696,5503	1445,1

3.5 Estimating Tail Probabilities

Tail probabilities can be estimated by simulating a data set with the estimated parameter values in Section 3.3, conditional on specific SOI values. A data

set of size 50000×12 is simulated and the following probability is estimated for a specific vector of SOI values:

$$P\begin{pmatrix} X_1 > 334; X_2 > 11263; X_3 > 6865; X_4 > 2524; X_5 > 1591; X_6 > 1328; \\ X_7 > 538; X_8 > 2148; X_9 > 3116; X_{10} > 5355; X_{11} > 6273; X_{12} > 4164 \end{pmatrix}, \tag{3.6}$$

$x_i, i = 1,..,12$ is some maximum extreme inflow for each month.

For the following SOI values:

SOI = $\begin{bmatrix} 0 \\ . \\ . \\ . \\ 0 \end{bmatrix}$, the $P\left(X_1 > x_1,..., X_{12} > x_{12} \mid \text{SOI}\right) = \frac{6}{50000} = 1,2 \times 10^{-04}.$

For the SOI values equal to

SOI = $\begin{bmatrix} -5 \\ . \\ . \\ . \\ -5 \end{bmatrix}$, the $P\left(X_1 > x_1,..., X_{12} > x_{12} \mid \text{SOI}\right) = \frac{0}{50000} = 0.$

For the SOI values of

SOI= $\begin{bmatrix} 5 \\ . \\ . \\ . \\ 5 \end{bmatrix}$, the $P\left(X_1 > x_1,..., X_{12} > x_{12} \mid \text{SOI}\right) = \frac{861}{50000} = 0,0172.$

And for the SOI values of

$$\text{SOI} = \begin{bmatrix} 6 \\ 7 \\ 4 \\ 3 \\ -2 \\ -1 \\ -2 \\ 2 \\ 5 \\ 6 \\ 5 \\ 7 \end{bmatrix}, \text{ the } P\left(X_1 > x_1, ..., X_{12} > x_{12} \mid \text{SOI}\right) = \frac{164}{50000} = 3,28 \times 10^{-03}.$$

Positive SOI values are an indication that it will rain, therefore extreme maximum inflow values are more likely with larger (between 7 and 10) SOI values.

The following tail probability

$$P\begin{pmatrix} X_1 > 1102,3; X_2 > 2976,92; X_3 > 1791,32; X_4 > 933,12; X_5 > 1258,61; X_6 > 347,4; \\ X_7 > 106,11; X_8 > 627,46; X_9 > 496,39; X_{10} > 416,97; X_{11} > 6273,48; X_{12} > 873,01 \end{pmatrix}$$

given the SOI values in Table 3.1 for the year 2006 is calculated as $\frac{4}{50000} = 8 \times 10^{-05}$.

3.6 Conclusion

From the predictions in Section 3.4, SOI seems to be an appropriate covariate to consider when the extreme values are ignored and when only the data under the threshold is considered. When working with the whole data set, including extreme values, it seems that not only the SOI should be considered but more covariates should be taken into consideration. There are a lot of different covariates that can play a role when modelling the maximum monthly inflows such as temperature, rainfall *etc*.

Chapter 4

The selection of a tail sample fraction

4.1 Introduction

In this chapter only data above a threshold *t* is considered, therefore the optimal sample size above a threshold is discussed and various methods are considered for choosing the optimal sample size.

In some cases there can be a clear physical reason to select a certain threshold. In such a case there is no preference for a certain statistical method. This chapter assumes that there is no clear indication, from the data or otherwise, where a threshold should be chosen.

The underlying distribution is assumed to be from the Pareto-type family of distributions discussed in Chapter 1, Section 1.1.3, where $\gamma > 0$. It was mentioned in Chapter 1 that the Fréchet distribution and the Burr (type III) distribution are examples of Pareto-type distributions. Therefore, in Section 4.2, a data set from a Fréchet distribution and a data set from a Burr (type III) distribution are simulated. In Section 4.3 a threshold (*t*) is selected by choosing an optimal *k* value. Four methods are discussed on how to select an optimal k value. In Section 4.4 the Generalized Pareto distribution (GPD) is fitted to the data above the threshold, and the underlying parameters of the GPD and large quantiles are estimated. Section 4.4 also includes QQ-plots of the data above the chosen thresholds to see if the methods for selecting a threshold are suitable. In Section 4.5 the goodness of the Dirichlet process is discussed in more detail by using a simulation study. A conclusion is given in Section 4.6.

4.2 Simulated data sets

Before simulating the observations, the quantile function is discussed. This discussion follows from Beirlant *et al*. (2004, p. 1-5). Typically, when working with sample data the properties of the distribution function

$$F(x) = P(X \leq x) \tag{4.1}$$

are studied. An alternative is to study the properties of the inverse function, called the quantile function, given by the equation

$$Q(p) = \inf\{x : F(x) \geq p\}, p \in (0,1). \tag{4.2}$$

First the quantile function of the exponential model is discussed. The standard exponential distribution has the general survival function

$$1 - F_{\lambda}(x) = \exp(-\lambda x). \tag{4.3}$$

The quantile function is given by the equation

$$Q_{\lambda}(p) = \frac{-1}{\lambda}\log(1-p), \text{ for } p \in (0,1). \tag{4.4}$$

A practical choice of values for p is $p \in \left\{\frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}, 1\right\}$.

In this chapter data sets are simulated from a Fréchet distribution and from a Burr (type III) distribution. In both cases the size of the simulated data set is $n = 1000$. The survival function of the Fréchet distribution is given by the equation

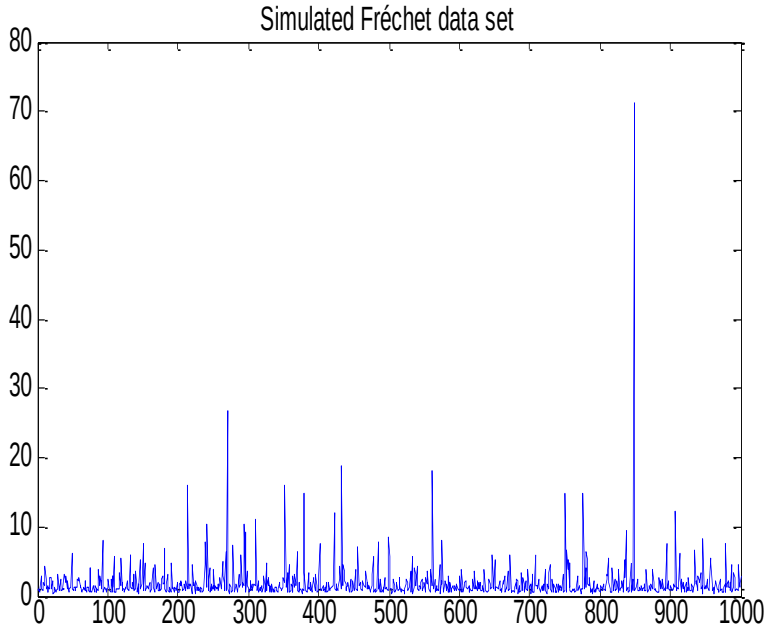
$$1 - F(x) = 1 - \exp\left(-x^{-\alpha}\right), x > 0, \alpha > 0 \tag{4.5}$$

(Beirlant *et al*. 2004, p. 59) and the quantile function of the Fréchet distribution is given as

$$Q(p) = \left[-\log(p)\right]^{\frac{-1}{\alpha}}, \alpha > 0; p \in (0,1). \tag{4.6}$$

The EVI, γ, of the Fréchet distribution is $\frac{1}{\alpha}$ (Beirlant *et al*. 2004, p. 59). A 1000 observations are simulated from a Fréchet distribution with $\alpha = 1{,}9$, $\gamma = 0{,}5263$, and this simulated data set is shown in Figure 4.1.

Fig. 4.1 A simulated data set from a Fréchet distribution



The survival function of the Burr (type III) distribution is given by the equation

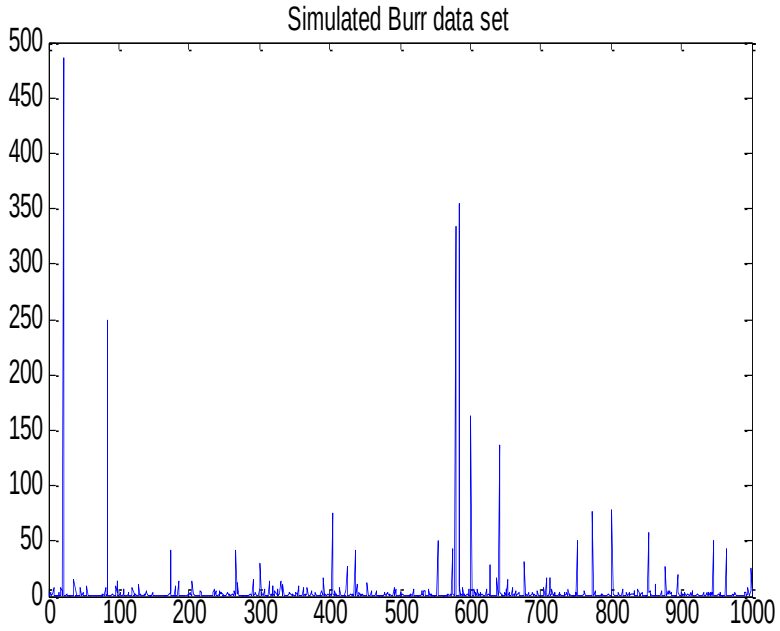
$$1 - F(x) = 1 - \left(\frac{\eta}{\eta + x^{-\tau}} \right)^{\lambda}, x > 0; \eta, \tau, \lambda > 0 \qquad (4.7)$$

(Beirlant *et al*. 2004, p. 59) and the quantile function of the Burr (type III) distribution is given as

$$Q(p) = \left[\eta \left(p^{\frac{-1}{\lambda}} - 1 \right) \right]^{-\frac{1}{\tau}}, \eta, \tau, \lambda > 0; p \in (0,1). \qquad (4.8)$$

The EVI, γ, of the Burr (type III) distribution is $\frac{1}{\tau}$ (Beirlant *et al*. 2004, p. 59). A 1000 observations are simulated from a Burr (type III) distribution with $\eta = 3,9, \tau = 1,2$ and $\lambda = 2,5$ and this simulated data set is shown in Figure 4.2.

Fig. 4.2 A simulated data set from a Burr (type III) distribution



4.3 Methods for selecting the optimum k value

Because the bias is the main problem of the Hill estimator (discussed in Section 1.1.3.3 of Chapter 1), $b_{n,k}$, the parameter that dominates the bias of the Hill estimator, is very useful in selecting a k value for which the bias in the Hill estimator is not too large.

For each simulated data set different methods are investigated to choose an optimum tail sample fraction, thus to select an optimum k value.

4.3.1 The Guillou and Hall method

This method is discussed by Beirlant *et al*. (2004, p. 123). Guillou and Hall (2001) proposed a method to choose $H_{\hat{k},n}$, the Hill estimator, where $\hat{k}$ is the smallest value of k for which

$$\sqrt{\frac{k}{12}}\frac{|\hat{b}_{LS}^{+}|}{H_{k,n}} > C_{crit} \tag{4.9}$$

where C_{crit} is some critical value such as 1,25 or 1,5.

After setting $\beta = 1$ in (1.30) of Chapter 1, $\hat{b}_{LS}^{+}$ is called the least squares estimator of $b_{n,k}$. From (1.32) of Chapter 1

$$\hat{b}_{LS}^{(+)} = \frac{12}{k}\sum_{j=1}^{k}\left(\left(\frac{j}{k+1}\right)-\frac{1}{2}\right)Z_j. \tag{4.10}$$

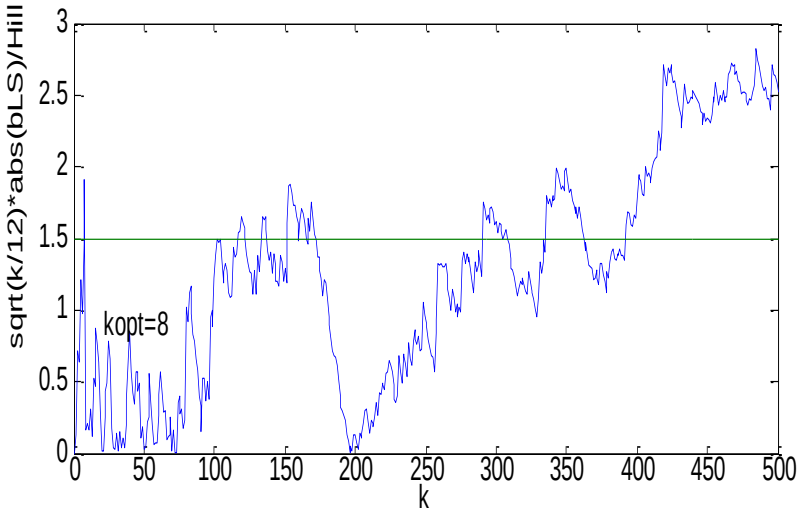
When using (1.29) of Chapter 1, it can be shown that, when Z_j is approximated by γf_j, $j = 1,...,k$,

$$Var\left(\hat{b}_{LS}^{(+)}\right) \sim 12\gamma^2/k \text{ for } k \to \infty. \tag{4.11}$$

This motivates (4.9) (Beirlant *et al*. 2004, p. 123).

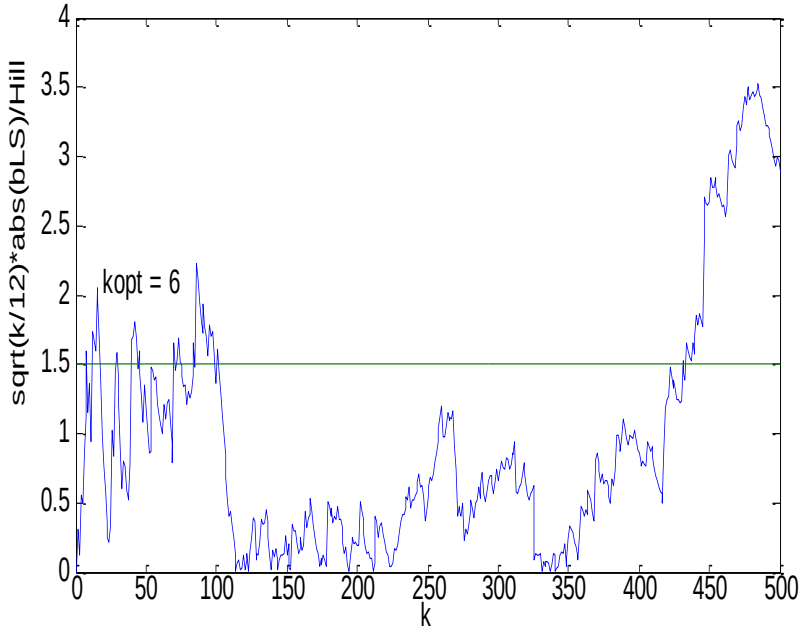
When applying this method to the Fréchet simulated data set k is chosen as 8. Therefore, only the top 8 ordered data entries are considered. Thus the threshold is the 991st ordered data entry and $t = 12{,}1735$. Figure 4.3 illustrates the Guillou and Hall method for the Fréchet simulated data set.

Fig. 4.3 **Choosing k for the simulated Fréchet data set by making use of the Guillou and Hall method**



When applying this method to the Burr (type III) simulated data set, k is chosen as 6. Therefore, only the top 6 ordered data entries are considered. The threshold is therefore the 993rd ordered data entry and $t = 76,9797$. Figure 4.4 illustrates the Guillou and Hall method for the Burr (type III) simulated data set.

Fig. 4.4 **Choosing k for the simulated Burr (type III) data set by making use of the Guillou and Hall method**



4.3.2 Method of minimizing the mean squared error

Beirlant *et al*. (2004, p. 125) discusses another way of obtaining an optimal value for k. This method consists of obtaining a value for k for which the asymptotic mean squared error of $H_{k,n}$ is minimized. The asymptotic mean squared error is given by

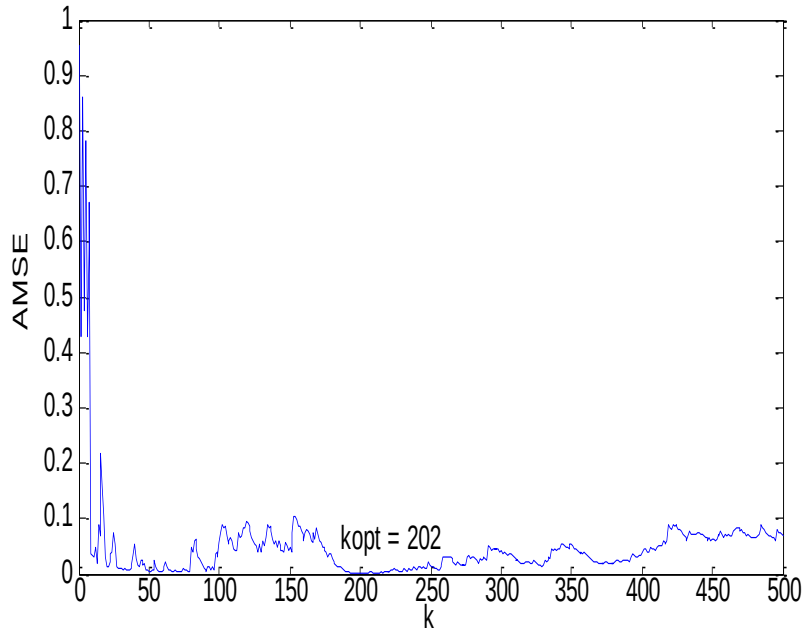
$$AMSE\left(H_{k,n}\right) = AVar\left(H_{k,n}\right) + ABias^2\left(H_{k,n}\right) = \frac{\gamma^2}{k} + \left(\frac{b_{n,k}}{1+\beta}\right)^2. \quad (4.12)$$

Therefore, one should find an estimator of k ($\hat{k}$) where the plot $\left\{\left(k, AMSE\left(H_{k,n}\right)\right), k = 1,...,n-1\right\}$ achieves a minimum.

When applying this method the least squares estimates are chosen as follows: $\hat{\beta} = 0$, $\hat{\gamma}$ is the Hill estimator and $\hat{b}_{n,k}$ is $\hat{b}_{LS}^{+}$, introduced in the Guillou

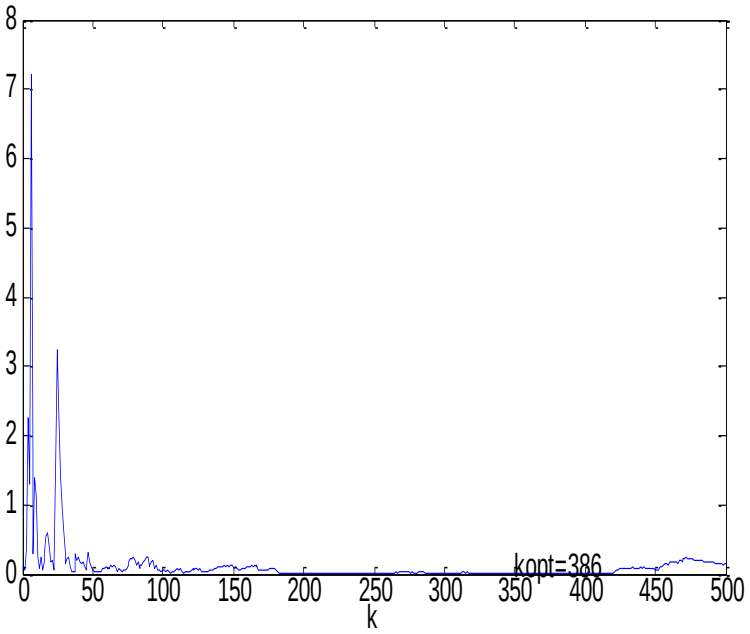
and Hall method. For the Fréchet simulated data set the mean squared error plot reaches a minimum at $k = 202$, therefore only the top 202 ordered data entries are considered. The threshold is the 797th ordered data entry; therefore $t = 2,2559$. Figure 4.5 shows that for $k = 202$ the AMSE plot reaches a minimum.

Fig. 4.5 Choosing k for the simulated Fréchet data set by making use of the mean squared error plot



For the Burr (type III) simulated data set the mean squared error plot reaches a minimum at $k = 386$, therefore only the top 386 ordered data entries are considered. The threshold is the 613th ordered data entry; therefore $t = 1,2239$. Figure 4.6 shows that for $k = 386$ the AMSE plot reaches a minimum.

Fig. 4.6 **Choosing k for the simulated Burr (type III) data set by making use of the mean squared error plot**



4.3.3 The method of the median of the optimal k-values

Beirlant *et al*. (2004, p. 125-126) shows that the asymptotic mean squared error of the Hill estimator is minimal for

$$k_{n,opt} \sim \left(b^2\left(n\right)\right)^{-\frac{1}{(1+2\beta)}} \left(\frac{\gamma^2\left(1+\beta\right)^2}{2\beta}\right)^{\frac{1}{(1+2\beta)}}, n \to \infty. \qquad (4.13)$$

Because of the regular varying behaviour of b the following is obtained:

$$k_{n,opt} \sim \left[b^2\left(\frac{n}{k_0}\right)\right]^{\frac{-1}{(1+2\beta)}} k_0^{\frac{2\beta}{(1+2\beta)}} \left(\frac{\gamma^2\left(1+\beta\right)^2}{2\beta}\right)^{\frac{1}{(1+2\beta)}} . \qquad (4.14)$$

Consistent estimators of b_{n,k_0}, β and γ are plugged into (4.14). For this method the estimates are as follows: $\hat{\gamma}$ is the Hill estimator, $\hat{b}_{n,k_0}$ is $\hat{b}^{+}_{LS}$ in the Guillou and Hall method and β=1.

Now, for each k_0 an estimator of $k_{n,opt}$ is obtained by the following equation:

$$\hat{k}_{n,k_0} = \left[\hat{b}^2_{n,k_0}\right]^{\frac{-1}{(1+2\hat{\beta})}} k_0^{\frac{2\hat{\beta}}{(1+2\hat{\beta})}} \left(\frac{\hat{\gamma}^2 \left(1+\hat{\beta}\right)^2}{2\hat{\beta}} \right)^{\frac{1}{(1+2\hat{\beta})}} . \qquad (4.15)$$

An overall estimate of $k_{n,opt}$ is the median of the first $\frac{n}{2}$ $\hat{k}_{n,k_0}$ values, calculated as follows:

$$\hat{k}_{n,med} = median\left\{ \hat{k}_{n,k_0} : k_0 = 3,...,\frac{n}{2} \right\} . \qquad (4.16)$$

In the case of the Fréchet simulated data set the median of the first $\frac{n}{2}$ $\hat{k}_{n,k_0}$ values is 155,8312. Therefore, the optimal k-value is 155 and only the top 155 ordered data entries are considered. Therefore, the threshold is the 849th data entry and $t = 2,6520$.

In the case of the Burr (type III) simulated data set the median of the first $\frac{n}{2}$ $\hat{k}_{n,k_0}$ values is 173,7058. Therefore, the optimal k-value is 173 and only the top 173 ordered data entries are considered. Therefore, the threshold is the 826th data entry and $t = 2,8963$.

4.3.4 Method of the Dirichlet process

The work discussed here is based on the work done by De Waal *et al*. (2007).

Let $z_j = \sqrt{\sum_{i=1}^{2} X_{ij}^2}$, $j = 1,\ldots,n$ and $i = 1,\ldots,d$ be the distance from each observation to the origin.

When working with the Dirichlet process an underlying distribution must be assumed. In this case the Strict Pareto distribution is assumed, the Strict Pareto has only one parameter γ, the EVI. If Z is Strict Pareto distributed it can be described by the distribution function

$$F\left(z \mid t\right) = 1 - \left(\frac{z}{t}\right)^{\frac{-1}{\gamma}}, \qquad (4.17)$$

where t denotes the threshold. Thus, only data above the threshold t is considered.

The Dirichlet process implies that a threshold t is selected through the entropy of the Dirichlet distribution. Suppose the true distribution function of Z conditional on t is $F^*(z|t)$ and we assume $F(z|t)$ given in (4.17) as our best belief to describe $F^*(z|t)$. $F^*(z|t)$ is described through a Dirichlet process with parameters $F(z|t)$ and β (Ferguson 1973,1983; Lo 1984; Mazzucchi 2002).

The Dirichlet process is a nonparametric process where the spacings in the probabilities of occurrences of the observations are assumed to be Dirichlet distributed. The concentration parameter β is chosen as $\beta = k+1$.

Let $\nu_1 = \beta F\left(z_1 \mid t\right)$, $\nu_2 = \beta\left\{F\left(z_2 \mid t\right) - F\left(z_3 \mid t\right)\right\}, \ldots, \nu_{k+1} = \beta\left\{1 - F\left(z_k \mid t\right)\right\}$ and $Y_i,\ i = 1,\ldots,k+1$ be the random variables.

For $Y_1 = F^*(z_1|t)$, $Y_2 = F^*(z_2|t) - F^*(z_3|t),\ldots,Y_{k+1} = 1 - F^*(z_k|t)$ the Dirichlet distribution defines $Y_1,\ldots,Y_{k+1}$ to be Dirichlet distributed with parameters $\nu_1,\ldots,\nu_{k+1}$, therefore $Y_1,\ldots,Y_{k+1} \sim D\left(\nu_1,\ldots,\nu_{k+1}\right)$.

The selection of the threshold *t* is done by considering the entropy of the Dirichlet distribution. The entropy captures information from the data; the idea is to look for the threshold where the information gathered from the data is optimal. This is equivalent to minimizing the negative differential entropy (NDE) relative to the lower bound. Honkela (2001) showed that the NDE for $y(Y_1,\ldots,Y_{k+1})$ is given by

$$\begin{aligned} J_{k+1} &= E\log y\left(Y_1,...,Y_{k+1}\right) \\ &= \log\Gamma\left(\nu_0\right)-\sum_{i=1}^{k+1}\log\Gamma\left(\nu_i\right)+\sum_{i=1}^{k+1}\left(\nu_i-1\right)\left\{\psi\left(\nu_i\right)-\psi\left(\nu_0\right)\right\} \end{aligned} \tag{4.18}$$

where $y(Y_1,\ldots,Y_{k+1})$ is the joint density of $Y_1,\ldots,Y_{k+1}$; $\nu_0=\sum_{i=1}^{k+1}\nu_i$ and $\psi(\nu)=\frac{\partial}{\partial\nu}\log\Gamma(\nu)$ (called the digamma function).

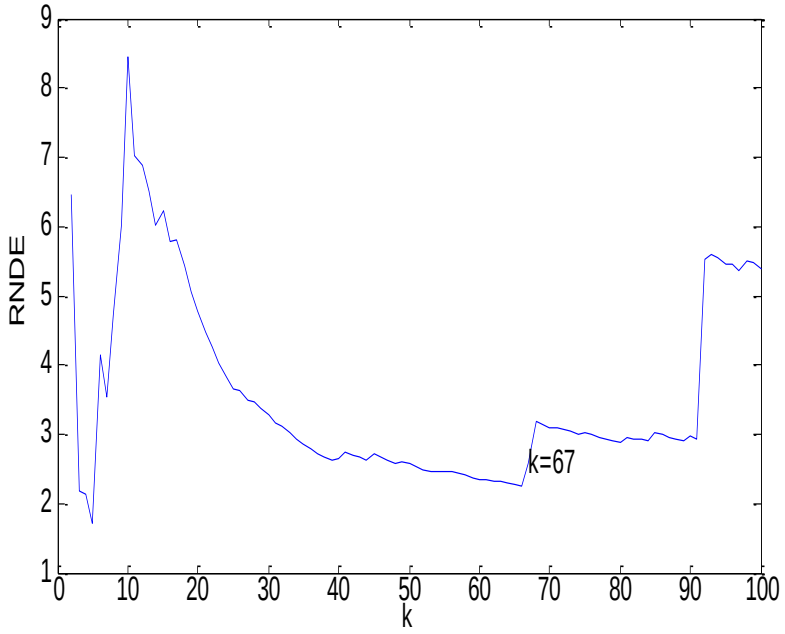
From (4.18) it follows that the lower bound (J_{k+1}) is reached when $\nu_i=1, i=1,...,k+1$, namely $\Gamma(k+1)$. This implies that $\beta=k+1$ and the spacings on the data are such that the information of the data to $F(z|t)$ is maximal.

The aim is, given a dataset $z_1,\ldots,z_k$, to select *k* (or threshold *t*) as large as possible such that the NDE is close to $\log\Gamma(k+1)$.

The Hill estimate is considered as an estimate for γ in (4.17).

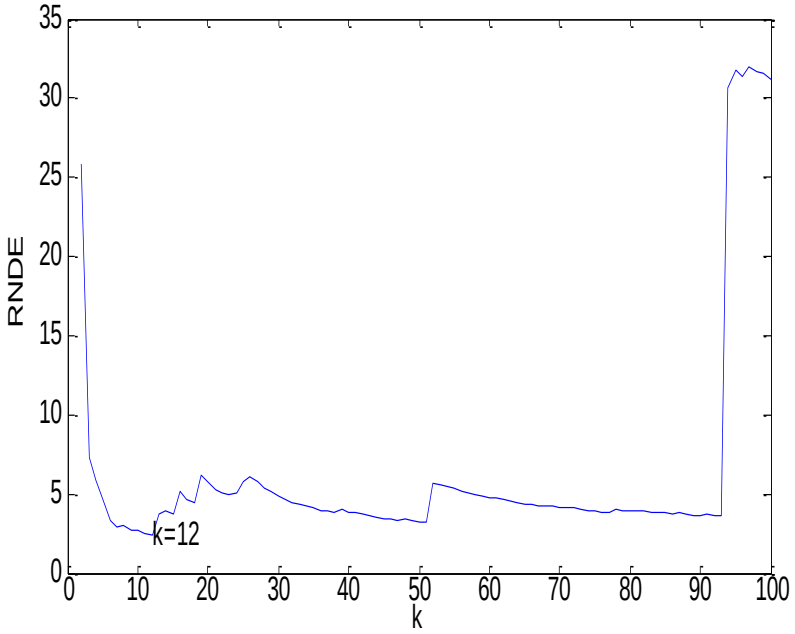
In the case of the Frechet simulated data set the NDE is the closest to $\log\Gamma(k+1)$ when $k=67$. Thus only the top 47 ordered data entries are considered. Therefore, the threshold is the 933rd ordered data entry and $t=4,4929$. Figure 4.7 shows the NDE of the Dirichlet process in the case of the simulated Fréchet data set.

Fig. 4.7 The NDE of the Dirichlet process in the case of the simulated Fréchet data set



In the case of the Burr (type III) simulated data set the NDE is the closest to $\log\Gamma(k+1)$ when $k=12$. Thus only the top 12 ordered data entries are considered. Therefore, the threshold is the 987th ordered data entry and $t=42,6215$. Figure 4.8 shows the NDE of the Dirichlet process in the case of the simulated Burr (type III) data set.

Fig. 4.8 **The NDE of the Dirichlet process in the case of the simulated Burr (type III) data set**



4.4 Goodness of fit

In this section the goodness of fit of the above mentioned methods is evaluated.

4.4.1 Estimation of parameters

In Section 1.1.4.3 of Chapter 1, the peaks over threshold (POT) method was discussed where the interest lies in the exceedances over a threshold. In the POT method the GPD in (1.43) is fitted to the data above the threshold t. The absolute exceedances are given by $Y_j = X_i - t$, where $X_i > t, j = 1,...,N_t$ and i is the index of the j^{th} exceedance in the original sample. N_t represents the number of exceedances. In this section, the GPD is fitted to the data above the thresholds of the simulated Fréchet and Burr (type III) distributions. The parameters of the GPD can be estimated by various methods such as the maximum likelihood method, the method of moments, a Bayesian method *etc*.

In the following sections the maximum likelihood method and the Bayesian method are discussed in more detail.

4.4.1.1 Maximum likelihood method

For a given sample $Y_1,...,Y_{N_t}$ of i.i.d. Generalized Pareto variables the log-likelihood is given by the equation

$$\log L(\sigma,\gamma) = -N_t \log\sigma - \left(\frac{1}{\gamma}+1\right)\sum \log\left(1+\frac{\gamma Y_i}{\sigma}\right), 1+\frac{\gamma Y_i}{\sigma} > 0; i = 1,..., N_t. \tag{4.19}$$

When maximizing $\log L(\sigma,\gamma)$ the following re-parameterization can be performed:

$(\sigma,\gamma) \to (\tau,\gamma)$ where $\tau = \frac{\gamma}{\sigma}$

therefore

$$\log L(\tau,\gamma) = -N_t \log\gamma + N_t \log\tau - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{N_t} \log(1+Y_i) \tag{4.20}$$

(Beirlant *et al*. 2004, p. 149).

The maximum likelihood estimators follow from

$$\frac{1}{\hat{\tau}} + \left(\frac{1}{\hat{\tau}}+1\right)\frac{1}{N_t}\sum_{i=1}^{N_t}\frac{Y_i}{1+\hat{\tau}Y_i} = 0 \tag{4.21}$$

where

$$\hat{\gamma} = \frac{1}{N_t}\sum_{i=1}^{N_t}\log\left(1+\hat{\tau}Y_i\right) \tag{4.22}$$

(Beirlant *et al*. 2004, p. 149).

4.4.1.2 Bayes estimation method

A more suitable way of estimating parameters is by using a Baysian approach.

Let $\mathbf{y} = \left(y_1,...,y_m\right)$ denote the observed data of a random variable Y with density function $f\left(y|\boldsymbol{\theta}\right)$, where $\boldsymbol{\theta}$ is a vector of parameters.

The prior distribution for $\boldsymbol{\theta}$ is given by $\pi\left(\boldsymbol{\theta}\right)$ and the likelihood of $\boldsymbol{\theta}$ is given by $f\left(\mathbf{y}|\boldsymbol{\theta}\right)$. According to Bayes' theorem the posterior distribution is given by

$$\pi\left(\boldsymbol{\theta},\mathbf{y}\right) = \frac{f\left(\mathbf{y}|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)}{\int_{\boldsymbol{\Omega}} f\left(\mathbf{y}|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)\partial\boldsymbol{\theta}} \propto f\left(\mathbf{y}|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right) \tag{4.23}$$

where the integral is taken over the parameter space $\boldsymbol{\Omega}$. Calculating the mode or the mean of the posterior is a method of obtaining estimates of $\boldsymbol{\theta}$ (Beirlant *et al*. 2004, p. 430).

When using a Bayesian approach a prior $\pi\left(\boldsymbol{\theta}\right)$ must be specified. With minimal available information an objective prior distribution can be assumed. Uniform priors are the simplest kind of objective priors, other kinds are Jeffreys' priors and maximal data information (MDI) priors. In this section the MDI prior of the GPD is assumed. It is given by the following equation:

$$\pi\left(\sigma,\gamma\right) \propto \frac{1}{\sigma}e^{-\gamma} \tag{4.24}$$

(Beirlant *et al*. 2004, p. 447).

The problem with Bayesian computation is the difficulty of the computation of the integrals in the posterior. These difficulties can be solved with simulation-based techniques called the Markov Chain Monte Carlo (MCMC) methods. Two popular MCMC methods are the Gibbs sampler and the Metropolis-Hastings algorithm (Beirlant *et al.* 2004, p. 430-433). The Metropolis-Hastings algorithm is considered in this study. The Metropolis-Hastings algorithm is now discussed.

4.4.1.2.1 The Metropolis-Hastings algorithm

The following procedure, from Beirlant *et al*. (2004, p. 433), shows how a sequence $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$ is simulated with the Metropolis-Hastings algorithm:

Decide on an initial point $\boldsymbol{\theta}_1$. Before the next state $\boldsymbol{\theta}^{(i+1)}$ is chosen a candidate point $\boldsymbol{\theta}^*$ is sampled from a proposed density $q\left(\boldsymbol{\theta}^* \middle| \boldsymbol{\theta}^{(i)}\right)$ that depends on the current state $\boldsymbol{\theta}^{(i)}$. The candidate $\boldsymbol{\theta}^*$ can be accepted or rejected. The candidate has probability α_i to be accepted where α_i is given by the following equation:

$$\alpha_i = \min\left\{\frac{\pi\left(\boldsymbol{\theta}^* \middle| \mathbf{y}\right) q\left(\boldsymbol{\theta}^{(i)} \middle| \boldsymbol{\theta}^*\right)}{\pi\left(\boldsymbol{\theta}^{(i)} \middle| \mathbf{y}\right) q\left(\boldsymbol{\theta}^* \middle| \boldsymbol{\theta}^{(\mathbf{i})}\right)}, 1\right\}. \qquad (4.25)$$

Given that the candidate is accepted the next state becomes $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^*$. If $\boldsymbol{\theta}^*$ is rejected the chain remains at $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)}$.

The advantage of this algorithm is that it depends on the posterior through a ratio; therefore the posterior is only needed as a proportionality (Beirlant *et al*. 2004, p. 433).

The parameters are estimated as the mean of the simulated sequence when it converges to a certain value.

Table 4.1 gives the Bayesian parameter estimates of the GPD that are obtained when the GPD is fitted to the Fréchet simulated data set above the various thresholds that were chosen with the methods in Section 4.3.

Table 4.1 Parameter estimates of the GPD fitted to the Fréchet simulated data above the thresholds. The Fréchet data set was simulated with $\gamma = 0.5263$

Method	**Threshold**	**Bayes estimators**
Guillou and Hall method	$t = 12{,}1735$	$\hat{\gamma}$ = 0,818 ; $\hat{\sigma}$ = 4,4073
Minimizing MSE method	$t = 2{,}2559$	$\hat{\gamma}$ = 0,5440; $\hat{\sigma}$ = 1,1972
Median of the optimal *k*-values method	$t = 2{,}6520$	$\hat{\gamma}$ = 0,4214; $\hat{\sigma}$ = 1,8995
Dirichlet method	$t = 4{,}4929$	$\hat{\gamma}$ = 0,6264; $\hat{\sigma}$ = 2,1671

Figures 4.9 and 4.10 show the simulated parameter values and the histograms of the posteriors when the GPD is fitted to the Fréchet simulated data set above the respective thresholds. Each row in Figures 4.9 and 4.10 represents the Guillou and Hall, the minimizing of the MSE, the median of the optimal k-values and the Dirichlet method respectively.

Fig. 4.9 **The simulated parameter values of the GPD fitted to the simulated Fréchet data set above the respective thresholds.**

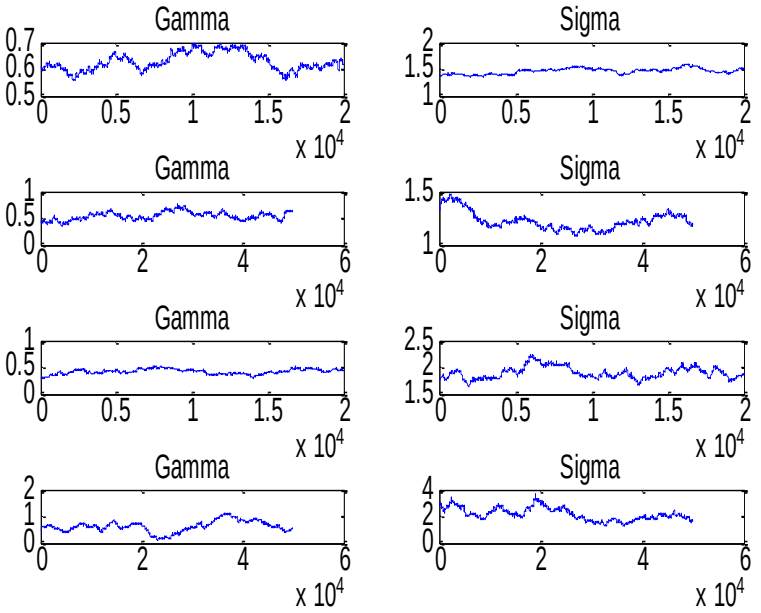


Fig. 4.10 **The histograms of the posterior distributions for the estimated parameters of the GPD fitted to the simulated Fréchet data set above the respective thresholds.**

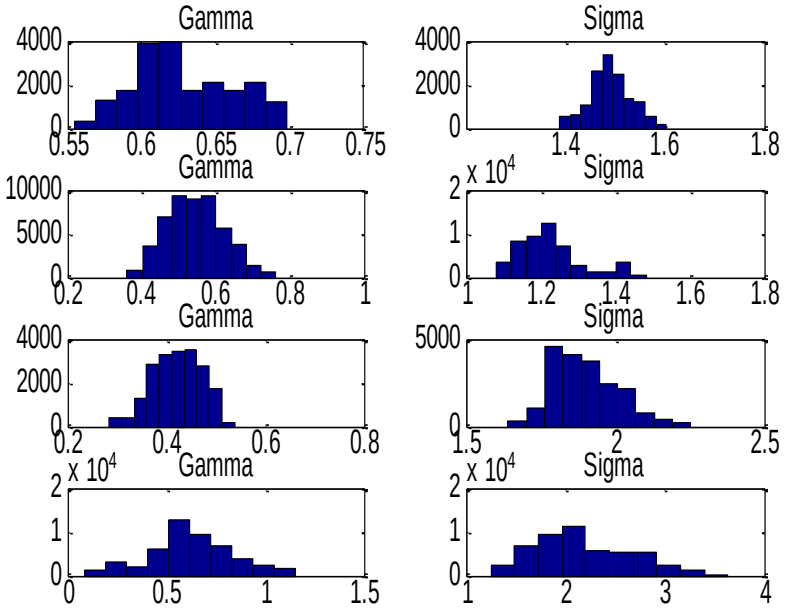


Table 4.2 gives the Bayesian parameter estimators of the GPD that are obtained when the GPD is fitted to the Burr (type III) simulated data set above the various thresholds that were chosen with the different methods in Section 4.3.

Table 4.2 Parameter estimates of the GPD fitted to the Burr (type III) simulated data above the thresholds. The Burr (type III) data set was simulated with $\gamma = 0.8333$

Method	**Threshold**	**Bayes estimators**
Guillou and Hall method	$t = 76,9797$	$\hat{\gamma}$ = 1,1966; $\hat{\sigma}$ = 17,442
Minimizing MSE method	$t = 1,2239$	$\hat{\gamma}$ = 0,9752; $\hat{\sigma}$ = 1,2997
Median of the optimal *k*-values method	$t = 2,8963$	$\hat{\gamma}$ = 1,2357; $\hat{\sigma}$ = 2,4291
Dirichlet method	$t = 42,6215$	$\hat{\gamma}$ = 1,1638; $\hat{\sigma}$ = 35,3059

Figures 4.11 and 4.12 show the simulated parameter values and the histograms of the posteriors when the GPD is fitted to the Burr (type III) simulated data set above the respective thresholds. Each row in Figures 4.11 and 4.12 represents the Guillou and Hall, the minimizing of the MSE, the median of the optimal k-values and the Dirichlet method respectively.

Fig. 4.11 **The simulated parameter values of the GPD fitted to the simulated Burr (type III) data set above the respective thresholds.**

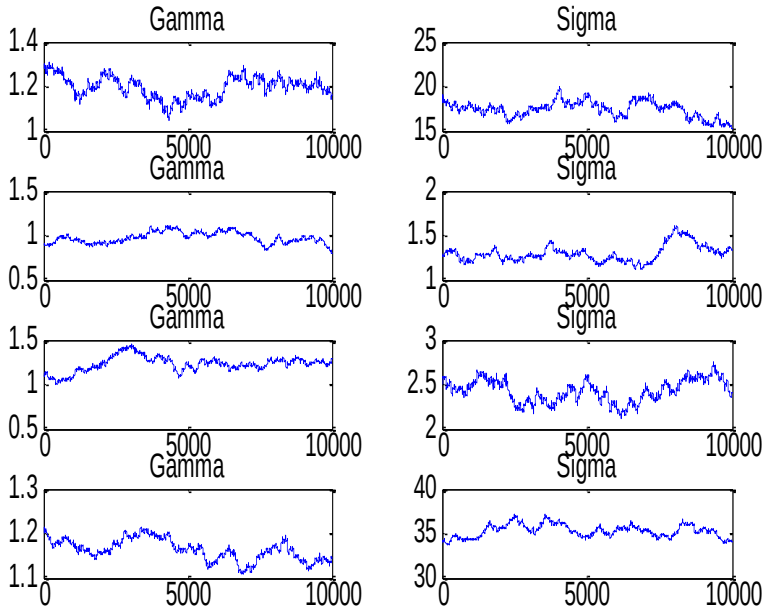
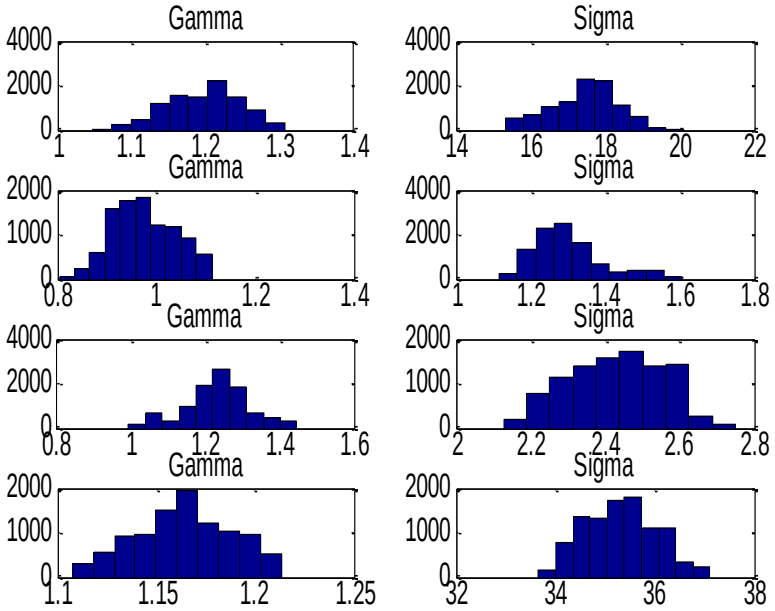


Fig. 4.12 **The histograms of the posterior distributions for the estimated parameters of the GPD fitted to the simulated Burr (type III) data set above the respective thresholds.**



4.4.2 Estimation of large quantiles

After estimating the parameters of the GPD the extreme quantiles of the GPD is estimated. This estimation is done by inverting (1.44) resulting in the following:

$$U\left(\frac{1}{p}\right)=\begin{cases}\frac{\sigma}{\gamma}\left(p^{-\gamma}-1\right), \gamma \neq 0 \\ -\sigma \log p, \gamma = 0\end{cases} \tag{4.26}$$

(Beirlant *et al*. 2004, p.158). (1.43) implies that if the data are not exact GPD and $x = t + y$ then

$$\bar{F}(x) \sim \bar{F}(t)\left(1+\frac{\gamma(x-t)}{\sigma}\right)^{\frac{-1}{\gamma}}. \tag{4.27}$$

$\bar{F}(t)$ is estimated by $\frac{N_t}{n}$ (Beirlant *et al*. 2004, p.158). Substituting the estimated parameter values and $\hat{\bar{F}}(t)$ into (4.27) gives

$$\hat{\bar{F}}(x)=\frac{N_t}{n}\left(1+\frac{\hat{\gamma}(x-t)}{\hat{\sigma}}\right)^{\frac{-1}{\gamma}}. \tag{4.28}$$

Inverting the right-hand side of (4.28) leads to a POT estimator for large quantiles given by the following equation:

$$\hat{Q}(1-p)=t+\frac{\hat{\sigma}}{\hat{\gamma}}\left(\left(\frac{np}{N_t}\right)^{-\hat{\gamma}}-1\right) \tag{4.29}$$

(Beirlant *et al*. 2004, p.158).

4.4.3 Quantile-quantile plots

Four methods were discussed on how to select the optimum threshold. To investigate whether these methods are appropriate one can make use of QQ-plots. A QQ-plot can provide information about the tail of a distribution; therefore a statistical fit is preformed on the data above a threshold and not on the whole data set. The QQ-plot is explained by Beilant *et al*. (2004, p. 5-7). Again the discussion starts with the exponential model, $Exp(\lambda)$, as was done in Section 4.2, where the quantile function of the exponential distribution is given in (4.4).

There exists a simple linear relationship between the quantiles of any exponential distribution and the standard exponential quantiles. This is shown by the following equation:

$$Q_\lambda(p) = \frac{1}{\lambda} Q_1(p) \text{ for } p \in (0,1). \tag{4.30}$$

For a given data set $x_1, ..., x_n$ the unknown population quantile function Q is replaced by empirical approximations $\hat{Q}_n$ where

$$\hat{Q}_n(p) = x_{i,n}, \frac{i-1}{n} \le p \le \frac{i}{n}; i = 1, ..., n. \tag{4.31}$$

The QQ-plots have points

$$\left(-\log(1-p), \hat{Q}_n(p)\right) \tag{4.32}$$

that are plotted for several values of $p \in (0,1)$. If the pattern of the scatter plot follows a straight line the underlying model provides a plausible statistical fit for the given statistical population. In general the intercept for the given model is 0. Often, as is the case in this study, only data above a threshold t is

available. In such a case the exponential model changes to a shifted exponential model because of the conditioning on the event $(X > t)$. Thus, in the case of the shifted exponential model,

$$P(X > x \mid X > t) = \frac{P(X > x)}{P(X > t)} = \exp\left(-\lambda(x - t)\right), x > t \ . \qquad (4.33)$$

Now the quantile function becomes

$$Q(p) = t - \frac{1}{\lambda}\log(1 - p), 0 < p < 1 \qquad (4.34)$$

and the QQ-plot has an intercept t at the value $p = 0$ (Beilant *et al*. 2004, p. 5-7).

Using the quantile function (4.29) the QQ-plot is constructed when the GPD quantiles (4.29) are plotted against the observed quantiles, $\hat{Q}_n(p)$ in (4.31).

Figure 4.13 (a - d) shows the QQ-plots of the quantiles of the GPD (fitted on the Fréchet simulated data above the respective thresholds) with the respective estimated parameter values against the observed quantiles.

Fig. 4.13 (a) **The QQ-plot with t = 12,1735, $\hat{\gamma}$ = 0,818 and $\hat{\sigma}$ = 4,4073**

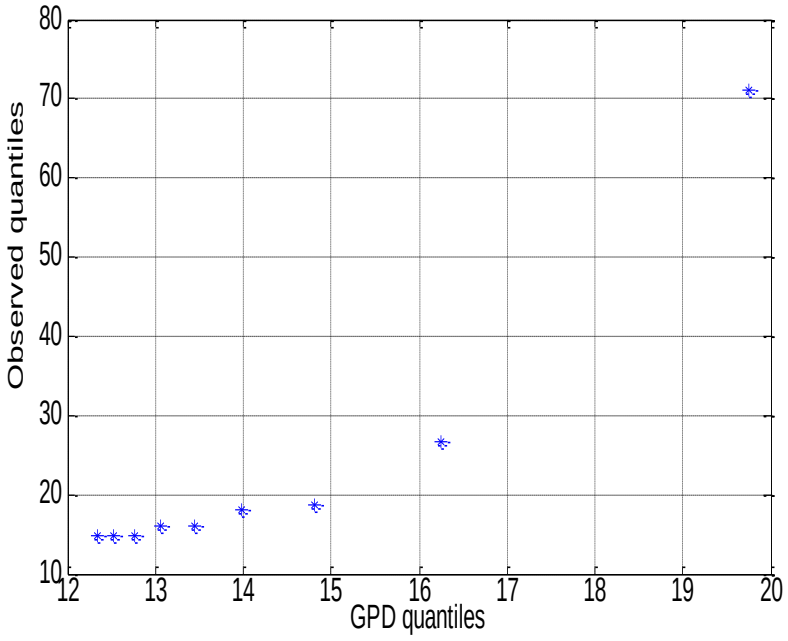


Fig. 4.13 (b) **The QQ-plot with t = 2,2559, $\hat{\gamma}$ = 0,5440 and $\hat{\sigma}$ = 1,1972**

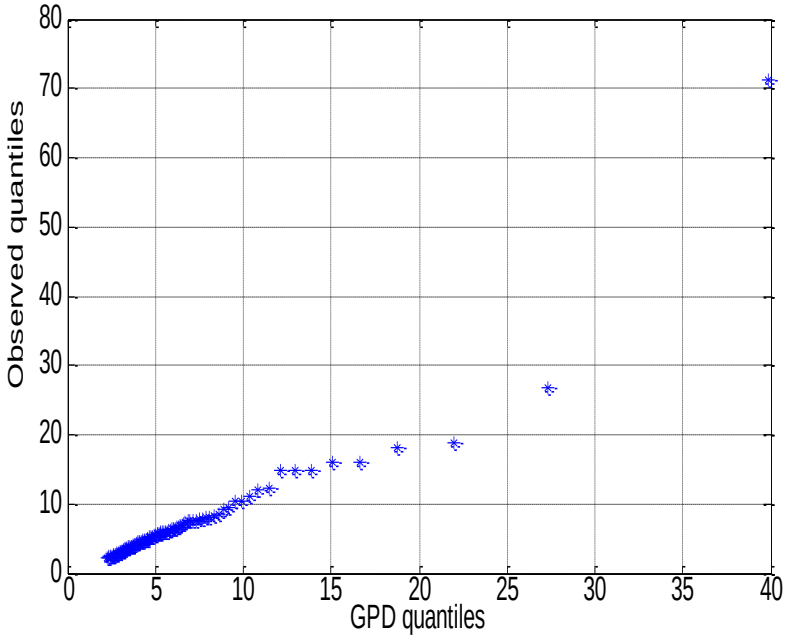


Fig. 4.13 (c) **The QQ-plot with t = 2,6520, $\hat{\gamma}$ = 0,4214 and $\hat{\sigma}$ = 1,8995**

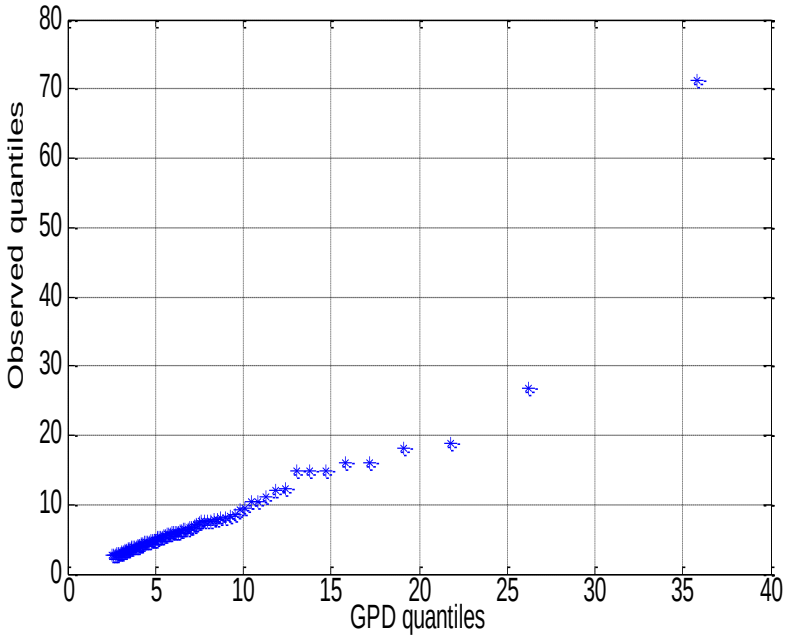


Fig. 4.13 (d) **The QQ-plot with t = 14,8267, $\hat{\gamma}$ = 0,6264 and $\hat{\sigma}$ = 2,1671**

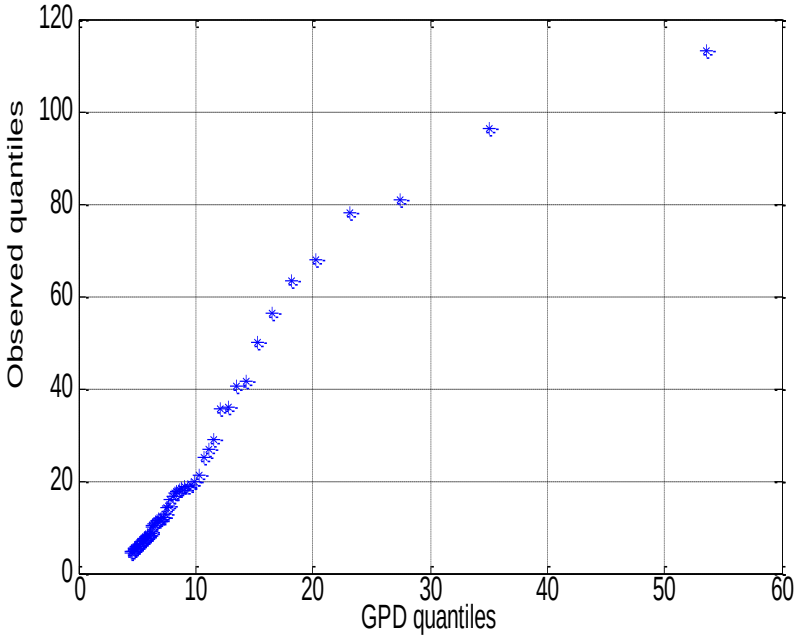


Figure 4.14 (a - d) shows the QQ-plots of the quantiles of the GPD (fitted on the Burr (type III) simulated data above the respective thresholds) with the respective estimated parameter values against the observed quantiles.

Fig. 4.14 (a) **The QQ-plot with t = 76,9797, $\hat{\gamma}$ = 1,1966 and $\hat{\sigma}$ = 17,442**

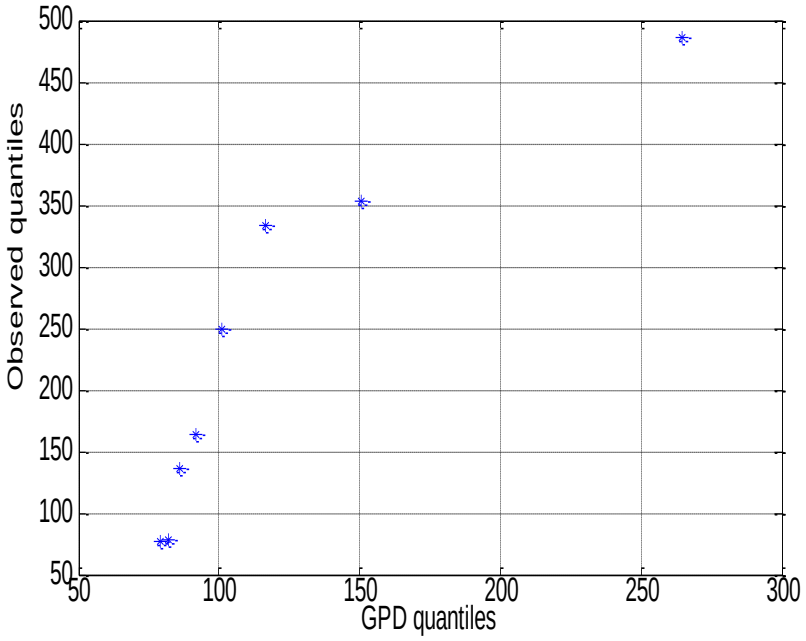


Fig. 4.14 (b) **The QQ- plot with t = 1,2239, $\hat{\gamma}$ = 0,9752 and $\hat{\sigma}$ = 1,2997**

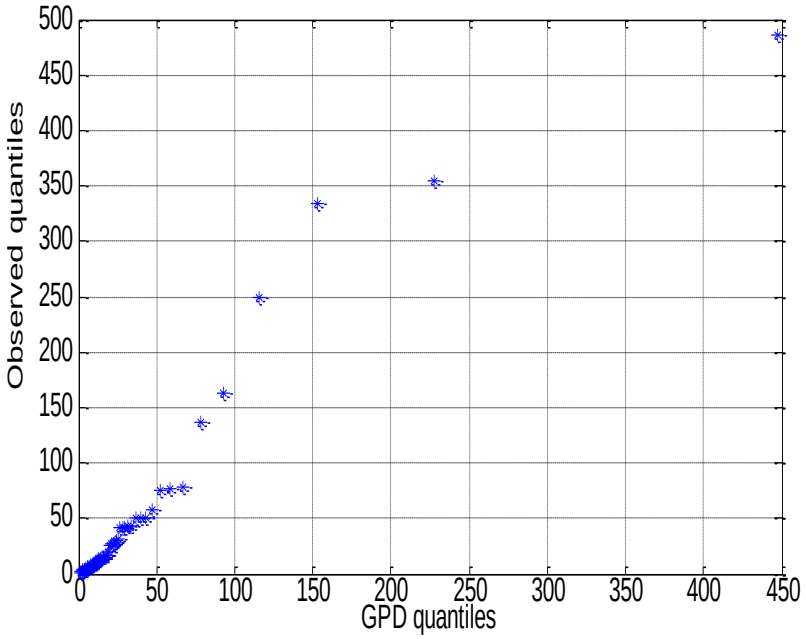


Fig. 4.14 (c) **The QQ-plot with t = 1,2357, $\hat{\gamma}$ = 1,2357 and $\hat{\sigma}$ = 2,4291**

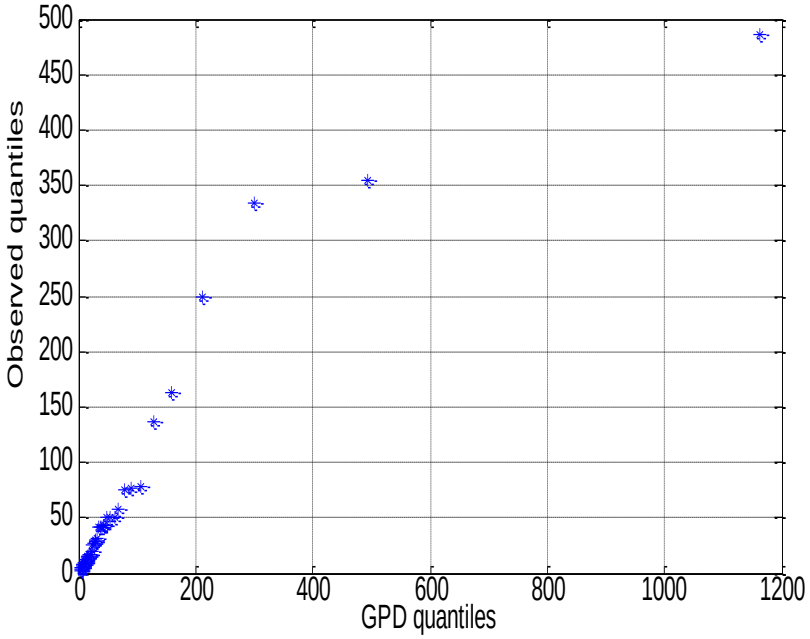
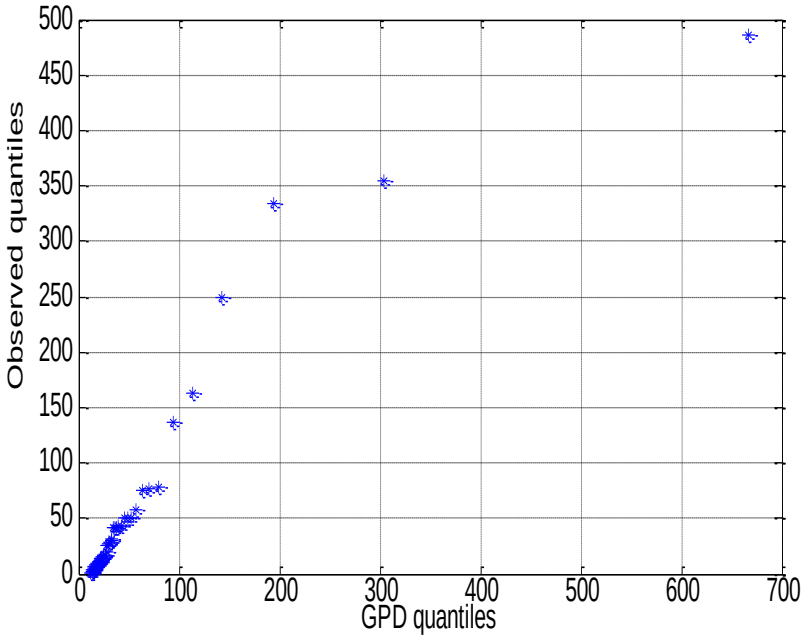


Fig. 4.14 (d) **The QQ-plot with t = 42.6215, $\hat{\gamma}$ = 1,1638 and $\hat{\sigma}$ = 35,3059**



4.5 A simulation study of the Dirichlet process

In this section the focus is specifically on the behaviour of the Dirichlet process. The behaviour of the Dirichlet process is investigated by using a simulation study. Fifty simulations are made from a Burr (type III) distribution with $\tau = 0,9$, $\eta = 4,2$ and $\lambda = 1,5$. For each simulation a threshold is chosen by using the NDE of the Dirichlet process (discussed in Section 4.3.4). After selecting a threshold, the Strict Pareto distribution is fitted to the data above the threshold. The Strict Pareto distribution is easier to work with than the GPD because there is only one parameter γ (the EVI). The survival function of the Strict Pareto is

$$\bar{F}(x) = x^{\frac{-1}{\gamma}}, (x > 1) \qquad (4.35)$$

(Beirlant *et al*. 2004, p. 59) and the log-likelihood is given by

$$\log L\left(Y_1,...,Y_{N_t}\right) = -N_t \log\gamma - \left(1+\frac{1}{\gamma}\right)\sum_{j=1}^{N_t}\log Y_j \qquad (4.36)$$

where $Y_j = \frac{X_i}{t}$ are the relative excesses over the threshold *t*, on condition that $X_i > t$, $j = 1,...,N_t$ (Beirlant *et al*. 2004, p.103). The MDI prior of the Strict Pareto distribution is

$$\pi(\gamma) = \left(\frac{1}{\gamma}\right)\exp(-\gamma) \qquad (4.37)$$

(Beirlant *et al*. 2004, p. 438).

The quantile function is given by

$$Q(p) = (1-p)^{-\gamma}. \qquad (4.38)$$

Taking logs on both sides, the quantile function becomes

$$\log Q(p) = -\gamma \log(1-p). \qquad (4.39)$$

The Pareto quantile plot is given by

$$\left(-\log\left(1-p_{i,n}\right), \log x_{i,n}\right), i = 1,..., p, \qquad (4.40)$$

which is the exponential QQ-plot after taking the log of the data (Beirlant *et al*. 2004, p.61).

The parameter γ of the Strict Pareto is estimated for each simulation; an estimate of γ is the mean or the mode of the posterior distribution. After estimating the parameter the QQ-plot is drawn to observe the goodness of fit.

Figure 4.15 shows the estimated parameter values (blue line) for each simulation when the mode of the posterior is obtained as an estimated value of the parameter for each simulation. The red line shows the estimated parameter values when the mean is used as an estimator of the parameter for each simulation. The actual value of $\gamma = \frac{1}{\tau} = 1{,}11$, this value is indicated with a green line in the figure.

Fig. 4.15 The estimated parameter values for each simulation

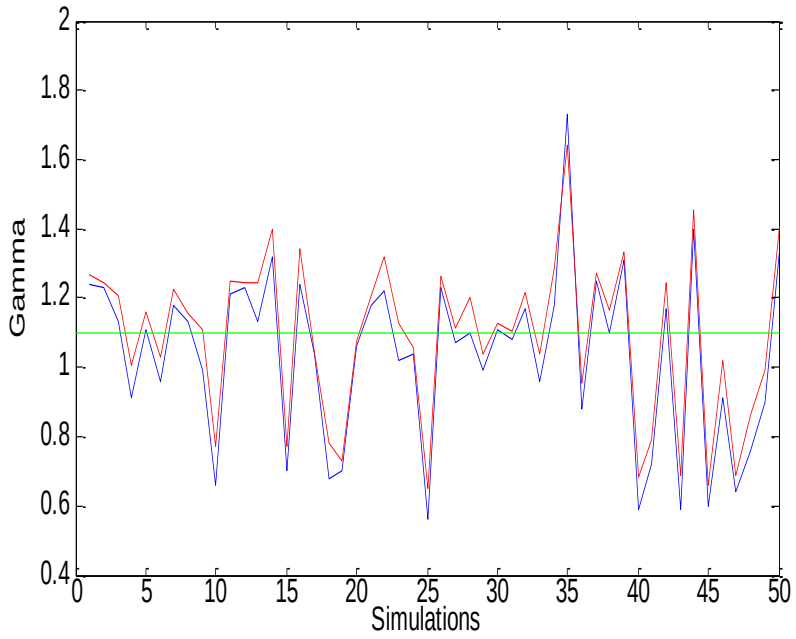
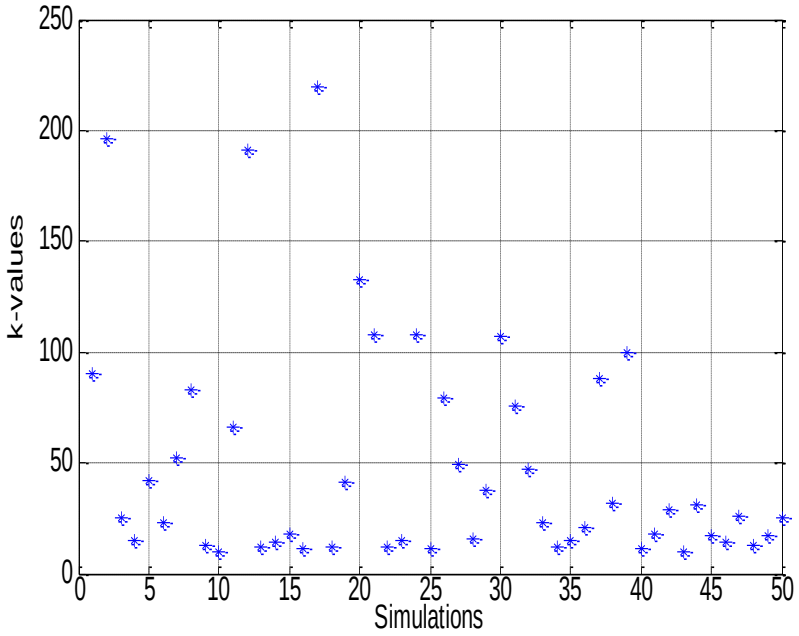


Figure 4.15 shows that the blue line and the red line is very similar and rather close to the actual value (the green line), therefore the mean or the mode of the posterior can be used to obtain an estimate for the parameter. For this study the mean of the posterior is used as an estimate.

Figure 4.16 shows the values of *k* for each simulation that is chosen by the NDE of the Dirichlet process.

Fig.4.16 The selected *k*-values for each simulation



Next a QQ-plot for all 50 simulations is drawn on the same graph, shown in Figure 4.17.

Fig. 4.17 The quantile-quantile plots for all the simulations

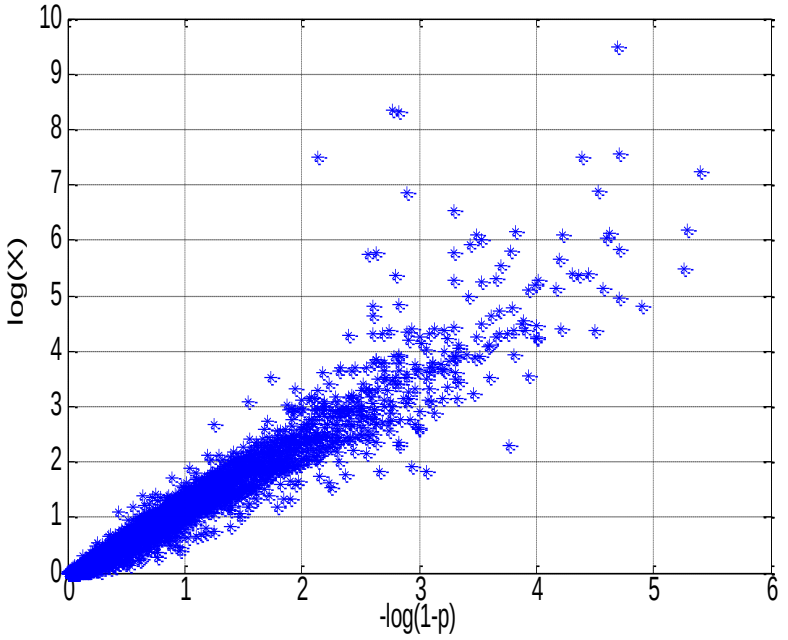


Figure 4.17 shows that, although the QQ-plots vary due to different simulations, the various QQ-plots lie more or less in the same area. This indicates that the Dirichlet process is a proper and consistent method to use when selecting thresholds.

4.6 Conclusion

From Figure 4.13 (b) and 4.13 (c) the GPD seems to be a good fit to the Fréchet simulated data above the thresholds $t = 2,2559$ and $t = 2,6520$. This is because the QQ-plots show a linear pattern. Therefore, the method of minimizing the MSE and the method of the median of the optimal *k*-values seem to be efficient in this case.

From Figure 4.14 (c) the GPD seems to be a good fit to the Burr (type III) simulated data above the threshold $t = 1,7311$. This is because the QQ-plot shows a linear pattern. Therefore, the method of the median of the optimal *k*-values seems to be efficient in this case. Figure 4.14 (d) is also close to a linear pattern, indicating that the GPD seems to be a reasonably good fit to the Burr (type III) simulated data above the threshold $t = 40,4139$. Therefore, the Dirichlet process is also efficient in this case.

The simulation study in Section 4.5 shows that the Dirichlet method is an appropriate method to use when choosing a threshold.

Chapter 5

Copulas and Dirichlet mixture models

In the first part of this chapter extreme values are modelled through extreme value copulas and in the second part of this chapter extreme values are modelled through a Dirichlet mixture model.

5.1 Introduction of copulas

Copula is a Latin word that refers to connecting or joining together. In this study copula refers to statistical concepts: the way random variables relate to each other. Copula functions give information on how univariate marginal distributions "come together" to determine a multivariate distribution (Dowd 2007).

Usually bivariate or multivariate models are constructed from known marginal distributions and dependence between random variables. There are various ways of measuring this dependence. In classical theory the correlation coefficient, Kendall's tau, Spearman's rho *etc*. are used as a dependence measure. These measures have limitations and give very little information specifically about the dependence in the tails. In extreme value theory dependencies between random variables in the tails are of most interest. For instance, if one extreme event will occur, what is the probability that another extreme event will also occur? Therefore we turn to copulas; copulas are a reliable measure of dependencies between variables. Copulas separate the dependence structure from the marginal distribution functions. Copulas were first introduced by Sklar (1959)**.** At first copulas were mainly used in the theory of probabilistic metric spaces; only later becoming popular in mathematical statistics as nonparametric measures of dependence between random variables.

In Section 5.2 the basic properties of copulas are shown. Section 5.3 discusses a specific family of copulas called the extreme value copulas. In Section 5.4 the fitting of copula models, with reference to the Gumbel and the Tawn copula, are discussed. Section 5.4 also includes a discussion on parameter estimation. A practical application is done in Section 5.5. And Section 5.6 discusses the calculating of tail probabilities.

5.2 Basic properties of copulas

The basic properties of copulas are discussed in this section. This discussion is based on the book "Operational Risk Modeling Analytics" by Panjer (2006). For further reading on copulas the reader is referred to "Introduction to Copulas" by Nelsen (2006).

Let a *d*-variate copula represent the joint distribution function of *d* Uniform (0,1) random variables. Thus, if the *d* random variables are given by $U_1,...,U_d$, then the copula *C* is as follows:

$$C(u_1,...,u_d) = \Pr(U_1 \le u_1,...,U_d \le u_d) \qquad (5.1)$$

(Panjér 2006, p.234). Therefore, copulas are the joint distribution functions of Uniform (0,1) random variables.

For continuous random variables $X_1,...,X_d$, having distribution functions $F_1,..,F_d$ respectively and a joint distribution function *F*, $F_1(x_1), F_2(x)_2,...,F_d(x_d)$ are uniform (0,1) random variables. This is due to the probability integral from basic probability theory. Therefore, a copula evaluated at $F_1(x_1), F_2(x)_2,...,F_d(x_d)$ is given as

$$C(F_1(x_1),...,F_d(x_d)) = \Pr(U_1 \le F_1(x_1),...,U_d \le F_d(x_d)) \qquad (5.2)$$

(Panjér 2006, p.234). The quantile function is defined as

$$F_j^{-1}(u) = \inf\left\{x : F_j(x) \geq u\right\},\ j = 1,...,d \tag{5.3}$$

(Panjér 2006, p.235). Now the copula in (5.1) is rewritten as

$$\begin{aligned} C\left(F_1(x_1),...,F_d(x_d)\right) &= \Pr\left(F_1^{-1}(U_1) \leq x_1,...,F_d^{-1}(U_d) \leq x_d\right) \\ &= \Pr\left(X_1 \leq x_1,...,X_d \leq x_d\right) \\ &= F\left(x_1,...,x_d\right), \end{aligned} \tag{5.4}$$

(5.4) is Sklar's theorem in a less mathematical sense (Panjér 2006, p.235).

Sklar's theorem states that there is a unique copula *C* for any joint distribution function *F* that satisfies the following:

$$F\left(x_1,...,x_d\right) = C\left(F_1(x_1),...,F_d(x_d)\right). \tag{5.5}$$

On the contrary, for any copula *C* and any distribution functions $F_1(x_1), F_2(x_2),..., F_d(x_d)$, $C\left(F_1(x_1), F_2(x_2),..., F_d(x_d)\right)$ is a joint distribution function with $F_1(x_1),..., F_d(x_d)$ as marginals (Panjér 2006, p.235). Hence, Sklar's theorem enables us to separate the dependence structure from the marginal distributions and enables us to construct joint distribution functions from marginal functions which take the dependence structure into account (Panjér 2006, p.235).

Certain basic probability arguments can be followed in the bivariate case. This is shown in the following equations, which lead to the Fréchet lower and upper bound:

$$\Pr\left(U_i > u_i, U_j > u_j\right) = 1 - u_i - u_j + C\left(u_i, u_j\right) \tag{5.6}$$

therefore

$$\begin{aligned} C\left(u_i,u_j\right) &= u_i+u_j-1+\Pr\left(U_i>u_i,U_j>u_j\right) \\ &\geq u_i+u_j-1 \end{aligned} \tag{5.7}$$

(Panjér 2006, p.236). From (5.7) and because u_i+u_j-1 can be negative, the Fréchet lower bound, on the copula cumulative distribution function, is

$$C\left(u_i,u_j\right)\geq \max\left\{0,u_i+u_j-1\right\}. \tag{5.8}$$

The Fréchet upper bound is obtained as follows:

$$\Pr\left(U_i\leq u_i,U_j\leq u_j\right)\leq \Pr\left(U_i\leq u_i\right)=u_i \tag{5.9}$$

and

$$\Pr\left(U_i\leq u_i,U_j\leq u_j\right)\leq \Pr\left(U_j\leq u_j\right)=u_j \tag{5.10}$$

therefore

$$\Pr\left(U_i\leq u_i,U_j\leq u_j\right)\leq \min\left\{u_i,u_j\right\} \tag{5.11}$$

(Panjér 2006, p.236-237).

As mentioned before, extreme value theory is especially interested in the tails of the distributions. An understanding the joint behaviour in extreme events is of great importance. There are certain developed measures that give the tail dependence and investigate the correlation in the upper tails.

The index of the upper tail dependence of two continuous random variables with marginal distributions $F\left(x\right)$ and $G\left(x\right)$ is given by the following equation:

$$\lambda_u = \lim_{u \to 1} \Pr\left\{ X > F^{-1}(u) \middle| Y > G^{-1}(u) \right\} \tag{5.12}$$

(Panjér 2006, p.239). This index explains more or less the probability that *X* is very large if it is known that *Y* is very large.

(5.12) is now rewritten as

$$\begin{aligned} \lambda_u &= \lim_{u \to 1} \Pr\left\{ F(X) > u \middle| G(Y) > u \right\} \\ &= \lim_{u \to 1} \Pr\left\{ U > u \middle| V > u \right\} \end{aligned} \tag{5.13}$$

where *U* and *V* are random variables (Panjér 2006, p.239).

Now

$$\begin{aligned} \lambda_u &= \lim_{u \to 1} \frac{1 - \Pr\{U \le u\} - \Pr\{V \le u\} + \Pr\{U \le u, V \le u\}}{1 - \Pr\{V \le u\}} \\ &= \lim_{u \to 1} \frac{1 - 2u + C(u,u)}{1 - u} \end{aligned} \tag{5.14}$$

(Panjér 2006, p. 240). (5.14) show that tail dependencies should rather be measured with copulas than with the original distributions.

There are various well known types and families of copulas. One such type is called the extreme value copula, which is discussed in the following section.

5.3 An overview of extreme value copulas

One particular class of copulas is called the extreme-value copulas. This class of copulas are the possible limits of copulas of component-wise maxima of i.i.d. samples. A copula is called an extreme value copula if it satisfies the following:

$$C\left(u_1^n,...,u_d^n\right)=C^n\left(u_1,...,u_d\right) \tag{5.15}$$

for all $\left(u_1,...,u_d\right)$ where *n > 0* (Panjer 2006, p. 257).

This is due to the fact that the extreme value copula has a max-stable property, explained by (Panjer 2006, p. 258) as follows:

In a bivariate case where $\left(X_1,Y_1\right),...,\left(X_n,Y_n\right)$ are i.i.d. random variables with joint distribution $F\left(x,y\right)$, marginal distributions $F_X\left(x\right)$ and $F_Y\left(y\right)$ and copula function $C\left(x,y\right)$, $M_X=\max\left(X_1,..,X_n\right)$ and $M_Y=\max\left(Y_1,..,Y_n\right)$ are called component-wise maximums.

The distribution function of $\left(M_X,M_Y\right)$ is given by

$$\begin{aligned}\Pr\left(M_X\le x,M_Y\le y\right)&=\Pr\left(X_i\le x,Y_i\le y\right),\text{for all } i\\&=F^n\left(x,y\right)\end{aligned} \tag{5.16}$$

and the marginals are $F_X^n\left(x\right)$ and $F_Y^n\left(y\right)$ respectively.

From (5.5) it is known that

$$F\left(x,y\right)=C\left(F_X\left(x\right),F_Y\left(y\right)\right), \tag{5.17}$$

therefore the joint distribution function is

$$\begin{aligned}F^n\left(x,y\right)&=C^n\left(F_X\left(x\right),F_Y\left(y\right)\right)\\&=C^n\left(\left[F_X^n\left(x\right)\right]^{\frac{1}{n}},\left[F_Y^n\left(y\right)\right]^{\frac{1}{n}}\right).\end{aligned} \tag{5.18}$$

From (5.18) the copula of maxima is given by

$$C_{\max}\left(u_1,u_2\right)=C^n\left(u_1^{\frac{1}{n}},u_2^{\frac{1}{n}}\right) \tag{5.19}$$

and this is equivalent to

$$C_{\max}\left(u_1^n,u_2^n\right)=C^n\left(u_1,u_2\right). \tag{5.20}$$

Because the maxima copula, $C_{\max}$, has the same form as the original copula *C*, it is said that the copula has a max-stable property (Panjer 2006, p. 258).

The following work was introduced by Pickands (1981). He showed that

$$C\left(u_1,u_2\right)=\exp\left\{\ln\left(u_1u_2\right)A\left(\frac{\ln u_1}{\ln\left(u_1u_2\right)}\right)\right\} \tag{5.21}$$

has a max-stable property, where *A* is known as the Pickands dependence function. It satisfies

$$A(\omega)=\int_0^1\max\left[x(1-\omega),\omega(1-x)\right]dH\left(x\right) \tag{5.22}$$

where $\omega\in[0,1]$ and *H* is a distribution function on [0,1] (Panjer 2006, p. 258).

$A(\omega)$ must also be a convex function that satisfies the following inequality:

$$\max\left(\omega,1-\omega\right)\le A(\omega)\le 1,0<\omega<1. \tag{5.23}$$

Therefore a copula can be constructed from any differentiable convex function $A(\omega)$ that satisfies (5.23) (Panjer 2006, p. 258).

.

The upper tail dependence index can now be rewritten in terms of $A(\omega)$ as

$$\begin{aligned}\lambda_u &= \lim_{u\to 1}\frac{1-2u+C(u,u)}{1-u}\\ &= \lim_{u\to 1}\frac{1-2u+u^{2A\left(\frac{1}{2}\right)}}{1-u}\\ &= \lim_{u\to 1} 2-2A\left(\frac{1}{2}\right)u^{2A\left(\frac{1}{2}\right)-1}\\ &= 2-2A\left(\frac{1}{2}\right)\end{aligned} \tag{5.24}$$

(Panjer 2006, p. 259).

Another dependence measure is given in Tawn (1988). Here the dependence between (U_1, U_2) is measured by

$$\rho = \int_0^1 \frac{\partial\omega}{\{A(\omega)\}^2} - 1. \tag{5.25}$$

This is always nonnegative because $A(\omega) \le 1$.

For any further statistical inference on the bivariate extreme-value copula *C,* one can concentrate on the inference on its dependence function *A*. There are mainly two approaches to inference: the first is to assume a parametric model for *A* and to estimate the unknown parameters, while the second is to construct non-parametric estimators of *A* in the full model.

When considering the second approach there are mainly two families of estimators: (i) Pickands (1981) estimator and variations and (ii) Capéraà *et al*. (1997) estimator and variations (Segers 2004, p. 2).

The non-parametric estimator that was given by Pickands (1981) is explained as follows:

If $(X_i, Y_i), 1 \le i \le n$ are i.i.d. pairs from a bivariate extreme value distribution with exponential margins and Z_i is defined as follows:

$$Z_i(\omega) = \min\left\{(1-\omega)^{-1} X_i, \omega^{-1} Y_i\right\}, 1 \le i \le n, 0 \le \omega \le 1 \tag{5.26}$$

then

$$\Pr\left\{Z_i(\omega) > Z\right\} = \exp\left\{-zA(\omega)\right\} \tag{5.27}$$

and the maximum likelihood estimator is

$$\hat{A}(\omega) = n\left\{\sum_{i=1}^{n} Z_i(\omega)\right\}^{-1} \tag{5.28}$$

(Tawn 1988, p. 400-401).

When the dependence structure is modelled via parametric models the models can either be differentiable or non-differentiable. This study only focuses on differentiable models for *A*. Two classes of copulas are considered in the next section: the Gumbel copula and the Tawn copula.

5.4 Fitting copula models

5.4.1 The Gumbel copula

The Gumbel copula is given by the following equation:

$$C(u_1, ..., u_d) = \exp\left\{-\left[(-\ln u_1)^{\theta} + ... + (-\ln u_d)^{\theta}\right]\right\}^{\frac{1}{\theta}} \tag{5.29}$$

(Panjer 2006, p. 244).

The dependence function is given as

$$A(\omega)=\left[\omega^{\theta}+(1-\omega)^{\theta}\right]^{\frac{1}{\theta}},\theta\geq 0. \tag{5.30}$$

When $\theta=\frac{1}{2}$ the index of the upper tail dependence is equal to $2-2^{\frac{1}{\theta}}$ (Panjer 2006, p. 244 & 259).

5.4.2 The Tawn copula

The Tawn copula is the extended version of the Gumbel copula with three parameters instead of one. The dependence function is

$$A(\omega)=(1-\alpha)\omega+(1-\beta)(1-\omega)+\left\{(\alpha\omega)^{\theta}+\left[\beta(1-\omega)\right]^{\frac{1}{\theta}}\right\},0\leq\alpha,\beta\leq 1 \tag{5.31}$$

(Panjer 2006, p. 261) and the copula function is given as

$$C(u_1,u_2)=u_1^{1-\alpha}u_2^{1-\beta}\exp\left\{-\left[(-\alpha\ln u_1)^{\theta}+(-\beta\ln u_2)^{\theta}\right]^{\frac{1}{\theta}}\right\} \tag{5.32}$$

(Panjer 2006, p. 261).

5.4.3 Parameter estimation

In the following sections the maximum likelihood and the Bayesian approach are discussed as methods for estimating the parameters.

5.4.3.1 Maximum likelihood estimation

This follows from Panjer (2006, p. 396-397). Given that a *d*-variate random variable $(X_1,X_2,...,X_d)$, with continuous marginal distributions, has

$f_1(x_1), f_2(x_2),..., f_d(x_d)$ as the respective probability density functions (pdfs) and $f(x_1, x_2,..., x_d)$ as a continuous multivariate joint density. The joint cumulative distribution function (cdf) is then

$$F(x_1,..., x_d) = C\left(F_1(x_1),..., F_d(x_d)\right) \tag{5.33}$$

where $C(u_1,...,u_d)$ is the cdf of the copula at the point $(u_1,...,u_d)$ (Panjer 2006, p. 396). After differentiating (5.33) the joint pdf is given by the equation

$$f(x_1,..., x_d) = f_1(x_1)\cdot f_2(x_2)\cdot...\cdot f_d(x_d)\cdot c\left(F_1(x_1),..., F_d(x_d)\right) \tag{5.34}$$

where $c(u_1,...,u_d)$ is the pdf of the copula at point $(u_1,...,u_d)$ (Panjer 2006, p. 396).

The maximum likelihood estimates of the copula parameters are unstable due to additional uncertainty caused by the estimation of the marginal distributions' parameters.

Taking the log of (5.34) gives the following:

$$\log f(x_1,..., x_d) = \sum_{i=1}^{d} \log f_i(x_i) + \log c\left(F_1(x_1),..., F_d(x_d)\right) \tag{5.35}$$

(Panjer 2006, p. 396).

For a given sample of *n* i.i.d. observations, $x_{i,j}$ will represent the i^{th} dimension of the j^{th} outcome. The log-likelihood function is given by the following equation:

$$l = \sum_{j=1}^{n} \log f\left(x_{1,j},...,x_{d,j}\right)$$
$$= \sum_{j=1}^{n}\sum_{i=1}^{d} \log f_i\left(x_{i,j}\right) + \sum_{j=1}^{n} \log c\left(F_1\left(x_{1,j}\right),...,F_d\left(x_{d,j}\right)\right) \quad (5.36)$$

(Panjer 2006, p. 397). The maximum likelihood estimates are the values of the parameters that maximize (5.36).

First the parameter estimates for the marginal distributions are obtained. The log-likelihood function of the i^{th} marginal distribution is given by the equation

$$l_i = \sum_{j=1}^{n} \log f_i\left(x_{i,j}\right), i = 1,...,d \quad (5.37)$$

(Panjer 2006, p. 397). All the parameters are estimated for the *d* different marginals. These maximum likelihood estimates (MLEs) are called pseudo-MLEs. These estimates are not very efficient because the parameter(s) of the copula function must still be estimated.

Pseudo-MLEs can also be obtained for the copula parameter(s). Assume $\tilde{u}_{i,j} = \tilde{F}_i\left(x_{i,j}\right)$ are the pseudo-estimate(s) (at each observed value) of the cdf of the marginal distributions.

$$\tilde{l}_c = \sum_{j=1}^{n} \log c\left(\tilde{u}_{1,j},...,\tilde{u}_{d,j}\right) \quad (5.38)$$

is then the pseudo-estimate(s) for the parameter(s) of the copula (Panjer 2006, p. 397). (5.38) is maximized to obtain the pseudo-estimate(s) for the parameter(s) of the copula.

To estimate the true MLE of all the parameters (5.36) is maximized. The pseudo-MLEs are used as starting values for the maximization.

5.4.3.2 Bayesian estimation

The previous section described the method of maximum likelihood to estimate parameters, in this section the estimation of parameters are discussed by using a Bayesian approach. Thus, in this section data are modelled by using copulas in a Bayesian context.

Two Bayesian approaches are considered. The first approach involves estimating the parameters of the marginal distributions and the copula separately and the second approach is to estimate the parameters jointly.

5.4.3.2.1 Separate estimation approach

This follows from Romeo *et al*. (2006, p. 210). The choice of marginal distributions does not depend on the choice of the copula, therefore the estimation of the marginals' parameters and the copula's dependence parameter(s) can be done separately.

This approach assumes independence and each marginal's parameter(s) are estimated separately. The marginals are assumed to be known and $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the vectors of parameters of each marginal distribution respectively. A prior, likelihood and posterior distribution are considered for each vector of parameters. The prior distribution is given as

$$\pi\left(\boldsymbol{\theta}_j\right), j=1,2, \tag{5.39}$$

the likelihood function is given as

$$L\left(\boldsymbol{\theta}\middle|\mathbf{x}_j\right)=\prod_{i=1}^{n} f_{j\theta_j}\left(x_{ij}\right), j=1,2 \tag{5.40}$$

and the posterior distribution is given as

$$\pi\left(\boldsymbol{\theta}_j \middle| \mathbf{x}_j\right) \propto L\left(\boldsymbol{\theta} \middle| \mathbf{x}_j\right)\pi\left(\boldsymbol{\theta}_j\right). \tag{5.41}$$

If $\tilde{\boldsymbol{\theta}}_j$ is the estimate of $\boldsymbol{\theta}_j$ in the posterior distribution (5.41) then the dependence parameter (α) can be estimated from the following posterior distribution:

$$\pi\left(\alpha \middle| \mathbf{x}_1, \mathbf{x}_2, \tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2\right) \propto L\left(\alpha \middle| \mathbf{x}_1, \mathbf{x}_2, \tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2\right)\pi(\alpha) \tag{5.42}$$

(Romeo *et al*. 2006, p. 210).

The likelihood in this case is called a pseudo likelihood and is given by

$$L\left(\alpha \middle| \mathbf{x}_1, \mathbf{x}_2, \tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2\right) \propto \prod_{i=1}^{n} f\left(x_{i1}, x_{i2}; \alpha\right) \tag{5.43}$$

where $f\left(x_{i1}, x_{i2}\right)$ is the bivariate density function of the copula (Romeo *et al*. 2006, p. 210).

5.4.3.2.2 Joint estimation approach

This follows from Romeo *et al*. (2006, p. 211). This approach suggests estimating the marginal parameters and the dependence parameter(s) jointly.

The joint posterior distribution is given by the following equation:

$$\pi\left(\alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \middle| \mathbf{x}_1, \mathbf{x}_2\right) \propto L\left(\alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \middle| \mathbf{x}_1, \mathbf{x}_2\right)\pi\left(\alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right) \tag{5.44}$$

where the likelihood function is

$$L\left(\alpha, \tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2 \middle| \mathbf{x}_1, \mathbf{x}_2\right) = \prod_{i=1}^{n} f_1\left(x_{i1}\right) f_2\left(x_{i2}\right) c\left(F_1\left(x_{i1}\right), F_2\left(x_{i2}\right)\right) \tag{5.45}$$

and the prior distribution is

$$\pi\left(\alpha,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\right) \tag{5.46}$$

(Romeo *et al*. 2006, p. 211).

5.5 Practical application

As a practical example the Gariep Dam data set, from previous chapters, is considered. $Y_1,...,Y_m$ denote the maximum monthly inflow values for January to December of each year. First the Gumbel copula is fitted to the data and then the Tawn copula. When fitting each copula the separate estimation approach is first considered and secondly, the joint estimation approach is considered.

5.5.1 Fitting the Gumbel copula

5.5.1.1 Applying the separate estimation approach

When applying the separate estimation approach the marginal distribution of each month is assumed to be Generalized Extreme Value (GEV) distributed. The distribution function of the GEV distribution is given by the following equation:

$$G\left(y_i,\sigma,\gamma,\mu\right)=\exp\left(-\left(1+\gamma\left(\frac{y_i-\mu}{\sigma}\right)\right)^{\frac{-1}{\gamma}}\right),1+\gamma\left(\frac{y_i-\mu}{\sigma}\right)>0,\gamma\neq 0,\sigma>0,i=1,...,m \tag{5.47}$$

(Beirlant *et al*. 2004, p. 132).

To estimate the parameters of each marginal distribution the Bayesian approach is considered.

For $\gamma \neq 0$ the log-likelihood of a sample of $Y_1,...,Y_m$ i.i.d GEV random variables is given by the following equation:

$$\log L(\sigma,\gamma,\mu) = -m\log\sigma - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{m}\log\left(1+\gamma\frac{Y_i-\mu}{\sigma}\right) - \sum_{i=1}^{m}\left(1+\gamma\frac{Y_i-\mu}{\sigma}\right)^{\frac{-1}{\gamma}},$$
$$1+\gamma\frac{Y_i-\mu}{\sigma} > 0, i = 1,...,m \tag{5.48}$$

(Beirlant *et al*. 2004, p. 132).

The MDI prior is given as

$$\pi(\sigma,\gamma,\mu) = \exp E\left\{\log f\left(Y|\sigma,\gamma,\mu\right)\right\} \tag{5.49}$$

(Beirlant *et al*. 2004, p. 435).

And the log of the prior is

$$\log\pi(\sigma,\gamma,\mu) \propto -\log\sigma - 1 - \left(1+\frac{1}{\gamma}\right)\psi(1), \tag{5.50}$$

where $\psi(1)$ = 0,5772 is Euler's constant.

The log of the posterior is then

$$\log\pi\left(\sigma,\gamma,\mu|y_i\right) = \log L(\sigma,\gamma,\mu) + \log\pi(\sigma,\gamma,\mu). \tag{5.51}$$

The accuracy of the parameters can be described through the highest posterior density (HPD) intervals (Gamerman & Lopes 2006, p. 48).

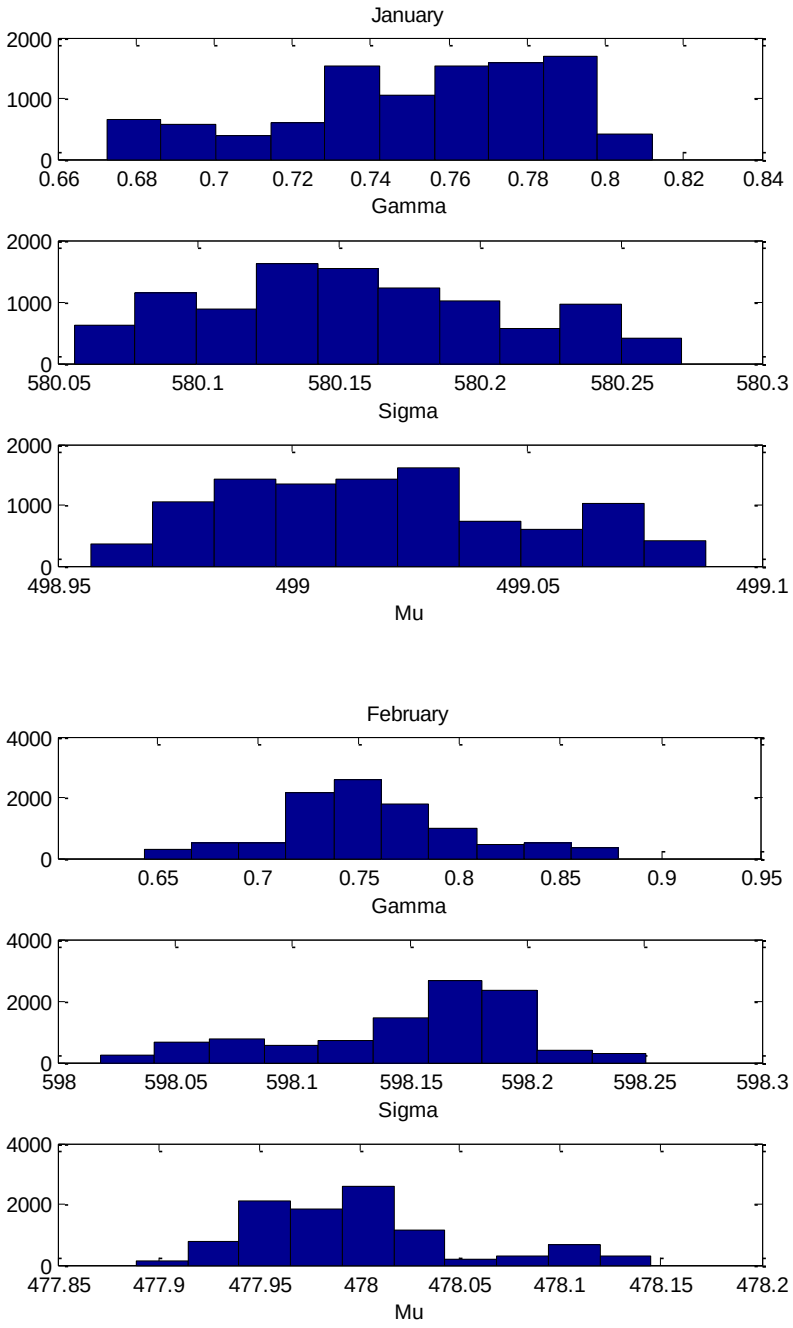
Table 5.1 shows the estimated parameter values for each month and the respective 90% HPD intervals.

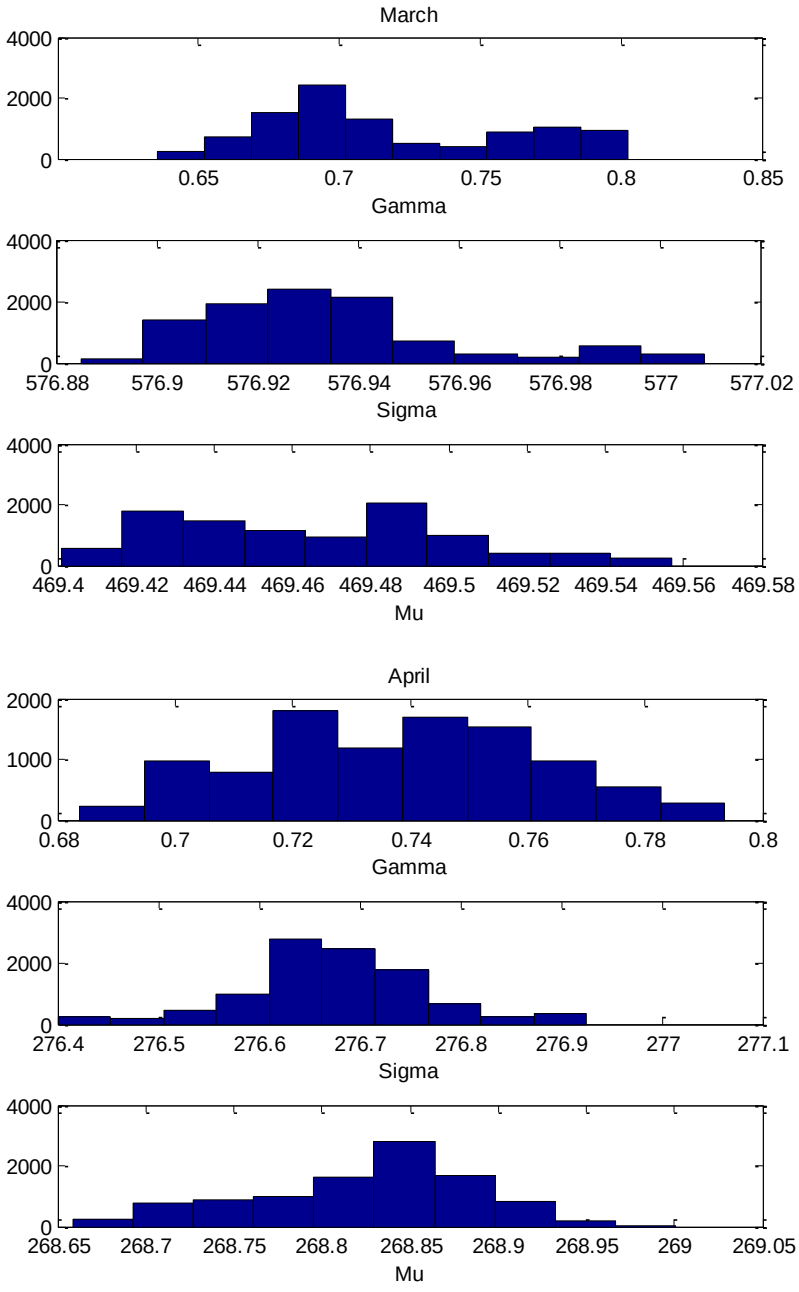
Table 5.1 Separately estimated marginal parameter values

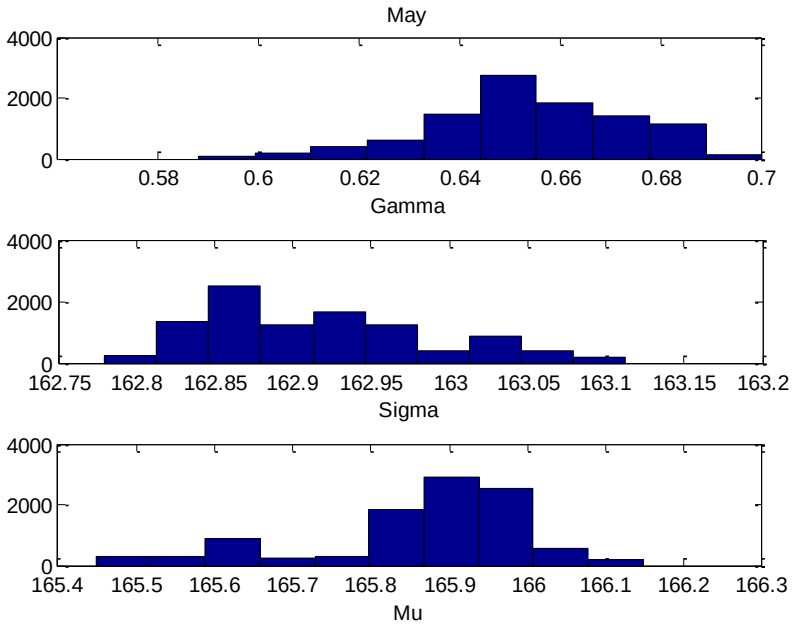
	γ	**90% HPD**	σ	**90% HPD**	μ	**90% HPD**
Jan	0,752	[0,6843; 0,7969]	580,1564	[580,1564; 580,2469]	499,0187	[498,9725; 499,0737]
Feb	0,7568	[0,6770; 0,8477]	598,1504	[598,0532; 598,2133]	477,9782	[477,9321; 478,1106]
Mar	0,7176	[0,6615; 0,7911]	576,9336	[576,9014; 576,9902]	469,4655	[469,4160; 469,5302]
Apr	0,7318	[0,6987; 0,7773]	276,6734	[276,5097; 276,8322]	268,8253	[268,7101; 268,9174]
May	0,6534	[0,6186; 0,6818]	162,9151	[162,8255; 163,0470]	165,8647	[165,5822; 166,0341]
June	0,6036	[0,5723; 0,6282]	139,4791	[139,4006; 139,5703]	70,1865	[70,0105; 70,4797]
July	0,6351	[0,5831; 0,7174]	39,9793	[39,9520; 40,0084]	39,9441	[39,8973; 39,9906]
Aug	0,5934	[0,5547; 0,6409]	150,7137	[150,3103; 151,3888]	82,0768	[81,7867; 82,3812]
Sept	0,6348	[0,593; 0,6664]	352,3933	[351,9014; 352,8726]	105,0991	[104,8119; 105,4879]
Oct	0,5852	[0,5011; 0,6655]	318,1884	[317,7815; 318,6110]	205,2203	[204,8383; 205,4684]
Nov	0,6505	[0,5622; 0,7235]	500,3164	[499,9310; 500,7252]	379,8821	[379,4282; 380,2970]
Dec	0,5616	[0,5130; 0,6247]	504,6988	[504,3956; 505,0051]	426,3184	[425,8315; 426,9137]

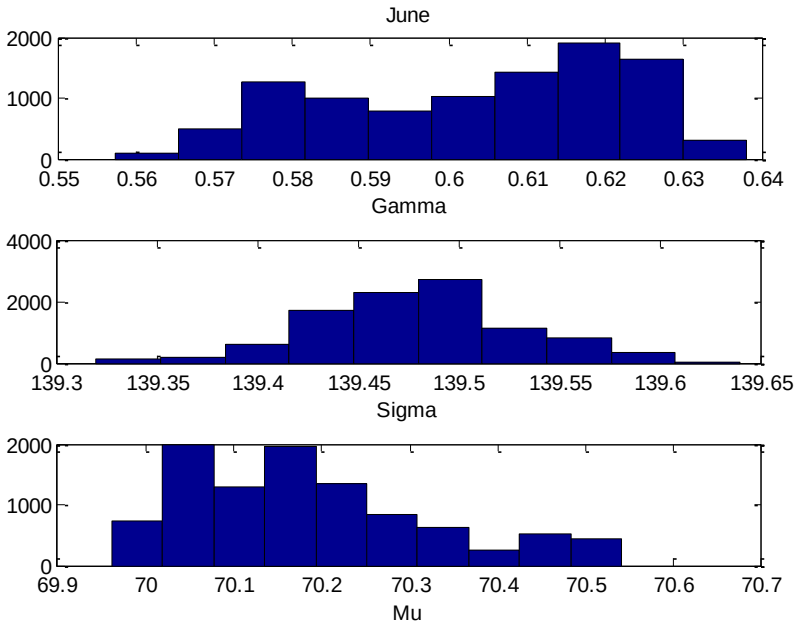
Figure 5.1 shows the histograms of the posterior distributions of the estimated parameter values for each month.

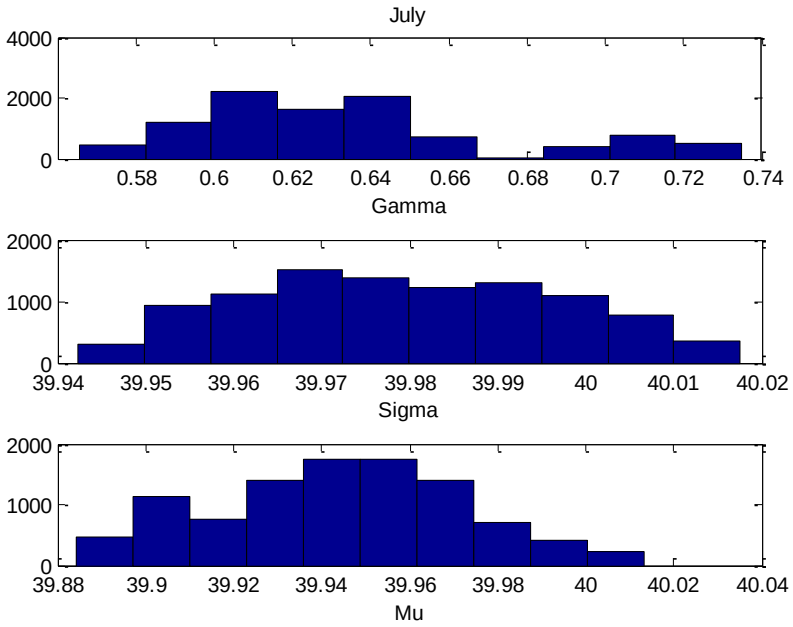
Fig. 5.1 Histograms of the posterior distributions

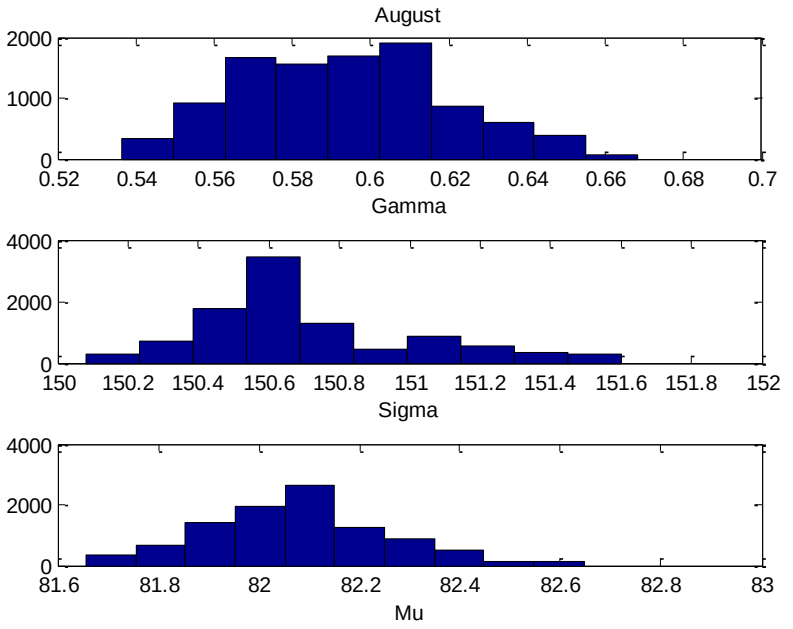


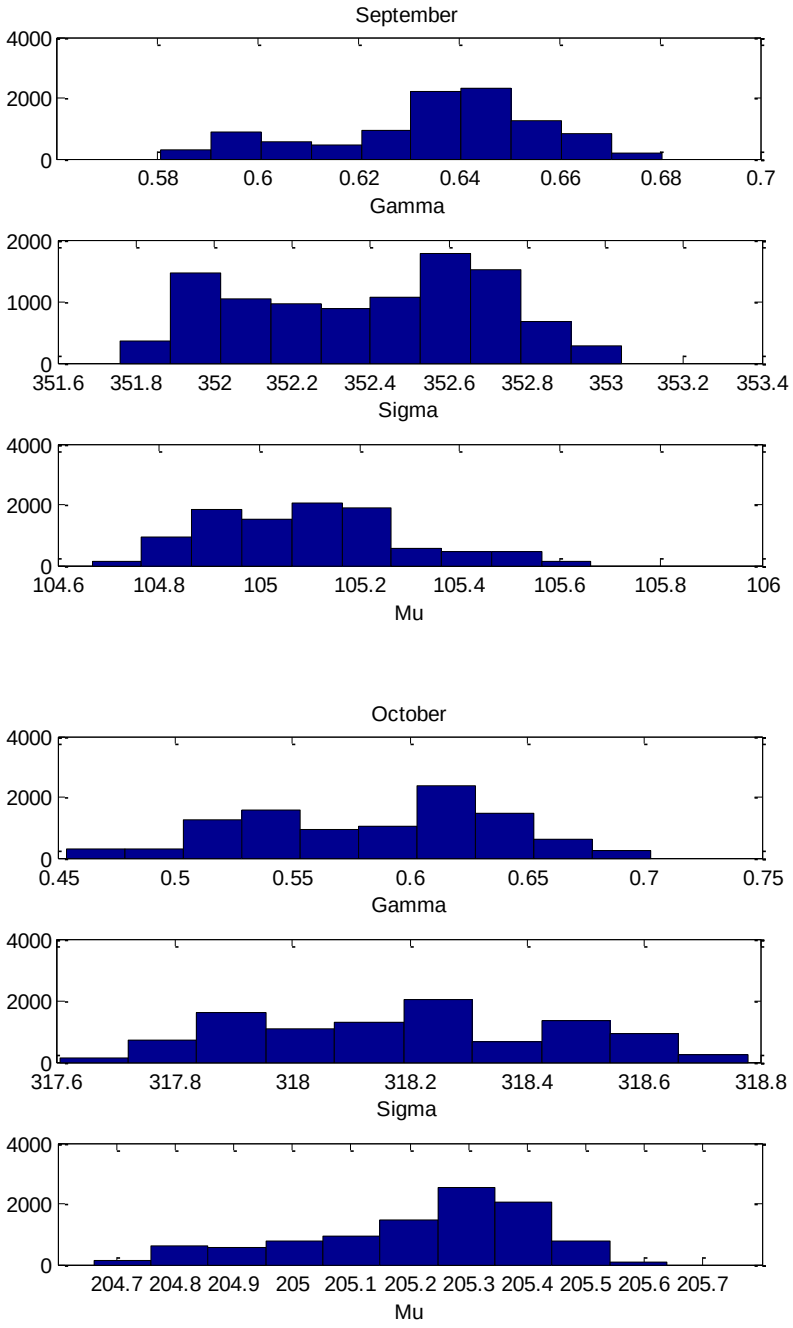

March
4000
2000
0
0.65
0.7
0.75
0.8
0.85
Gamma
4000
2000
0
576.88
576.9
576.92
576.94
576.96
576.98
577
577.02
Sigma
4000
2000
0
469.4
469.42
469.44
469.46
469.48
469.5
469.52
469.54
469.56
469.58
Mu
April
2000
1000
0
0.68
0.7
0.72
0.74
0.76
0.78
0.8
Gamma
4000
2000
0
276.4
276.5
276.6
276.7
276.8
276.9
277
277.1
Sigma
4000
2000
0
268.65
268.7
268.75
268.8
268.85
268.9
268.95
269
269.05
Mu

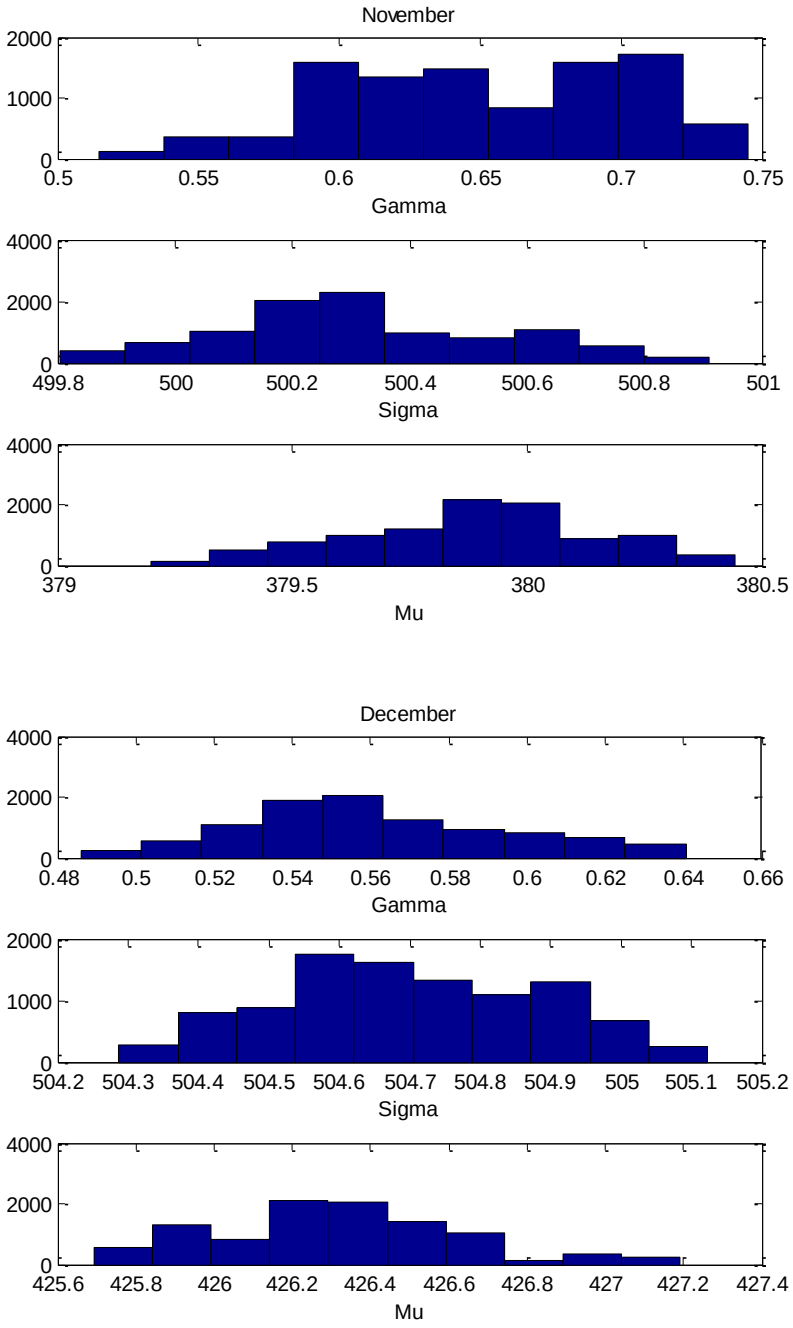

May
4000
2000
0
0.58
0.6
0.62
0.64
0.66
0.68
0.7
Gamma
4000
2000
0
162.75
162.8
162.85
162.9
162.95
163
163.05
163.1
163.15
163.2
Sigma
4000
2000
0
165.4
165.5
165.6
165.7
165.8
165.9
166
166.1
166.2
166.3
Mu


June
2000
1000
0
0.55
0.56
0.57
0.58
0.59
0.6
0.61
0.62
0.63
0.64
Gamma
4000
2000
0
139.3
139.35
139.4
139.45
139.5
139.55
139.6
139.65
Sigma
2000
1000
0
69.9
70
70.1
70.2
70.3
70.4
70.5
70.6
70.7
Mu


July
4000
2000
0
0.58 0.6 0.62 0.64 0.66 0.68 0.7 0.72 0.74
Gamma
2000
1000
0
39.94 39.95 39.96 39.97 39.98 39.99 40 40.01 40.02
Sigma
2000
1000
0
39.88 39.9 39.92 39.94 39.96 39.98 40 40.02 40.04
Mu


August
2000
1000
0
0.52 0.54 0.56 0.58 0.6 0.62 0.64 0.66 0.68 0.7
Gamma
4000
2000
0
150 150.2 150.4 150.6 150.8 151 151.2 151.4 151.6 151.8 152
Sigma
4000
2000
0
81.6 81.8 82 82.2 82.4 82.6 82.8 83
Mu


September
4000
2000
0
0.58
0.6
0.62
0.64
0.66
0.68
0.7
Gamma
2000
1000
0
351.6
351.8
352
352.2
352.4
352.6
352.8
353
353.2
353.4
Sigma
4000
2000
0
104.6
104.8
105
105.2
105.4
105.6
105.8
106
Mu
October
4000
2000
0
0.45
0.5
0.55
0.6
0.65
0.7
0.75
Gamma
4000
2000
0
317.6
317.8
318
318.2
318.4
318.6
318.8
Sigma
4000
2000
0
204.7
204.8
204.9
205
205.1
205.2
205.3
205.4
205.5
205.6
205.7
Mu



After estimating the parameters of the marginal distributions the dependence parameter of the Gumbel copula is estimated. Recall that the Gumbel copula (5.29) was given as

$$C(u_1,...,u_d) = \exp\left\{-\left[(-\ln u_1)^{\theta} + ... + (-\ln u_d)^{\theta}\right]\right\}^{\frac{1}{\theta}} \qquad (5.52)$$

where u_j represents the GEV marginal distributions:

$$u_j = \exp\left(-\left(1+\gamma_j \left(\frac{\mathbf{x}_j - \mu_j}{\sigma_j} \right) \right)^{\frac{-1}{\gamma_j}} \right), j = 1,...,d \,. \tag{5.53}$$

Now the Gumbel copula is given as

$$C(u_1,...,u_d) = \exp\left\{ -\left[\left[\left(1+\gamma_1 \left(\frac{\mathbf{x}_1 - \mu_1}{\sigma_1} \right) \right)^{\frac{-1}{\gamma_1}} \right]^{\theta} + ... + \left[\left(1+\gamma_d \left(\frac{\mathbf{x}_d - \mu_d}{\sigma_d} \right) \right)^{\frac{-1}{\gamma_d}} \right]^{\theta} \right]^{\frac{1}{\theta}} \right\} \,. \tag{5.54}$$

Let $v_j = \left[1+\gamma_j \left(\frac{\mathbf{x}_j - \mu_j}{\sigma_j} \right) \right]^{\frac{-1}{\gamma_j}}$, then $C(v_1,...,v_d) = \exp\left\{ -\left[(v_1)^{\theta} + ... + (v_d)^{\theta} \right] \right\}^{\frac{1}{\theta}}$.

The Gumbel copula is now considered for joining two marginals, firstly the marginals that represent the maximum monthly inflow of January and February, with a correlation of 0,2016, and secondly the marginals that represent the maximum monthly inflow of June and July, with a correlation of 0,2971.

The density function of the bivariate Gumbel copula, where the marginals represent the maximum monthly inflow of January and February, is derived as follows:

$$\frac{\partial^2 C(v_1,v_2)}{\partial v_1 \partial v_2} = -v_1^{\theta-1} \exp\left\{ -\left[v_1^{\theta} + v_2^{\theta} \right] \right\}^{\frac{1}{\theta}} \left\{ -\left[v_1^{\theta} + v_2^{\theta} \right] \right\}^{\frac{1}{\theta}-2} v_2^{\theta-1} \left\{ -\left(\frac{1}{\theta} - 1 \right) \theta + \left\{ -\left[v_1^{\theta} + v_2^{\theta} \right] \right\}^{\frac{1}{\theta}} \right\} \tag{5.55}$$

where $v_1 = \left[1+\gamma_1\left(\frac{\mathbf{x}_1-\mu_1}{\sigma_1}\right)\right]^{\frac{-1}{\gamma_1}}$ and $v_2 = \left[1+\gamma_2\left(\frac{\mathbf{x}_2-\mu_2}{\sigma_2}\right)\right]^{\frac{-1}{\gamma_2}}$.

The density function of the bivariate Gumbel copula where the marginals represent the maximum monthly inflow of June and July is derived as

$$\frac{\partial^2 C(v_6,v_7)}{\partial v_6 \partial v_7} = -v_6^{\theta-1}\exp\left\{-\left[v_6^\theta+v_7^\theta\right]\right\}^{\frac{1}{\theta}}\left\{-\left[v_6^\theta+v_6^\theta\right]\right\}^{\frac{1}{\theta}-2} v_7^{\theta-1}\left\{-\left(\frac{1}{\theta}-1\right)\theta+\left\{-\left[v_6^\theta+v_7^\theta\right]\right\}^{\frac{1}{\theta}}\right\} \quad (5.56)$$

where $v_6 = \left[1+\gamma_6\left(\frac{\mathbf{x}_6-\mu_6}{\sigma_6}\right)\right]^{\frac{-1}{\gamma_6}}$ and $v_7 = \left[1+\gamma_7\left(\frac{\mathbf{x}_7-\mu_7}{\sigma_7}\right)\right]^{\frac{-1}{\gamma_7}}$.

The Bayesian approach is also considered for estimating the dependence parameter θ. The prior on θ is assumed to be a log-uniform prior, $\log\pi(\theta)\propto 1$.

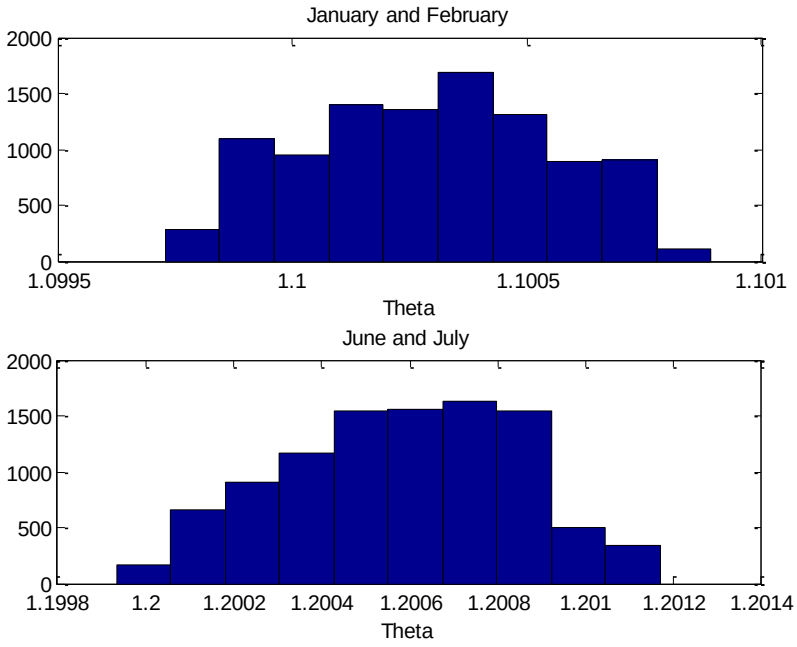
Table 5.2 shows the estimated dependence parameter and the 90% HPD interval, firstly between January and February and secondly between June and July.

Table 5.2 Separately estimated dependence parameter

	θ	**90% HPD**
January and February	1,1003	[1,0999;1,1007]
June and July	1,2006	[1,2001;1,2010]

Figure 5.2 shows the histograms of the posterior distributions for the estimated dependence parameter of the Gumbel copula between January and February and between June and July.

Fig. 5.2 Histograms of the posterior distributions



5.5.1.2 Applying the joint estimation approach

When applying the joint estimation approach to estimate the parameters the marginals are again assumed to be GEV distributed. Again the Gumbel copula is used for joining the marginal distributions.

First the parameters are estimated jointly by assuming that the marginals represent the maximum monthly inflow of January and February. Then the parameters are estimated jointly by assuming that the marginals represent the maximum monthly inflow of June and July. The density of the Gumbel copula is derived in (5.55) and (5.56). An MDI prior is assumed on the marginal parameters and a uniform prior is assumed on the dependence parameter, the log of the joint prior distribution is

$$\log \pi\left(\sigma_1,\gamma_1,\mu_1,\sigma_2,\gamma_2,\mu_2,\theta\right) \propto -\log\sigma_1 - 1 - \left(1+\frac{1}{\gamma_1}\right)0,5772 - \log\sigma_2 - 1 - \left(1+\frac{1}{\gamma_2}\right)0,5772. \tag{5.57}$$

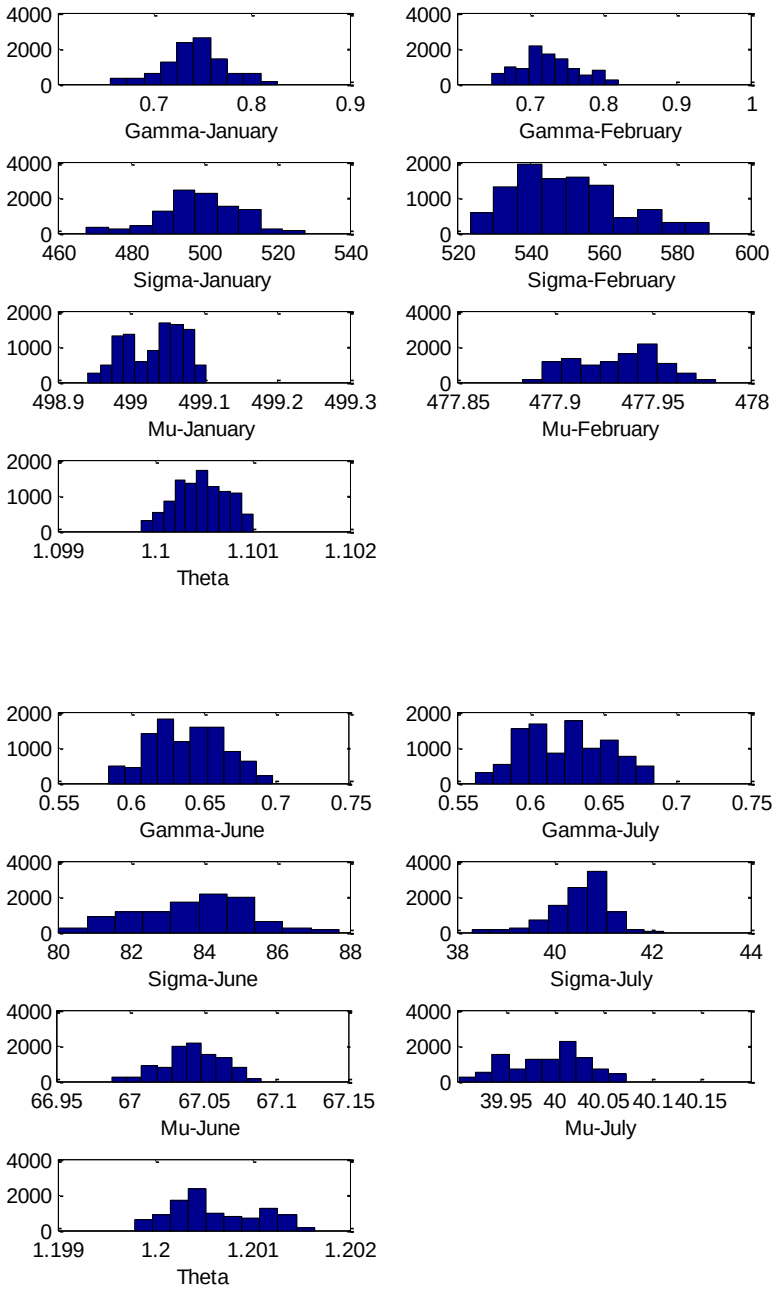
Table 5.3 shows the jointly estimated parameters, for January and February and for June and July, and the respective 90% HPD intervals.

Table 5.3 Joint estimated parameters

January and February							
	$\hat{\gamma}_{Jan}$	$\hat{\sigma}_{Jan}$	$\hat{\mu}_{Jan}$	$\hat{\gamma}_{Feb}$	$\hat{\sigma}_{Feb}$	$\hat{\mu}_{Feb}$	$\hat{\theta}$
	0,740	498,806	499,033	0,725	549,860	477,932	1,101
90% HPD	[0,680 ; 0,795]	[478,578 ; 514,743]	[498,970 ; 499,087]	[0,663 ; 0,797]	[529,783 ; 578,198]	[477,897 ; 477,965]	[1,100 ; 1,101]
June and July							
	$\hat{\gamma}_{June}$	$\hat{\sigma}_{June}$	$\hat{\mu}_{June}$	$\hat{\gamma}_{July}$	$\hat{\sigma}_{July}$	$\hat{\mu}_{July}$	$\hat{\theta}$
	0,639	83,641	67,042	0,624	40,53	39,926	1,201
90% HPD	[0,597 ; 0,679]	[81,04 ; 85,743]	[67,007 ; 67,073]	[0,583 ; 0,672]	[39,483 ; 41,233]	[39,93 ; 40,05]	[1,200 ; 1,201]

Figure 5.3 shows the histograms of the posterior distributions of the jointly estimated parameters, with the Gumbel copula, for January and February and for June and July.

Fig. 5.3 Histograms of the posterior distributions



5.5.2 Fitting the Tawn copula

5.5.2.1 Applying the separate estimation approach

The marginal distributions are again assumed to be GEV distributed as shown in (5.47). The same marginal parameter estimates given in Table 5.1 hold. The dependence parameters of the Tawn copula are now estimated. Recall that the Tawn copula (5.32) was given as

$$C(u_1,u_2)=u_1^{1-\alpha}u_2^{1-\beta}\exp\left\{-\left[\left(-\alpha\ln u_1\right)^{\theta}+\left(-\beta\ln u_2\right)^{\theta}\right]^{\frac{1}{\theta}}\right\} \tag{5.58}$$

where u_j represent the GEV distributions shown by the following equation:

$$u_j=\exp\left(-\left(1+\gamma_j\left(\frac{\mathbf{x}_j-\mu_j}{\sigma_j}\right)\right)^{\frac{-1}{\gamma_j}}\right),\ j=1,...,d\ . \tag{5.59}$$

Here the Tawn copula is considered for joining two marginals, firstly the marginals that represents the maximum monthly inflow of January and February and secondly the marginals that represents the maximum monthly inflow of June and July.

The density function of the bivariate Tawn copula, where the marginals represent the maximum monthly inflow of January and February, is derived as follows:

$$\frac{\partial^2 C(u_1,u_2)}{\partial u_1 \partial u_2} = u_1^{-\alpha} \exp\left\{-\left[(-\alpha \ln u_1)^{\theta} + (-\phi \ln u_2)^{\theta}\right]^{\frac{1}{\theta}}\right\}$$
$$\left[(1-\alpha)u_2^{-\beta}\left\{(1-\beta) - u_2\left[(-\alpha \ln u_1)^{\theta} + (-\phi \ln u_2)^{\theta}\right]^{\frac{1}{\theta}-1}(-\beta \log u_2)^{\theta}(\log u_2)^{-1}\right\}\right.$$
$$+(-\alpha \log u_1)^{\theta}(\log u_1)^{-1}\left[(-\alpha \ln u_1)^{\theta} + (-\phi \ln u_2)^{\theta}\right]^{\frac{1}{\theta}-1}$$
$$\left\{u_2^{1-\beta}(-\beta \log u_2)^{\theta}(\log u_2)^{-1}\left[(-\alpha \ln u_1)^{\theta} + (-\phi \ln u_2)^{\theta}\right]^{-1}\right.$$
$$\left.\left.\left[\left[(-\alpha \ln u_1)^{\theta} + (-\phi \ln u_2)^{\theta}\right]^{\frac{1}{\theta}} - \left(\frac{1}{\theta} - 1\right)\right] - (1-\beta)u_2^{-\beta}\right\}\right] \quad (5.60)$$

where $u_1 = \exp\left(-\left(1+\gamma_1\left(\frac{\mathbf{x}_1 - \mu_1}{\sigma_1}\right)\right)^{\frac{-1}{\gamma_1}}\right)$ and $u_2 = \exp\left(-\left(1+\gamma_2\left(\frac{\mathbf{x}_2 - \mu_2}{\sigma_2}\right)\right)^{\frac{-1}{\gamma_2}}\right)$.

And the density function of the bivariate Tawn copula, where the marginals represent the maximum monthly inflow of June and July, is derived as follows:

$$\frac{\partial^2 C(u_6,u_7)}{\partial u_6 \partial u_7} = u_6^{-\alpha} \exp\left\{-\left[(-\alpha \ln u_6)^{\theta} + (-\phi \ln u_7)^{\theta}\right]^{\frac{1}{\theta}}\right\}$$
$$\left[(1-\alpha)u_7^{-\beta}\left\{(1-\beta) - u_7\left[(-\alpha \ln u_6)^{\theta} + (-\phi \ln u_7)^{\theta}\right]^{\frac{1}{\theta}-1}(-\beta \log u_7)^{\theta}(\log u_7)^{-1}\right\}\right.$$
$$+(-\alpha \log u_6)^{\theta}(\log u_6)^{-1}\left[(-\alpha \ln u_6)^{\theta} + (-\phi \ln u_7)^{\theta}\right]^{\frac{1}{\theta}-1}$$
$$\left.\left\{u_7^{1-\beta}(-\beta \log u_7)^{\theta}(\log u_7)^{-1}\left[(-\alpha \ln u_6)^{\theta} + (-\phi \ln u_7)^{\theta}\right]^{-1}\left[\left[(-\alpha \ln u_6)^{\theta} + (-\phi \ln u_7)^{\theta}\right]^{\frac{1}{\theta}} - \left(\frac{1}{\theta} - 1\right)\right] - (1-\beta)u_7^{-\beta}\right\}\right] \quad (5.61)$$

where $u_6 = \exp\left(-\left(1+\gamma_6\left(\frac{\mathbf{x}_6 - \mu_6}{\sigma_6}\right)\right)^{\frac{-1}{\gamma_6}}\right)$ and $u_7 = \exp\left(-\left(1+\gamma_7\left(\frac{\mathbf{x}_7 - \mu_7}{\sigma_7}\right)\right)^{\frac{-1}{\gamma_7}}\right)$.

The Bayesian approach is again considered for estimating the dependence parameters θ, α and β. The prior on the dependence parameters is assumed to be a log-uniform prior, $\log \pi(\theta, \alpha, \beta) \propto 1$.

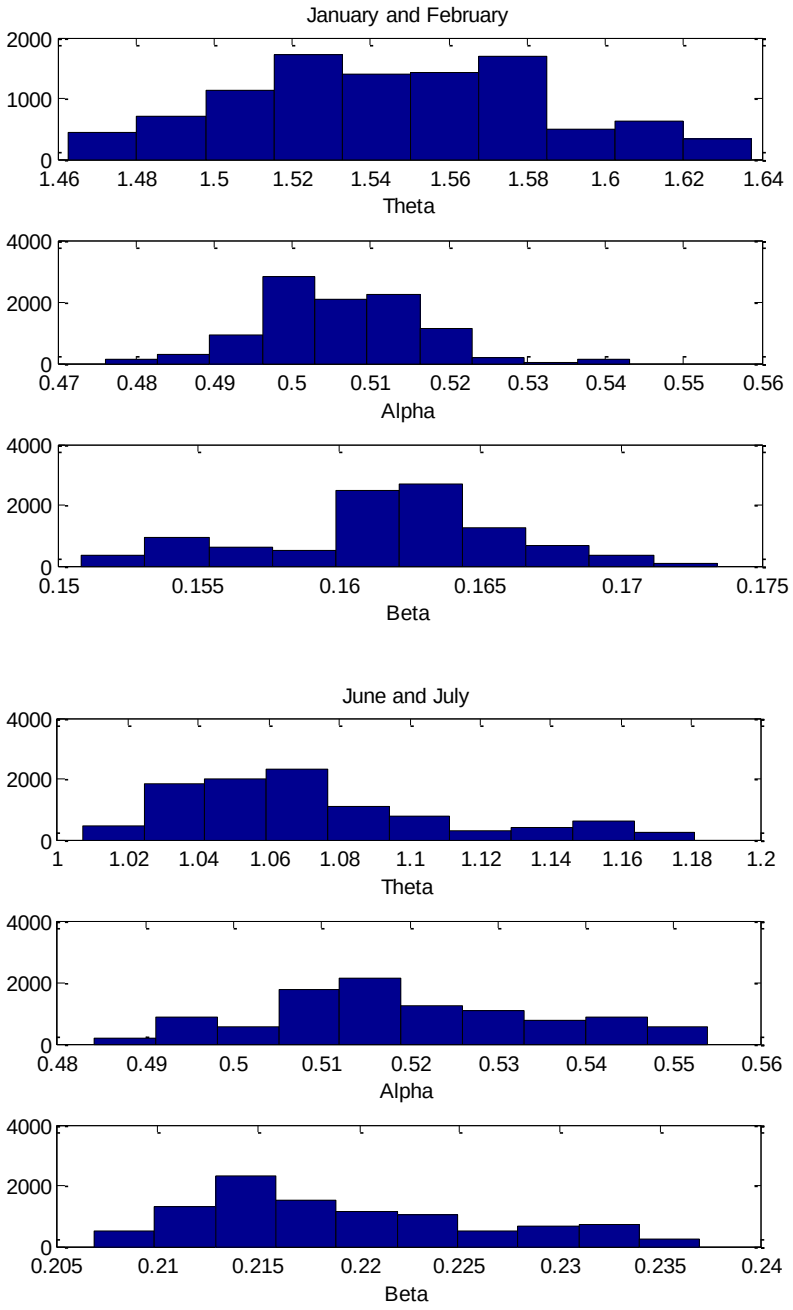
Table 5.4 shows the estimated dependence parameters for January and February and for June and July and the respective 90% HPD intervals.

Table 5.4 Separately estimated dependence parameters

	$\hat{\theta}$	**90% HPD**	$\hat{\alpha}$	**90% HPD**	β	**90% HPD**
Jan & Feb	1,0722	[1,0250; 1,1556]	0,5194	[0,4937; 0,5372]	0,2192	[0,2099; 0,2329]
June & July	1,5458	[1,4818; 1,6129]	0,5063	[0,4907; 0,5211]	0,1617	[0,1534; 0,1687]

Figure 5.4 shows the histograms of the posterior distributions of the estimated dependence parameters, with the Tawn copula, for January and February and for June and July.

Fig. 5.4 Histograms of the posterior distributions



5.5.2.2 Applying the joint estimation approach

When applying the joint estimation approach for estimating the parameters the marginals are again assumed to be GEV distributed. Again the Tawn copula is used to join the marginal distributions. First the parameters are estimated jointly by assuming that the marginals represent the maximum monthly inflow

of January and February and then by assuming that the marginals represent the maximum monthly inflow of June and July. The density of the Tawn copula is derived in (5.60) and (5.61). An MDI prior is assumed on the marginal parameters and a uniform prior is assumed on the dependence parameters, the log of the joint prior is

$$\log \pi\left(\sigma_1,\gamma_1,\mu_1,\sigma_2,\gamma_2,\mu_2,\theta,\alpha,\beta\right) \propto -\log\sigma_1 - 1 - \left(1+\frac{1}{\gamma_1}\right)0,5772 - \log\sigma_2 - 1 - \left(1+\frac{1}{\gamma_2}\right)0,5772 \quad (5.62)$$

Table 5.5 shows the jointly estimated parameter values between January and February and between June and July respectively, and the respective 90% HPD intervals.

Table 5.5 Joint estimated parameters

January and February							
	$\hat{\gamma}_{\text{Jan}}$	$\hat{\sigma}_{\text{Jan}}$	$\hat{\mu}_{\text{Jan}}$	$\hat{\gamma}_{\text{Feb}}$	$\hat{\sigma}_{\text{Feb}}$	$\hat{\mu}_{\text{Feb}}$	$\hat{\theta}$
	0,6710	851,6723	499,0167	0,7275	690,5699	476,2534	1,4443
90% HPD	[0,6176 ; 0,7101]	[822,4612 ; 877,0449]	[498,9683 ; 499,0551]	[0,7037 ; 0,7508]	[663,7029 ; 717,0113]	[476,1999 ; 476,3311]	[1,3767 ; 1,5006]

	$\hat{\alpha}$	$\hat{\beta}$
	0,3833	0,2039
90% HPD	[0,3589 ; 0,4010]	[0,1967 ; 0,2112]

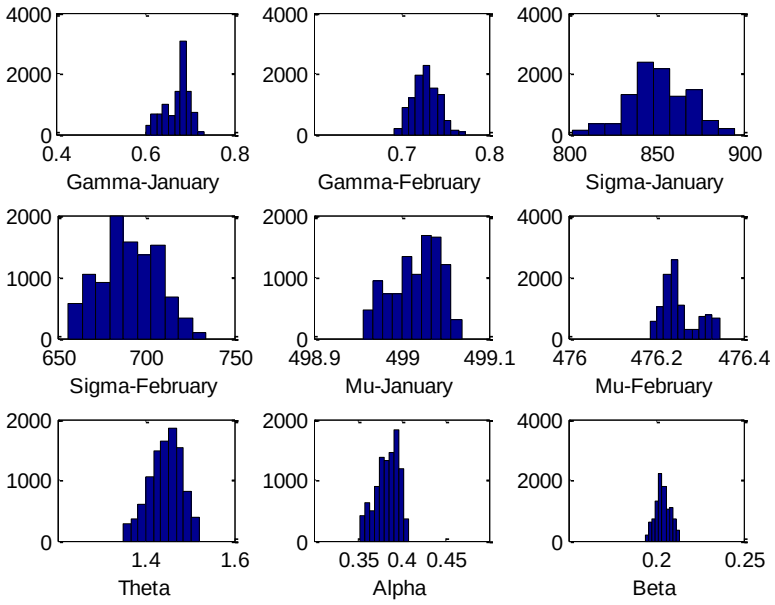
June and July							
	$\hat{\gamma}_{\text{June}}$	$\hat{\sigma}_{\text{June}}$	$\hat{\mu}_{\text{June}}$	$\hat{\gamma}_{\text{July}}$	$\hat{\sigma}_{\text{July}}$	$\hat{\mu}_{\text{July}}$	$\hat{\theta}$
	0,5995	50,9266	66,9925	0,5821	41,4228	40,9958	1,5930

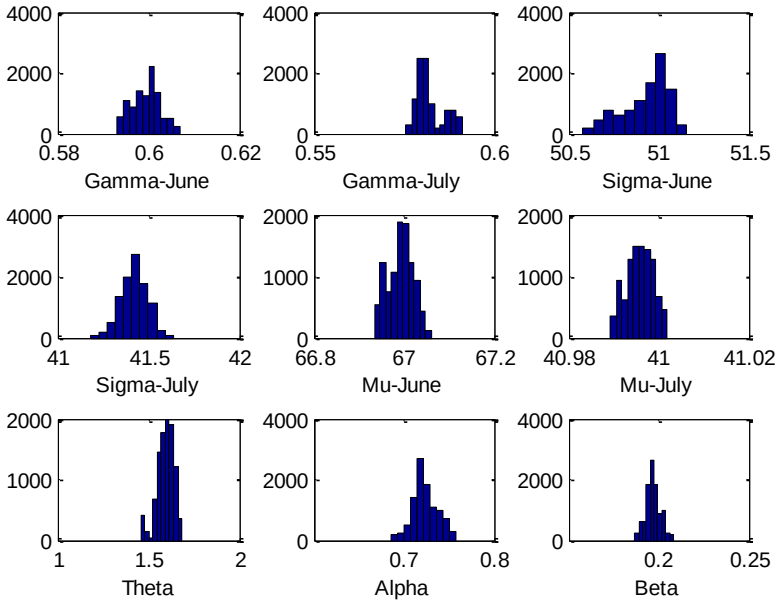
90% HPD	[0,5943 ; 0,6048]	[50,6769 ; 51,0796]	[66,9464 ; 67,0370]	[0,5778 ; 0,5896]	[41,2984 ; 41,5345]	[40,9907 ; 41,0006]	[1,4841 ; 1,6562]

	$\hat{\alpha}$	$\hat{\beta}$
	0,7239	0,1969
90% HPD	[0,7023 ; 0,7485]	[0,1898 ; 0,2036]

Figure 5.5 shows the histograms of the posterior distributions of the jointly estimated parameters for January and February and for June and July.

Fig. 5.5 Histograms of the posterior distributions





5.6 Calculating tail probabilities

Tail probabilities, such as the probability that the maximum inflow of January exceeds some high value and that the maximum inflow of February also exceeds a high value, are of interest. This tail probability can be calculated as follows:

$$P\left(X_1 > 3000, X_2 > 11000\right) = 1 - P\left(X_1 < 3000\right) - P\left(X_2 < 11000\right) + P\left(X_1 < 3000, X_2 < 11000\right) \tag{5.63}$$

where X_1 represents the maximum inflow of January and X_2 represents the maximum inflow of February. Because the marginals are GEV distributed (5.63) is transformed to

$$P\left(U_1 > 0{,}8887; U_2 > 0{,}9761\right) \tag{5.64}$$

where U_1 and U_1 are GEV distributed with the estimated parameters given in Table 5.3 and 5.5 respectively. In the case of the Gumbel copula (5.63) is calculated as

$$P(U_1 > 0,8638; U_2 > 0,9707) = 1 + C(0,8638; 0,9707) - C(0,8176; 1) - C(1; 0,9707)$$
$$= 1 + 0,9034 - 0,9171 - 0,9850$$
$$= 0,0013.$$

And in the case of the Tawn copula (5.63) is calculated as

$$P(U_1 > 0,8209; U_2 > 0,9680) = 1 + C(0,8209; 0,9680) - C(0,8209; 1) - C(1; 0,9680)$$
$$= 1 + 0,7987 - 0,8209 - 0,9680$$
$$= 0,0098.$$

For the Gariep Dam data set the probability $P(X_1 > 3000; X_2 > 11000)$ is calculated from the actual data. $X_1 > 3000$ and $X_2 > 11000$ applies to 1 out of the 29 data points, therefore the empirical probability is 0,0344. The Tawn copula's probability is the closest to the empirical probability. The Tawn copula is therefore a better fit to this data set than the Gumbel copula.

It may be possible to apply other criteria, such as the Bayesian information criteria (BIC) or the Akaike's information criteria (AIC) (McQuarrie and Tsai, 1998) to select the best copula; but this is not considered in this thesis.

5.7 A Dirichlet mixture model

From this section onwards the modelling of extreme values is considered through a Dirichlet mixture model.

5.7.1. Introduction

The Dirichlet mixture model is another way of modelling multivariate extreme values. The Dirichlet mixture model is based on mixtures of Dirichlet distributions.

Here the estimation of the joint tail of a multivariate distribution depends on the spectral distribution function. The spectral distribution function mainly

measures the dependence among extreme values. Detailed theory on the spectral measure is reviewed by Beirlant *et al*. (2004, Chapters 8 and 9) and Kotz and Nadarajah (2000, Chapter 3). Parametric models and nonparametric approaches are suggested for estimating the spectral distribution function.

The use of parametric models for the modelling of a joint tail for a large number of dimensions is problematic due to insufficient flexibility (Boldi and Davison 2007, p. 217).

Nonparametric approaches are useful but only in a two or three dimensional case; therefore Boldi and Davison (2007) proposed a semi parametric model with mixtures of Dirichlet distributions. Section 5.7.2 discusses the properties of the mixture model and Section 5.7.3 shows a practical application of the Dirichlet mixture model. A conclusion is given in Section 5.8.

5.7.2 Properties of the mixture model

Let $Y_1,...,Y_n \in \mathbb{R}^p$ be i.i.d random vectors with distribution function *F* and Fréchet marginal distributions. If the renormalized maximum $M_n = \max(Y_1,...,Y_n)/n$ on the *p*-dimensional vectors, $Y_i, i = 1,...,n$, converges in distribution to a non-degenerate distribution function *G*, then it has the form

$$G(y) = \lim_{x\to\infty}\{P(M_n \le y)\} = \exp\left\{-p\int_{S_p}\max\left(\frac{w_1}{y_1},...,\frac{w_p}{y_p}\right)H(\partial w)\right\} \quad (5.65)$$

(Boldi and Davison 2007, p. 218).

H is called the spectral distribution function on the *p*-dimensional unit simplex $S_p = \left\{w \in \mathbb{R}_+^p : \sum_{j=1}^{p} w_j = 1\right\}$ and *H* also satisfies the mean constraint

$$\int_{S_p} w_j H(dw) = p^{-1}, j = 1,..., p \tag{5.66}$$

(Boldi and Davison 2007, p. 218).

Further, a datum $(y_1,..., y_p)$ is extreme if its pseudo-radius, given by

$$r = \sum_{j=1}^{p} y_j \tag{5.67}$$

is large. In such a case the pseudo-angle, given by

$$w = \frac{y}{r} \tag{5.68}$$

is independent of *r* and distributed according to *H* (Boldi and Davison 2007, p. 218).

When constructing spectral distribution functions it is important that (5.66) is satisfied. Boldi and Davison (2006, p. 219) suggests that the Dirichlet mixture model imposes the best spectral density $h(w)$. This density, known as the mixture model, is given by

$$h(w) = \sum_{m=1}^{k} \pi_m \frac{\Gamma(\nu_m)}{\prod_{j=1}^{p} \Gamma(\nu_m \mu_j^{(m)})} \prod_{j=1}^{p} w_j^{\nu_m \mu_j^{(m)} - 1}, w \in S_p \,. \tag{5.69}$$

For $j = 1,..., p$ and $m = 1,..., k$, the constraints on the mixture model are $\pi_m \geq 0, \sum_{m=1}^{k} \pi_m = 1, \nu_m > 0, \mu_j^{(m)} \geq 0, \sum_{j=1}^{p} \mu_j^{(m)} = 1$ and (5.66) is satisfied if $\sum_{m=1}^{k} \pi_m \mu_j^{(m)} = p^{-1}$.

The Bayesian information criteria (BIC) are used to select the number of components (*k*) for the finite mixtures. The BIC is given by the following equation:

$$BIC(k) = -2\hat{l}_k + d\log(n), \qquad (5.70)$$

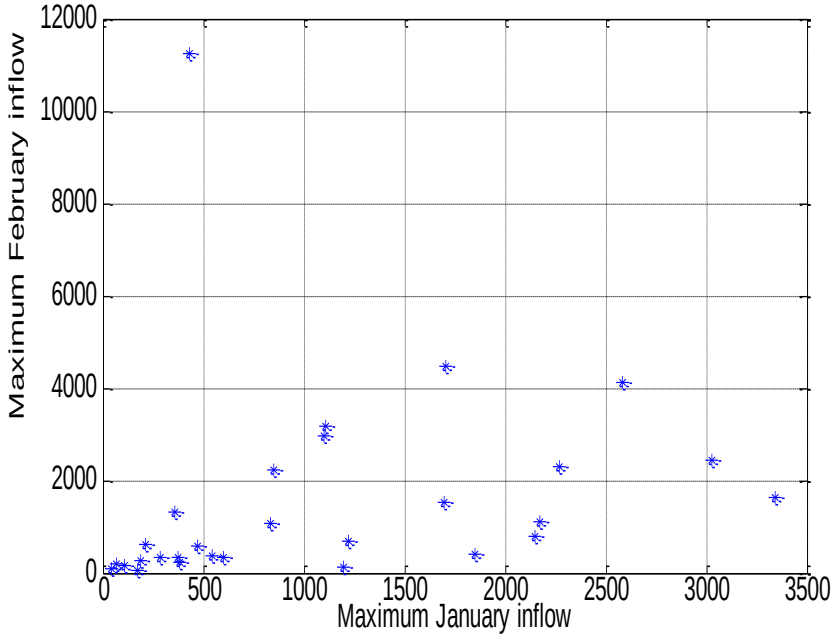
where $d = k + p(k-1)$ is the number of parameters, $\hat{l}_k$ is the maximized likelihood and *n* is the sample size (Boldi and Davision 2007, p. 221-222). The optimum value for *k* will be the value that gives the smallest BIC (McQuarrie and Tsai 1998).

5.7.3 Practical application

5.7.3.1 The data set

Again the Gariep Dam data set is considered. $X_1,...,X_p$ denotes the maximum inflow of each month, January to December. First a two dimensional case is considered, where *p* = 2. The aim is to model $X = (X_1, X_2)$, where X_1 represents the maximum inflow of January over a period of 29 years and X_2 represents the maximum inflow of February over the same period of 29 years. Again we assume that the marginal distributions are GEV distributed. The maximum inflow data for January and February are shown in Figure 5.6.

Fig. 5.6 The maximum inflow values of January and February over a period of 29 years



5.7.3.2 Marginal distributions

A GEV distribution is fitted to each of the marginals. If $X \sim GEV(\mu, \sigma, \gamma)$ then the distribution function, as discussed before, is given by

$$G(x) = \exp\left(-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{\frac{-1}{\gamma}}\right), x > 0. \qquad (5.71)$$

The estimated parameter values are given in Table 5.6. This is an extract from Table 5.1.

Table 5.6 The estimated parameter values

	$\hat{\gamma}$	**90% HPD**	$\hat{\sigma}$	**90% HPD**	$\hat{\mu}$	**90% HPD**
Jan	0,752	[0,6843; 0,7969]	580,1564	[580,0751; 580,2489]	499,0187	[498,9725; 499,0737]
Feb	0,7568	[0,6770; 0,8477]	598,1504	[598,0532; 598,2133]	477,9782	[477,9321; 478,1106]

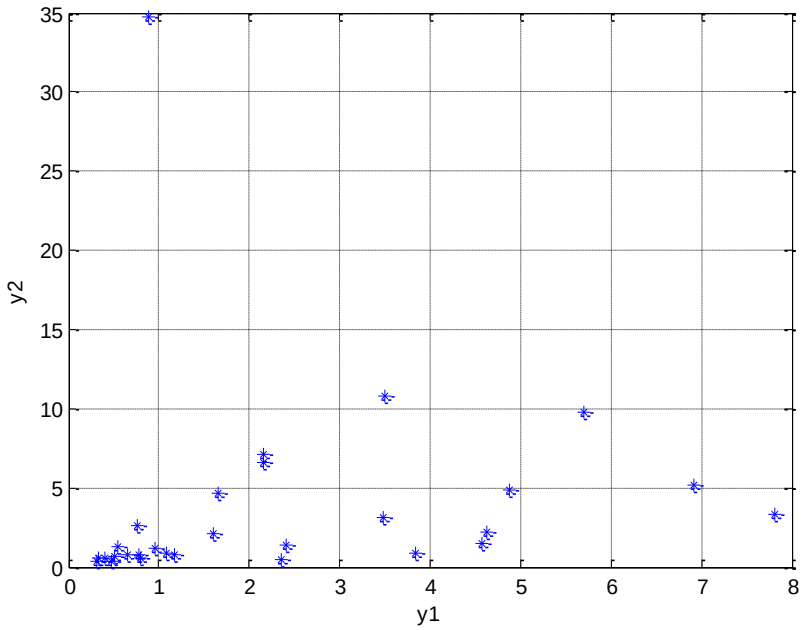
5.7.3.3 Fréchet transformations

When working with the mixture model the marginals undergo a Fréchet transformation:

$$Y = -\frac{1}{\log\left(F\left(X\right)\right)}, \tag{5.72}$$

where *F* denotes the respective GEV marginal distribution function. The newly transformed variables are denoted by Y_1 and Y_2 for the maximum inflows of January and February respectively. Figure 5.7 shows the Fréchet transformed data**.**

Fig 5.7 Fréchet transformed maximum inflows y_1 and y_2



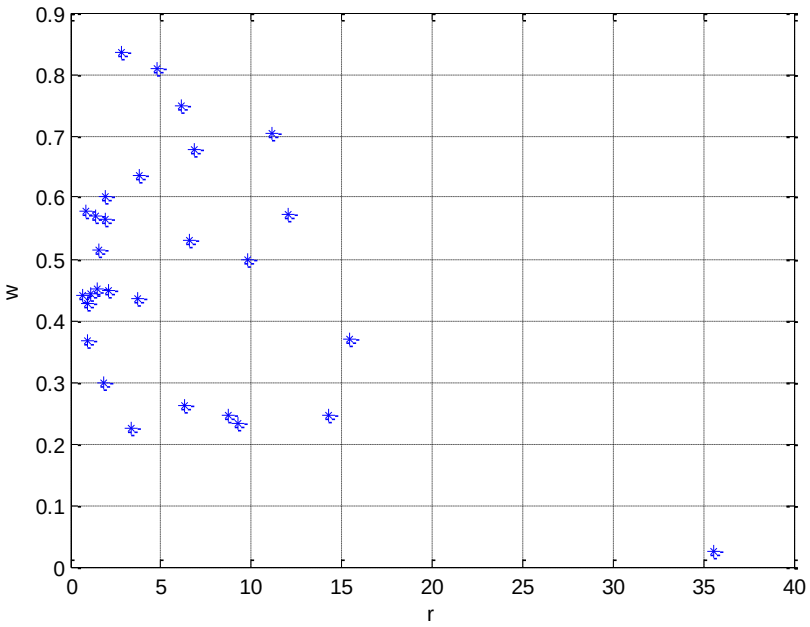
5.7.3.4 Fitting the spectral density and estimating the parameters

Fitting the spectral density depends of the selection of some high threshold (r_0) so that the pseudo angles are asymptotical independent of the radius ($R > r_0$) and so that the corresponding angles of the datum points are considered as an independent sample (Boldi and Davison 2006, p. 219). The parameters of the *k* mixtures are estimated jointly with the other parameters. For this study all the data is taken into consideration when fitting the mixture model to the w's. The Metropolis Hastings MCMC algorithm is used to estimate the parameters and a uniform prior is assumed. v_m and $\mu_j^{(m)}$ are estimated jointly as $a_j^{(m)}$, and solved separately from the constraints afterwards (De Waal, 2007).

In the two dimensional case, where $p = 2$, the Dirichlet distributions become Beta distributions. Therefore, a mixture of Beta distributions is fitted to the pseudo angles.

Figure 5.8 shows the pseudo radius *r* plotted against the pseudo angles w for the two dimensional case.

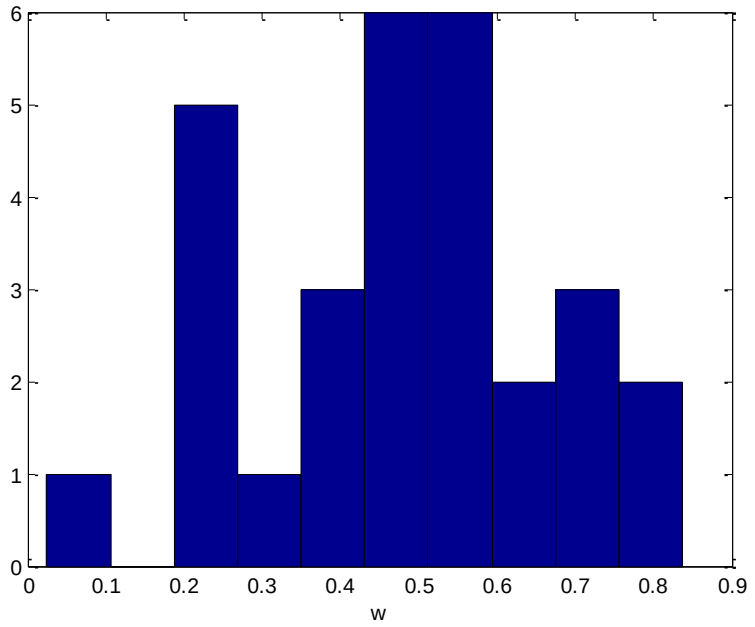
Fig 5.8 Plot of pseudo radius *r* against pseudo angles w



As expected, Figure 5.8 shows little dependence between *r* and w.

Figure 5.9 shows the histogram of the pseudo angles w.

Fig. 5.9 The histogram of the pseudo angles w



From Figure 5.9 it seems that a mixture of 2 or 3 Beta distributions will best fit the data. The BIC (5.58) is used as a method for deciding on the number of Beta distributions to include into the mixture. In this case the mixture of 2 Beta distributions, 3 Beta distributions and 4 Beta distributions are investigated. Therefore, for each value of *k* (2, 3 and 4) the BIC is calculated and then the mixture with the smallest BIC value is chosen. The mixture of 2 Beta distributions is given by

$$h(w) = \pi_1 Beta(w|a_1, b_1) + \pi_2 Beta(w|a_2, b_2) \qquad (5.73)$$

(De Waal, 2007).

The estimates of the parameters of the mixture of the 2 Beta distributions are obtained through the Metropolis Hastings algorithm assuming uniform priors on the parameters π, a_j and $b_j, j = 1,2$. The estimates are given in Table 5.7.

Table 5.7 Estimated parameters of the mixture of the 2 Beta distributions

$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{a}_1$	$\hat{b}_1$	$\hat{a}_2$	$\hat{b}_2$
0,3785	0,6215	39,7301	30,7709	3,0238	3,3835

When imposing the constraints, $v_m = a_m + b_m; m = 1,2$ the estimates for $\hat{v}_m$ are obtained and from there the estimates for $\hat{\mu}_j^{(m)}$ are calculated. The estimates are shown in Table 5.8.

Table 5.8 Estimated parameters of the mixture of the 2 Beta distributions

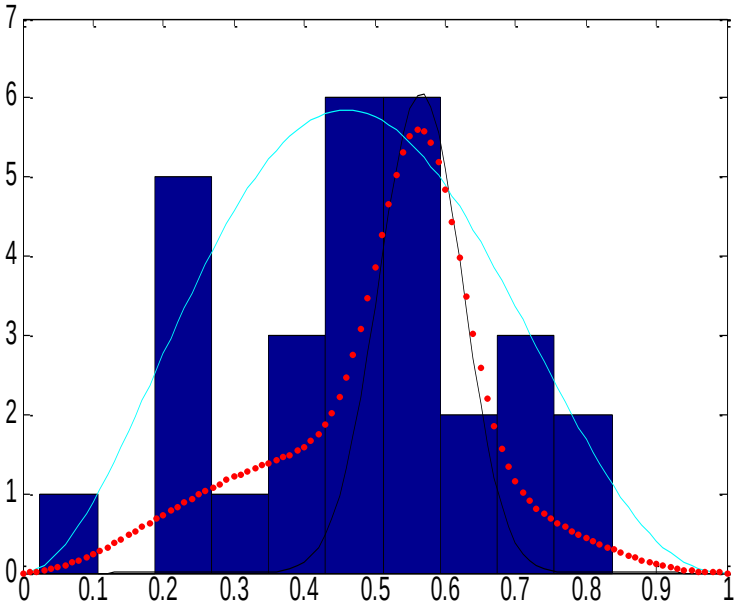
$\hat{v}_1$	$\hat{v}_2$	$\hat{\mu}_1^{(1)}$	$\hat{\mu}_2^{(1)}$	$\hat{\mu}_1^{(2)}$	$\hat{\mu}_2^{(2)}$
70,5010	6,4073	0,5635	0,4365	0,4719	0,5281

The BIC is calculated as 5,6852.

The estimates of the parameters of the mixture of 3 Beta distributions and 4 Beta distributions are also obtained through the Metropolis Hastings algorithm assuming uniform priors on the parameters π, a_j and $b_j, j = 1,2,3$ and π, a_j and $b_j, j = 1,2,3,4$ respectively. The BIC for the mixture of 3 Beta distributions is calculated as 30,7079 and for the mixture of 4 Beta distributions the BIC is calculated as 32,8039. Judging the BIC values the mixture of 2 Beta distributions has the lowest BIC value; therefore the mixture of 2 Beta distributions is used.

Figure 5.10 shows the mixture distribution in (5.73) fitted to the $w's$.

Fig. 5.10 **The mixture of the two Beta's (red dotted line) fitted to the pseudo angles. The blue and black lines show the two separate Beta densities**



After estimating the parameters, probabilities are calculated. Tail probabilities such as the following are calculated:

$$P\left(X_1 > 3000, X_2 > 11000\right) = 1 - P\left(X_1 < 3000\right) - P\left(X_2 < 11000\right) + P\left(X_1 < 3000, X_2 < 11000\right) \tag{5.74}$$

The actual maximum inflow of January was 3 344,30 m^3/s and 11 263 m^3/s for February.

These variables first undergo the Fréchet transformations as shown below.

$$F\left(X_1\right)=\exp\left\{-\left[1+\frac{0,752\left(3000-499,0187\right)}{580,1564}\right]^{\frac{-1}{0,752}}\right\}$$
$$=0,8638; \tag{5.75}$$
$$x_1^*=\frac{-1}{\log\left(F\left(X_1\right)\right)}=6,8299$$

and

$$F\left(X_2\right)=\exp\left\{-\left[1+\frac{0,7568\left(11000-477,9782\right)}{598,1504}\right]^{\frac{-1}{0,7568}}\right\}$$
$$=0,9707; \tag{5.76}$$
$$x_2^*=\frac{-1}{\log\left(F\left(X_2\right)\right)}=33,6272.$$

Then (5.65) is used to calculate the joint probability as follows:

$$P\left(X_1^*<x_1^*,X_2^*<x_2^*\right)=G\left(x\right)$$
$$=\exp\left\{-p\int_{S_p}\max\left(\frac{w}{x_1^*},\frac{1-w}{x_2^*}\right)H\left(\partial w\right)\right\}$$
$$=\exp\left\{-p\left(\int_0^{0,11}\frac{1-w}{x_2^*}H\left(\partial w\right)+\int_{0,11}^{1}\frac{w}{x_1^*}H\left(\partial w\right)\right)\right\}$$
$$=\frac{\pi_1}{x_2^*}\frac{\Gamma\left(a_1+b_1\right)}{\Gamma\left(a_1\right)\Gamma\left(b_1\right)}\frac{\Gamma\left(a_1\right)\Gamma\left(b_1+1\right)}{\Gamma\left(a_1+b_1+1\right)}\int_0^{0,1688}w^{a_1-1}\left(1-w\right)^{b_1+1-1}\partial w+$$
$$\frac{\pi_2}{x_2^*}\frac{\Gamma\left(a_2+b_2\right)}{\Gamma\left(a_2\right)\Gamma\left(b_2\right)}\frac{\Gamma\left(a_2\right)\Gamma\left(b_2+1\right)}{\Gamma\left(a_2+b_2+1\right)}\int_0^{0,1688}w^{a_2-1}\left(1-w\right)^{b_2+1-1}\partial w$$
$$\frac{\pi_1}{x_1^*}\frac{\Gamma\left(a_1+b_1\right)}{\Gamma\left(a_1\right)\Gamma\left(b_1\right)}\frac{\Gamma\left(a_1+1\right)\Gamma\left(b_1\right)}{\Gamma\left(a_1+1+b_1\right)}\int_{0,1688}^{1}w^{a_1+1-1}\left(1-w\right)^{b_1-1}\partial w+$$
$$\frac{\pi_2}{x_1^*}\frac{\Gamma\left(a_2+b_2\right)}{\Gamma\left(a_2\right)\Gamma\left(b_2\right)}\frac{\Gamma\left(a_2+1\right)\Gamma\left(b_2\right)}{\Gamma\left(a_2+1+b_2\right)}\int_{0,1688}^{1}w^{a_2+1-1}\left(1-w\right)^{b_2-1}\partial w$$

(5.77)

And the joint tail probability is

$$\begin{aligned} P(X_1 > 3000, X_2 > 11000) &= 1 - P(X_1 < 3000) - P(X_2 < 11000) + P(X_1 < 3000, X_2 < 11000) \\ &= 1 - 0,8638 - 0,9707 + 0,8617 \\ &= 0,0272 \end{aligned}$$

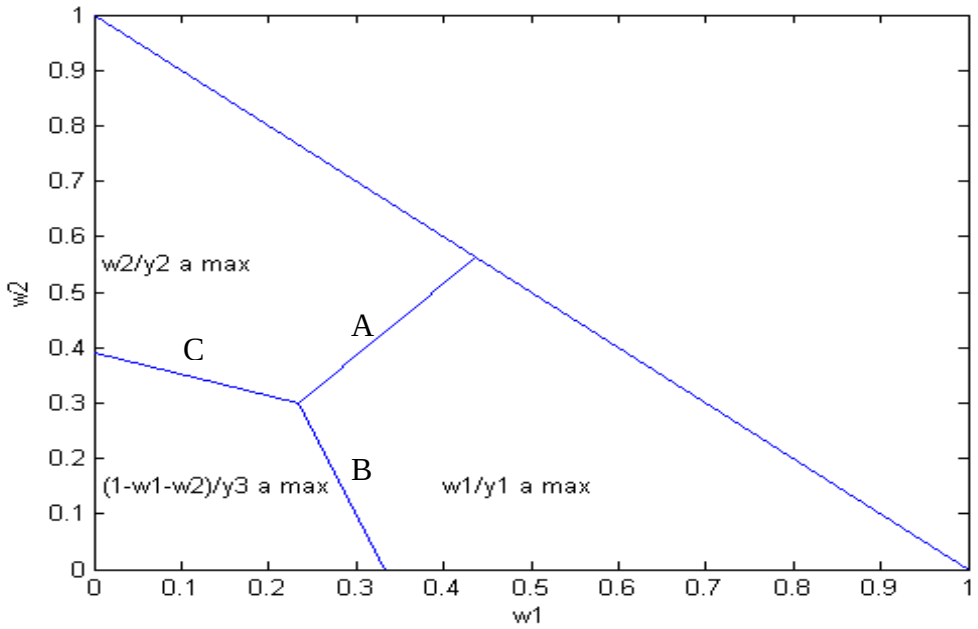
This is very close to the empirical tail probability of 0,0344.

5.7.4 A multidimensional application

Now the 2 dimensional case is expanded to the three dimensional case, where *p* = 3. The aim is to model $X = (X_1, X_2, X_3)$ where X_1 represents the maximum inflow of January, X_2 represents the maximum inflow of February and X_3 represents the maximum inflow of March, over a period of 29 years.

The three dimensional sample space is shown in Figure 5.11. The two dimensional sample space is divided into three regions defined by $\max\left(\frac{w_1}{y_1}, \frac{w_2}{y_2}, \frac{w_3}{y_3}\right)$ in (5.65). The sample space is two dimensional because w_3 is expressed in terms of w_1 and w_2; $w_3 = 1 - w_1 - w_2$.

Fig. 5.11 The 3 dimensional sample space of the pseudo angles w



$$\text{A: } w_2 = \frac{y_2}{y_1} w_1 \qquad \text{B: } w_2 = 1 - \left(1 + \frac{y_3}{y_1}\right) w_1 \qquad \text{C: } w_2 = \frac{1 - w_1}{y_3} \left(\frac{1}{y_3} + \frac{1}{y_2}\right)^{-1}$$

Again it is assumed that the marginals are GEV distributed (5.71). The estimated marginal parameter values are given in Table 5.9. This is an extract from Table 5.1.

Table 5.9 The estimated parameter values

	$\hat{\gamma}$	**90% HPD**	$\hat{\sigma}$	**90% HPD**	$\hat{\mu}$	**90% HPD**
Jan	0,752	[0.6843; 0.7969]	580,1564	[580,0751; 580,2489]	499,0187	[498,9725; 499,0737]
Feb	0,7568	[0.6770; 0.8477]	598,1504	[598,0532; 598,2133]	477,9782	[477,9321; 478,1106]
Mar	0,7176	[0,6615; 0,7911]	576,9336	[576,9014; 576,9902]	469,4655	[469,4160; 469,5302]

Again the marginals undergo a Fréchet transformation (5.60). The newly transformed variables are denoted by (Y_1, Y_2, Y_3).

An appropriate value for *k* must be chosen. For k = 2, 3 and 4 the parameters are estimated and for each mixture the BIC is calculated. The smallest BIC is obtained when *k* = 2 and has a value of 8,8273. The estimates of the parameters of the mixture of the 2 Dirichlet distributions are obtained through the Metropolis Hastings algorithm assuming uniform priors on the parameters π and $a_j^{(m)}, m = 1, 2,\ j = 1, 2, 3$. The estimates are given in Table 5.10.

Table 5.10 Estimated parameters of the mixture of the 2 Dirichlet distributions

$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{a}_1^{(1)}$	$\hat{a}_2^{(1)}$	$\hat{a}_3^{(1)}$	$\hat{a}_1^{(2)}$	$\hat{a}_2^{(2)}$	$\hat{a}_3^{(2)}$
0,6434	0,3566	10,2439	10,9804	7,1485	14,4497	12,0608	24,5730

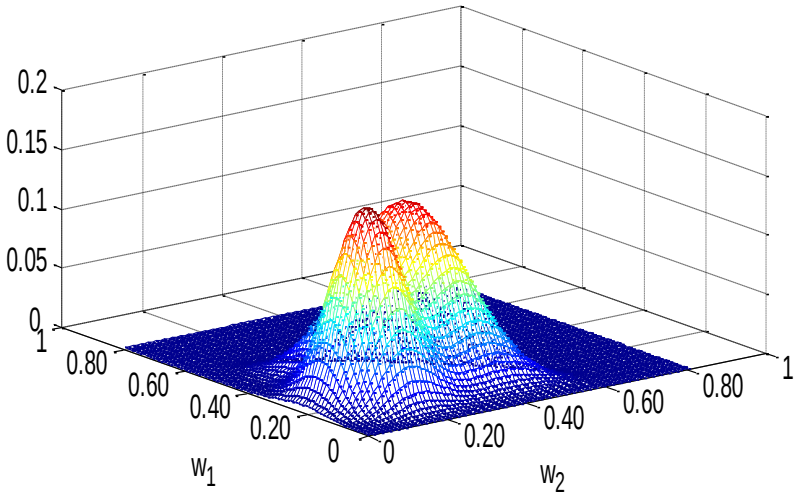
When imposing the constraints $v_m = \sum_{j=1}^{12} a_j^{(m)}, m = 1, 2$ the estimates of $\hat{v}_m$ are obtained, and from there the estimates of $\hat{\mu}_j^{(m)}$ are calculated. The estimates are shown in Table 5.11.

Table 5.11 Estimated parameters of the mixture of the 2 Dirichlet distributions

$\hat{v}_1$	$\hat{v}_2$	$\hat{\mu}_1^{(1)}$	$\hat{\mu}_2^{(1)}$	$\hat{\mu}_3^{(1)}$	$\hat{\mu}_1^{(2)}$	$\hat{\mu}_2^{(2)}$	$\hat{\mu}_3^{(2)}$
28,3727	51,0834	0,3610	0,3870	0,2519	0,2829	0,2361	0,4810

Figure 5.12 shows the mesh of the spectral density $h(w)$, in the 3 dimensional case, with a mixture of 2 Dirichlet distributions.

Fig. 5.12 The mesh of the spectral density $h(w)$



As in the two dimensional case the mixture distribution is used to calculate probabilities. The following probability is calculated:

$$P\left(X_1 < 3000, X_2 < 11000, X_3 < 6000\right). \tag{5.78}$$

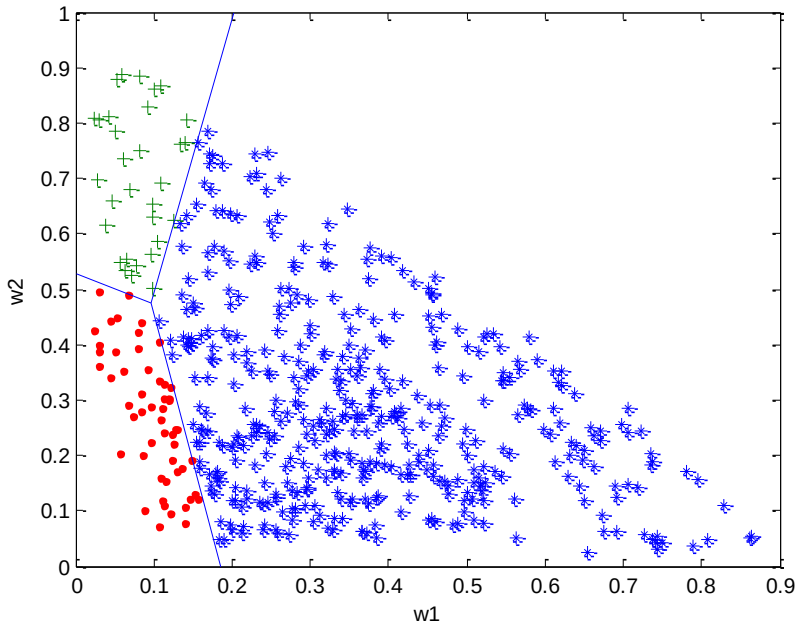
As before, these variables first undergo Fréchet transformations, as was done in the two dimensional case. The Fréchet transformation for X_3 is shown below, while the transformations for the other two variables were shown in (5.75) and (5.76),

$$\begin{aligned} F\left(X_3\right) &= \exp\left\{-\left[1+\frac{0,7176\left(6000-469,4655\right)}{576,9336}\right]^{\frac{-1}{0,7176}}\right\} \\ &= 0,9674; \\ x_3^* &= \frac{-1}{\log\left(F\left(X_3\right)\right)} = 30,1721. \end{aligned} \tag{5.79}$$

Calculating the probability $P\left(Y_1 < 6,8299, Y_2 < 33,6272, Y_3 < 30,1721\right)$ by using integration, as was done in (5.77), becomes difficult in the multidimensional case. An easier method is to simulate from a mixture of two Dirichlet distributions. 500 pseudo angles are simulated from the mixture of two

Dirichlet distributions. Figure 5.13 shows the number of simulated angles on the sample space that falls into the three different regions.

Fig. 5.13 The simulated angle values on the sample space



Dividing the proportion of data in each region relative to the corresponding $y_i, i = 1,..,3$ value by 500 and adding them together gives an estimate of $\frac{-\log\left(G(y)\right)}{3} = 0{,}1054$. From here the $P\left(Y_1 < 6{,}8299, Y_2 < 33{,}6272, Y_3 < 30{,}1721\right)$ is calculated as 0,7500.

According to Boldi and Davison (2007) the quantity ξ is interpreted as a measure of the asymptotic dependence among the variables. The following equation define ξ:

$$\xi = \int_{S_p} \min\left(w_1,...,w_p\right) H\left(dw\right), 0 \le \xi \le \frac{1}{2} \qquad (5.80)$$

(Boldi and Davison 2007, p. 221). $\xi = 0$ indicates complete independence and $\xi = \frac{1}{2}$ indicates complete dependence. In the three dimensional case ξ is calculated as 0,0959 which means that the variables are 19,18%, $\left\{\left(\frac{0,0959}{0,5}\right)\times 100\right\}$, dependent on each other in the tails. Therefore, if a high inflow value occurs in January the chance is 19,18% that high inflows will also occur in February and March.

5.8 Conclusion

This chapter introduces extreme value copulas, specifically the Gumbel and the Tawn copula, for modelling multivariate extreme values. A Bayesian approach is mainly used for estimating the parameters. The application of extreme value copulas is an appropriate method for modelling multivariate extreme values due to the small estimated joint tail probabilities (Section 5.6) and the simulated posterior distributions shown in Figures 5.3 and 5.5. This chapter also introduces the Dirichlet mixture model as another method for modelling multivariate extreme value. Again the parameter estimation is based on a Bayesian approach. The BIC value is calculated to choose the optimum number of Dirichlet distributions in the mixture. The Dirichlet mixture model is also an appropriate method for modelling multivariate extreme values due to the small estimated tail probabilities.

Chapter 6

Modelling of ordered multivariate extreme values

6.1 Introduction

Over recent years there has been some major break through in extreme value theory, especially in multivariate extreme value theory. Multivariate extreme value theory is very useful in environmental problems. In this chapter an important characteristic of multivariate extreme value theory is discussed: the ordering of extreme variables.

If $\{X_i(d)\}$ denotes water inflow aggregated over *d* days for time period *i* and if *d*' is a multiple of *d* then the following restriction must hold:

$$\max_i \{X_i(d)\} \leq \max_i \{X_i(d')\} \leq \frac{d'}{d} \max_i \{X_i(d)\} \qquad (6.1)$$

(Nadarajah *et al*. 1998, p. 474). (6.1) is an ordered restriction and is necessary in some multivariate extreme value models.

Section 6.2 discusses the selection of threshold values. Section 6.3 discusses the estimation of parameter values conditional on the ordered restrictions. Section 6.4 shows a practical application of ordered bivariate modelling. Section 6.5 shows a simulation study to investigate whether the ordering restrictions have an influence on the estimation of the parameters. A conclusion is given in Section 6.6.

6.2 Selecting thresholds

In this chapter only bivariate extreme values are considered, but the discussions can be extended to multivariate extremes. Again, the analysis of bivariate extremes lies in the tail of the bivariate distribution. Therefore, the

tail of the bivariate distribution is modelled. For the selection of the threshold, univariate threshold methods are first considered. Let $(X_1,...,X_n)$ be a univariate sequence of i.i.d. observations from some unknown distribution *F*. The estimation of the tail of a distribution consists of modelling the data points that exceeds a threshold *t*. If it is assumed that *F* is in the domain of attraction of an extreme value distribution (refer to Chapter 1) then the distribution of excesses over the threshold is modelled by using an extreme value distribution such as a Generalized Pareto distribution (GPD) or a Strict Pareto distribution. The GPD and the Strict Pareto distribution are given respectively by the following equations:

$$F_x(x \mid t_x) \approx 1-\left(1+\gamma\frac{x-t}{\sigma}\right)^{\frac{-1}{\gamma}} = G(x)$$

$$F_x(x \mid t_x) \approx 1-\left(\frac{x}{t_x}\right)^{\frac{-1}{\gamma}} = G(x). \qquad (6.2)$$

If the approximate distribution in (6.2) is taken to be exact for $x > t$, then the estimation of the tail of *F* leads to the estimation of the parameters of *G*. In the case of the GPD there are two parameters γ and σ, and in the case of the Strict Pareto distribution there is only one parameter, γ, the EVI.

Bivariate threshold methods, that estimate the joint tail of the distribution, are now considered.

Let $\{(X_{i1}, X_{i2}); i = 1,...,n\}$ be a sequence of i.i.d. observations from an unknown joint distribution function *F*. Again it is assumed that *F* is in the domain of attraction of a bivariate extreme value distribution (refer to Chapter 1). Let G_* be an approximate distribution for *F* where G_* is given by the equation

$$G_*(y_1, y_2) = \exp\left\{-\int_{[0,1]} \max\left(\frac{\omega}{y_1}, \frac{1-\omega}{y_2}\right) S(\partial\omega)\right\} \qquad (6.3)$$

(Nadarajah *et al*. 1998, p. 476).

G_* is a bivariate extreme value distribution with y_1 and y_2 as unit Fréchet marginals. S is a positive dependence measure on [0, 1] satisfying the following equation:

$$\int_{[0,1]} \omega S(\partial\omega) = 1 = \int_{[0,1]} (1-\omega) S(\partial\omega) \tag{6.4}$$

(Nadarajah *et al*. 1998, p. 476).

$$\begin{aligned} F(x_1, x_2) &\approx G_*\left(U_1(x_1), U_2(x_2)\right) \\ &= G(x_1, x_2) \end{aligned} \tag{6.5}$$

is now the approximated joint tail model of F which is assumed to be exact for $(x_1, x_2) \in (t_1, \infty) \times (t_2, \infty)$, where t_1 and t_2 are the thresholds selected in the univariate case (Nadarajah *et al*. 1998, p. 476).

The Logistic copula is a parametric model that satisfies (6.4) (Tawn 1988). The Logistic copula is given by the equation

$$G(u,v) = \exp\left\{-\left(u^{\phi} + v^{\phi}\right)^{\frac{1}{\phi}}\right\}, u, v > 0, \phi \geq 1 \tag{6.6}$$

where ϕ is a dependence parameter (Nadarajah *et al*. 1998, p. 477). $\phi = 1$ indicates independence. Estimation of the joint tail of F is reduced to estimating the marginal tail parameters $\boldsymbol{\theta}_j, j = 1, 2$ and the dependence parameter ϕ from the Logistic copula.

Nadarajah *et al.* (1998, p. 477) suggests to first transform the marginal variables to unit Fréchet variables by using the transformation

$$Y_j = \frac{-1}{\log\left\{F_j\left(X_j\right)\right\}}, j = 1,2, \tag{6.7}$$

where F_j is the marginal distribution function of X_j.

6.3 Parameter estimation

When the bivariate extreme value distribution

$$P\left(Y_1 \le y_1, Y_2 \le y_2\right) = G\left(y_1, y_2\right), \tag{6.8}$$

where Y_1 and Y_2 are Fréchet transformed variables above the respective thresholds t_1 and t_2, is used to model the joint distribution of $\left(X_1, X_2\right)$ the following order relationship must hold:

$$X_1 \le X_2 \le mX_1 \tag{6.9}$$

where $m \ge 1$ (Nadarajah *et al*. 1998, p. 478). (6.9) follow from (6.1). This imposes some restrictions on the estimation of the parameters. In this study a Bayesian approach is considered for estimating the parameters. Due to the ordered restriction on the data an *a priori* restriction on the parameter is stated before the estimation of the parameter takes place. In the case where the marginal distributions are assumed to be Strict Pareto distributed, $\gamma_{x_1} \ge \gamma_{x_2}$ is known *a priori*. This follows from the fact that $X_1 \le X_2$. Let X_1 represent individual elements and X_2 the aggregate of more than one element, then the variance of the aggregates is smaller than the variance of the individual elements. Therefore, it is expected that the EVI of the aggregate (γ_{x_2}) is smaller than the EVI of the individual elements (γ_{x_1}). This result was tested by simulating from several Pareto distributions, such as the Fréchet and t-distributions. The value of γ decreased as the aggregated block size increased. In the case where the marginal distributions are assumed to be

GPD distributed, $\frac{\sigma_{x_2}}{1-\gamma_{x_2}} > \frac{\sigma_{x_1}}{1-\gamma_{x_1}}$ is known *a priori*, where $\frac{\sigma_{x_1}}{1-\gamma_{x_1}}$ is the expected value of X_1 and $\frac{\sigma_{x_2}}{1-\gamma_{x_2}}$ is the expected value of X_2. This follows from the fact that $X_1 \le X_2$; therefore the $E(X_1) \le E(X_2)$.

6.4 Practical application

The data set considered here is the daily inflow of the water into the Gariep Dam, measured in m^3/s. It covers the period from January 01, 1976 until June 30, 2005, excluding 1980 and 1982 due to large data losses. The variable *X* is considered as the annual one day maximum and the variable *Y* is considered as the seven day aggregate annual maximum. The data are now divided into blocks (groups) of size 20 (20 days in a block). *X* is now the maximum of each block. *Y* is constructed by aggregating the consecutive seven data values in each block and then obtaining the maximum aggregate value for each block. Here a block size of 20 is chosen because the maximum amount of days that a flood will last is more or less 20 days.

The main purpose of this application is to consider the joint distribution of *X* and *Y* through the Logistic copula (Tawn 1988) as was done by Nadarajah *et al*. (1998).

Recall that the distribution function of the Logistic copula (6.6) is

$$G(u,v) = \exp\left\{-\left(u^{\phi} + v^{\phi}\right)^{\frac{1}{\phi}}\right\}, u,v > 0, \phi > 1 \qquad (6.10)$$

where the random variables *U* and *V* are the following Fréchet marginals:

$$U = \frac{-1}{\log\left(F_X\left(x \mid t_x\right)\right)}, V = \frac{-1}{\log\left(F_Y\left(y \mid t_y\right)\right)} \qquad (6.11)$$

(Nadarajah *et al*. 1998, p. 477).

$F_x\left(x \mid t_x\right)$ and $F_Y\left(y \mid t_y\right)$ are the marginal distributions of *X* and *Y* bounded by thresholds t_x and t_y respectively. Although the data are block maxima the marginal distributions are considered to be Strict Pareto distributed for illustration purposes. The marginal distributions are given by:

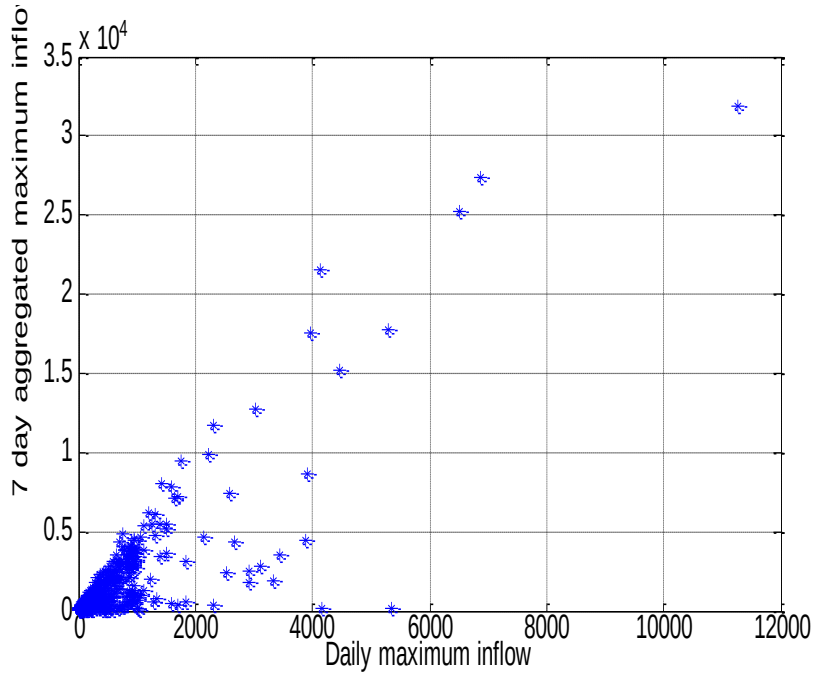
$$F_x\left(x \mid t_x\right) = 1 - \left(\frac{x}{t_x}\right)^{\frac{-1}{\gamma_x}}, x > t_x,$$
$$F_y\left(x \mid t_y\right) = 1 - \left(\frac{y}{t_y}\right)^{\frac{-1}{\gamma_y}}, y > t_y. \qquad (6.12)$$

γ_x and γ_y are the extreme value indices (EVI) for the daily maximum and the seven day aggregate maximum respectively.

$\gamma_x \geq \gamma_y$ is stated *a priori*, as discussed in Section 6.3. In this application γ_x and γ_y are estimated jointly.

Figure 6.1 shows the one day annual maxima plotted against the seven day aggregate annual maxima.

Fig 6.1 **One day annual maxima vs seven day aggregate annual maxima**



As mentioned before, ϕ in (6.10) is called the dependence parameter of the Logistic copula and $\phi = 1$ indicates independence. ϕ can be estimated by calculating the mode of the posterior distribution of ϕ. The MDI prior of ϕ is

$$\pi(\phi) \propto \exp\left\{E \log\left[g(U,V)\right]\right\} \qquad (6.13)$$

(De Waal *et al*. 2007, p. 624).

The joint density of *U* and *V* is derived as follows:

$$G(u,v) = \exp\left\{-\left(u^{-\phi} + v^{-\phi}\right)^{\frac{1}{\phi}}\right\}.$$

Therefore

$$\frac{\partial^2 G}{\partial u \partial v} = \frac{\partial}{\partial v}\left[G(u,v)\left(u^{-\phi}+v^{-\phi}\right)^{\frac{1}{\phi}-1} u^{-\phi-1} \right]$$

$$= G(u,v)\left(\frac{1}{\phi}-1\right)\left(u^{-\phi}+v^{-\phi}\right)^{\frac{1}{\phi}-2}(-\phi)v^{-\phi-1}u^{-\phi-1} +$$

$$G(u,v)\left(u^{-\phi}+v^{-\phi}\right)^{\frac{1}{\phi}-1} v^{-\phi-1}\left(u^{-\phi}+v^{-\phi}\right)^{\frac{1}{\phi}-1} u^{-\phi-1}$$

$$= G(u,v)\left(\frac{1}{\phi}-1\right)\left(u^{-\phi}+v^{-\phi}\right)^{\frac{1}{\phi}-2}(-\phi)v^{-\phi-1}u^{-\phi-1} + G(u,v)\left(u^{-\phi}+v^{-\phi}\right)^{\frac{2}{\phi}-2} v^{-\phi-1}u^{-\phi-1}$$

$$= G(u,v)\left(u^{-\phi}+v^{-\phi}\right)^{\frac{1}{\phi}-2} v^{-\phi-1}u^{-\phi-1}\left[\left(u^{-\phi}+v^{-\phi}\right)^{\frac{1}{\phi}} + (\phi-1)\right].$$

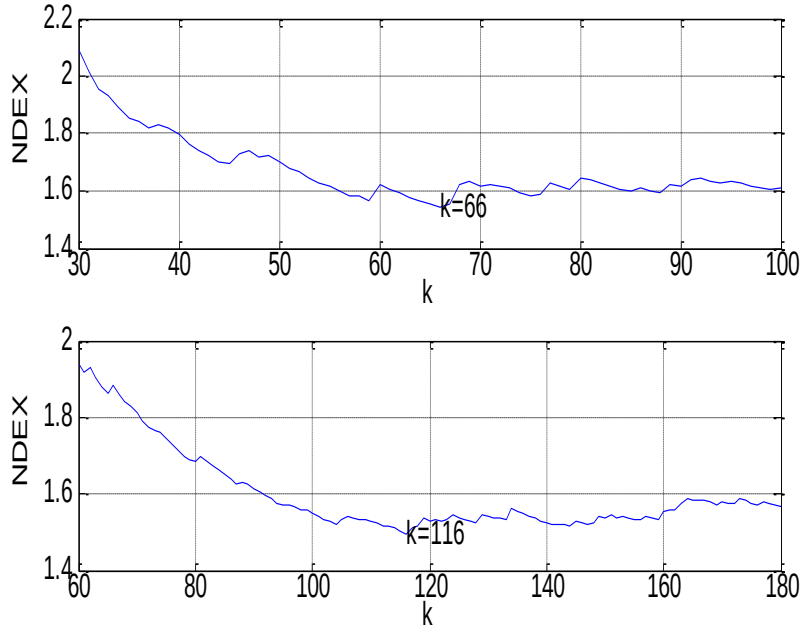
(6.14)

The sample space is discretised at m^2 points $\left(u_i, v_j\right)$ such that $u_i \in (0,7)$ and $v_j \in (0,7), i = 1,...,m,\ j = 1,...,m$ for large *m*. When $E\left\{\log\left[g(U,V)\right]\right\} = \sum_{i=1}^{m}\sum_{j=1}^{m} \log g\left(u_i, v_j\right) g\left(u_i, v_j\right)$ is calculated the resulting prior for ϕ is close to a uniform distribution (De Waal *et al*. 2007, p. 624). In this study the prior on ϕ is assumed to be uniform.

6.4.1 Choosing a threshold

As previously mentioned, the 1-day annual maximum (*X*) and the aggregated 7-day annual maximum (*Y*) are Strict Pareto distributed, as shown in (6.12). Choosing a threshold by using the entropy of the Dirichlet distribution was discussed in Chapter 4, Section 4.3.4. When applying this method to the 1-day annual maximum (*X*) and the aggregated 7-day annual maximum (*Y*) the following thresholds are obtained: $t_x = 938{,}19$ at *k* = 66 and $t_y = 2906{,}4$ at *k* = 116. This is shown in Figure 6.2.

Fig 6.2 The estimated negative differential entropies on the 1-day annual maxima and the 7-day annual maxima



6.4.2 Fitting the Logistic copula

The two parameters of the Strict Pareto marginals γ_x and γ_y and the dependence parameter ϕ of the Logistic copula are estimated jointly by using MCMC and the Metropolis Hastings algorithm.

The joint likelihood of the observations $(x_i, y_i), i = 1,...,k_p$ that exceed the thresholds t_x and t_y jointly is

$$\text{lik}\left(\gamma_x, \gamma_y, \phi \mid x_i, y_i, i = 1,...,k\right) = \prod G\left(u_i, v_i\right) u_i^{-\phi-1} v_i^{-\phi-1} \left(u_i^{-\phi} + v_i^{-\phi}\right)^{\frac{1}{\phi}-2} \left(u_i^{-\phi} + v_i^{-\phi}\right)^{\frac{1}{\phi}} + (\phi - 1)\left(u_i^{-\phi} + v_i^{-\phi}\right) \tag{6.15}$$

u_i and v_i are calculated using (6.11).

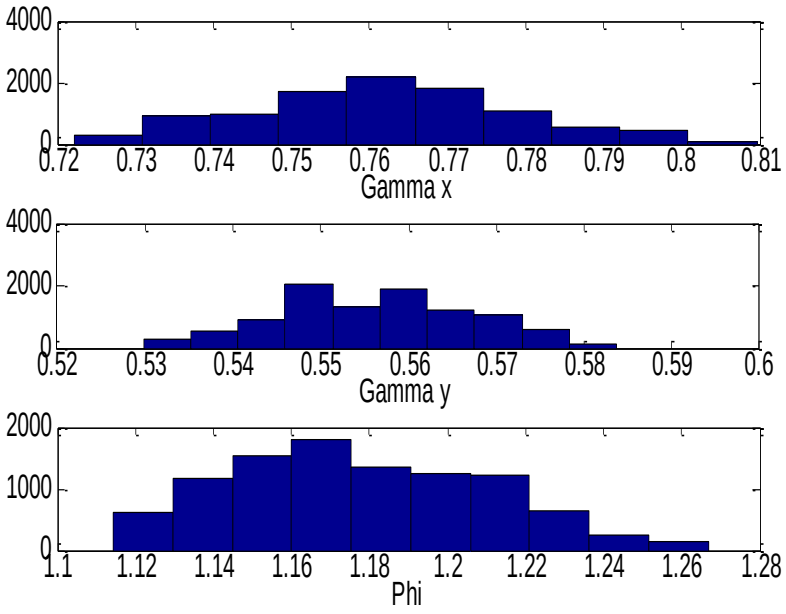
The joint prior is

$$\pi(\gamma_x, \gamma_y, \phi) \propto (\gamma_x \gamma_y)^{-1} e^{-(\gamma_x + \gamma_y)} \pi(\phi), \gamma_x > \gamma_y . \quad (6.16)$$

This prior is the product of the MDI priors on γ_x and γ_y and the uniform prior on ϕ, where $\gamma_x^{-1} e^{-\gamma_x}$ and $\gamma_y^{-1} e^{-\gamma_y}$ are the MDI priors on γ_x and γ_y respectively for the Strict Pareto in (6.12) (De Waal *et al*. 2007, p. 628). The uniform prior on ϕ is proportional to 1. The parameters are estimated under the ordering restriction $\gamma_x > \gamma_y$.

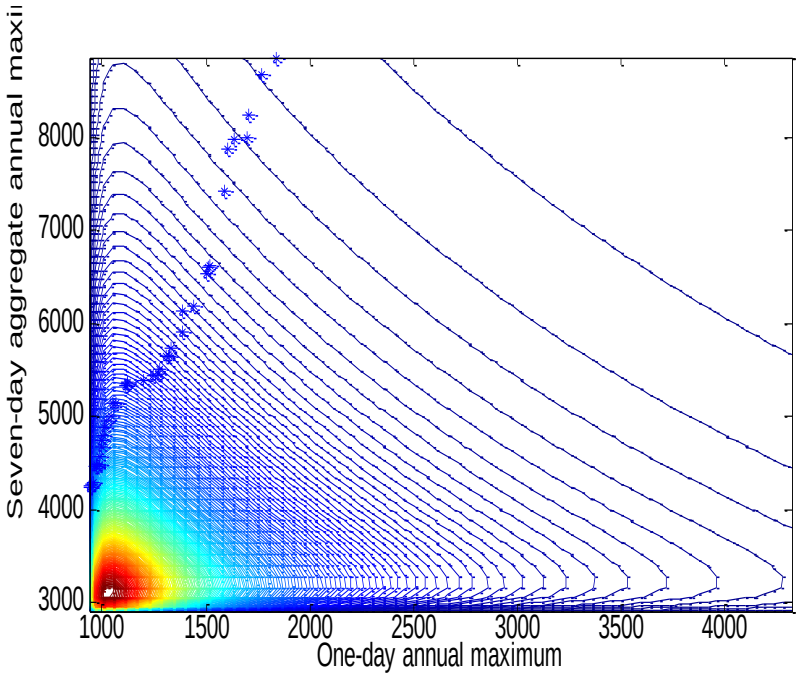
$\gamma_x = 0,6,\ \gamma_y = 0,5$ and $\phi = 2$ are used as starting values in the Metropolis Hastings algorithm. Figure 6.3 shows the simulated posterior distribution after 10 000 iterations. The mean of each simulated posterior is taken as the estimate of its associated parameter. The estimated parameter values are $\hat{\gamma}_x = 0,7618,\ \hat{\gamma}_y = 0,5562$ and $\hat{\phi} = 1,1773$.

Fig. 6.3 Simulated posterior distributions for the estimated parameters



A contour of the Logistic copula in terms of the original data that exceeds the thresholds is shown in Figure 6.4.

Fig. 6.4 Contours of the density with original data above both thresholds indicated by *



From Figure 6.4 the model seems to be a good fit to the data above the thresholds.

6.4.3 Calculating tail probabilities

Joint tail probabilities such as $P(X > 11000, Y > 30000)$ are calculated as follows:

$$\begin{aligned} P(X > 11000, Y > 30000) &= P(U > 24,8131; V > 65,9881) \\ &= 1 + G(24,8131; 65,9881) - G(24,8131; \infty) - G(\infty; 65,9881) \\ &= 1 + 0,9504 - 0,9605 - 0,9850 \\ &= 0,0049 \end{aligned} \tag{6.17}$$

where U and V are Fréchet transformed variables,

$G(24,8131; \infty) = \exp(-24,8131^{-1})$ and $G(\infty; 65,9881) = \exp(-65,9881^{-1})$.

The empirical probability is calculated from the actual data set. $X > 11000$ and $Y > 3000$ applies to only one observation in the data set that is divided into block of 20 days. The empirical probability is 0,002.

6.5 Simulated application

This Chapter focuses on the estimation of parameters under the ordered restriction (6.1); an *a priori* restriction on the parameters can be stated, as was shown in Section 6.3. In Section 6.4 a practical application was done where the marginal distributions were assumed to be Strict Pareto distributed and due to the ordering restriction, the parameter restriction $\gamma_x > \gamma_y$ was imposed.

In this section the question of whether the ordering restrictions on the parameters have any influence on the estimation of the parameters is investigated. For this investigation two marginal distributions are simulated from a GPD distribution. First the parameters are estimated separately without the ordered restriction on the parameters and then the parameters are estimated jointly where the ordered restriction is taken into account.

6.5.1 Simulation of the marginal distributions

Let *X* and *Y* be two variables where $X \sim \text{GPD}(\sigma_x, \gamma_x)$ and $Y \sim \text{GPD}(\sigma_y, \gamma_y)$, where *X* > *Y*. The following are the marginal distribution functions:

$$F(y) = 1 - \left(1 + \frac{\gamma_y y}{\sigma_y}\right)^{\frac{-1}{\gamma_y}}, y > 0;$$

$$F(x) = 1 - \left(1 + \frac{\gamma_x (x-y)}{\sigma_x}\right)^{\frac{-1}{\gamma_x}}, x > y. \qquad (6.18)$$

1000 values are simulated by using the quantile functions of the two GPD marginal distributions:

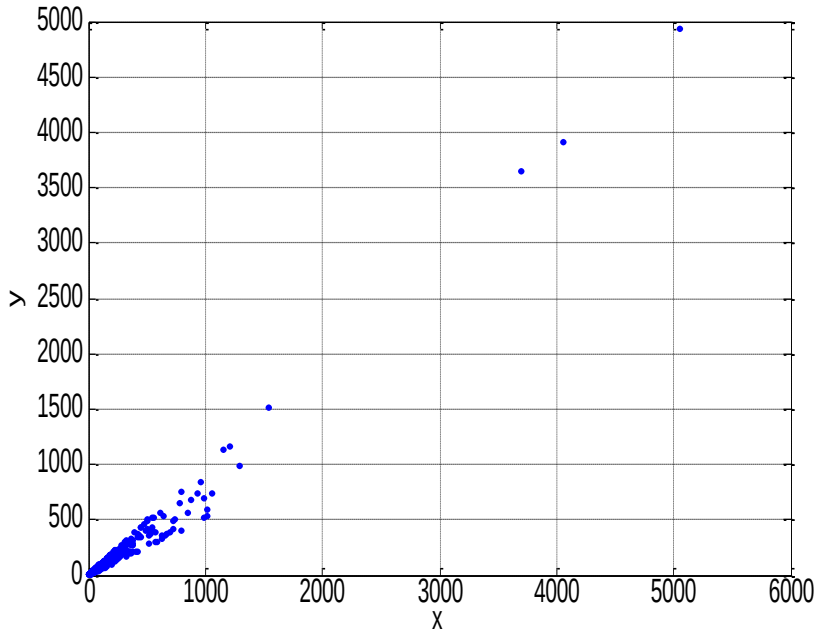
$$Q(p)=\frac{\sigma_y}{\gamma_y}\left((1-p)^{-\gamma_y}-1\right), p\in(0,1)$$

$$Q(p)=\frac{\sigma_x}{\gamma_x}\left((1-p)^{-\gamma_x}-1\right)+y, p\in(0,1) \quad (6.19)$$

with $\gamma_x=0,7;\sigma_x=100;\gamma_y=0,5$ and $\sigma_y=50$.

Only simulated values that satisfy the constraint $y<x<2y$ are considered. Figure 6.5 shows the simulated values that satisfy this constraint.

Fig. 6.5 The simulated values from the GPD distributions



6.5.2 Estimation of parameters

First the parameters are estimated separately for each marginal distribution. A Bayesian approach together with the Metropolis Hastings algorithm is used for estimation. MDI priors are assumed where the log of the priors are

$$\log \pi\left(\sigma_x, \gamma_x\right) \propto \log\left(\frac{1}{\sigma_x}\right) - \gamma_x;$$
$$\log \pi\left(\sigma_y, \gamma_y\right) \propto \log\left(\frac{1}{\sigma_y}\right) - \gamma_y \qquad (6.20)$$

(Beirlant *et al*. 2004, p. 447).

The log-likelihood is given as

$$l\log(l) = -N_t \log \sigma_x - \left(1 + \frac{1}{\gamma_x}\right)\sum\left(\log\left(1 + \frac{\gamma_x x}{\sigma_x}\right)\right);$$
$$l\log(l) = -N_t \log \sigma_y - \left(1 + \frac{1}{\gamma_y}\right)\sum\left(\log\left(1 + \frac{\gamma_y y}{\sigma_y}\right)\right) \qquad (6.21)$$

(Beirlant *et al*. 2004, p. 149).

The log of the posterior distribution is proportional to the sum of the log of the prior distribution and the log of the likelihood function.

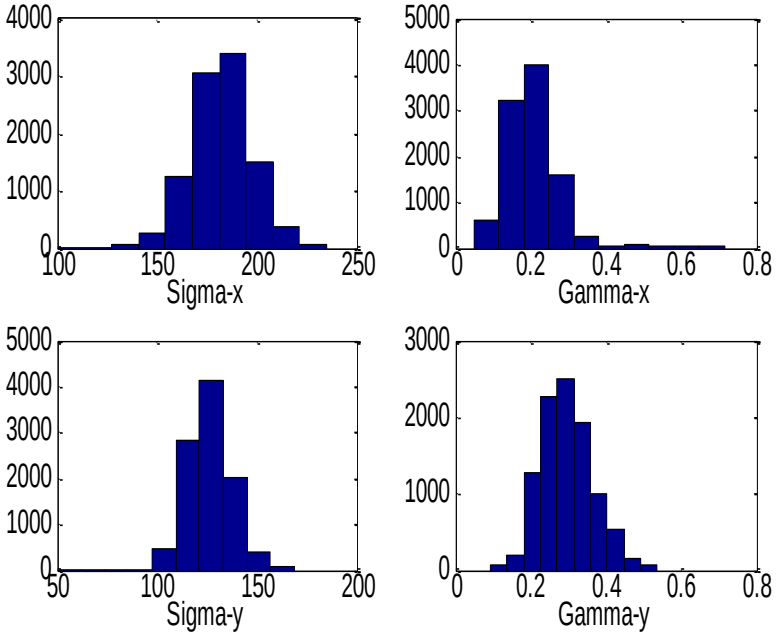
Table 6.1 gives the estimated parameter values after 10 000 iterations.

Table 6.1 Separately estimated parameter values

γ_x	σ_x	γ_y	σ_y
0,2104	182,1279	0,2961	126,0299

The simulated posterior distributions of the estimated parameters are shown in Figure 6.6.

Fig. 6.6 Simulated posteriors distribution for the separately estimated parameters



Now the parameter values are estimated jointly. When the parameters are estimated jointly the same simulated data set is used but now the following ordered restriction is placed on the parameters:

$$\frac{\sigma_x}{1-\gamma_x} > \frac{\sigma_y}{1-\gamma_y}, \tag{6.22}$$

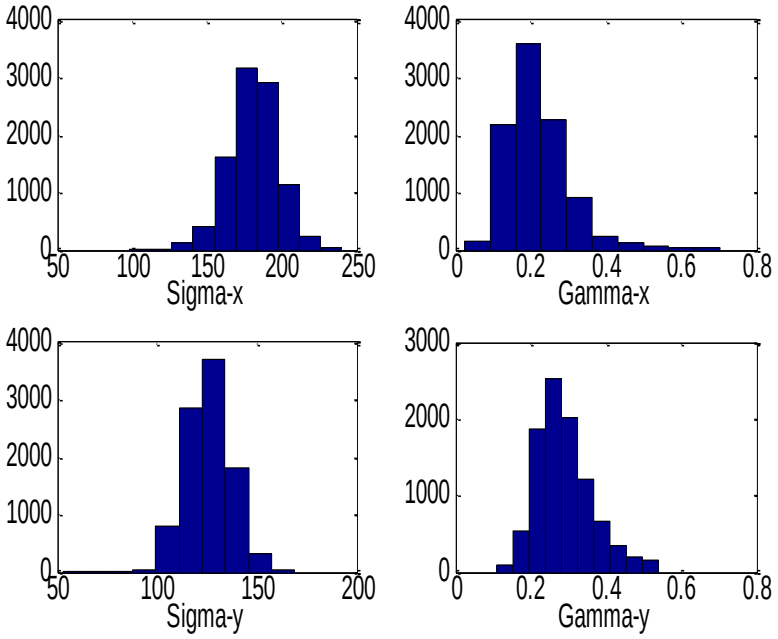
where $\frac{\sigma_x}{1-\gamma_x}$ is the expected value of *X* and $\frac{\sigma_y}{1-\gamma_y}$ is the expected value of *Y*. Because $X > Y$ it follows that $E(X) > E(Y)$. The parameters are estimated through the Metropolis Hastings algorithm under constraint (6.22). Table 6.2 gives the jointly estimated parameter values after 10 000 iterations.

Table 6.2 Jointly estimated parameter values

γ_x	σ_x	γ_y	σ_y
0,2201	180,8774	0,2883	125,8001

The simulated posterior distributions of the jointly estimated parameters are shown in Figure 6.7.

Fig. 6.7 Simulated posterior distributions for the jointly estimated parameters



The separately estimated parameter values and the jointly estimated parameter values, that take the ordering restriction into account, are very similar to each other. Therefore, for this simulation study, it seems that the ordering restrictions on the parameters have very little or no influence on the estimation of the parameters.

Now, for further investigation, 50 data sets are simulated with $\gamma_x = 0,7; \sigma_x = 100; \gamma_y = 0,5$ and $\sigma_y = 50$. For each data set the parameters are estimated separately and jointly.

Figures 6.8 - 6.11 show the separately estimated parameter values together with the jointly estimated parameter values.

Fig. 6.8 γ_x **estimated separately (blue line) together with** γ_x **estimated jointly (red dashed line)**

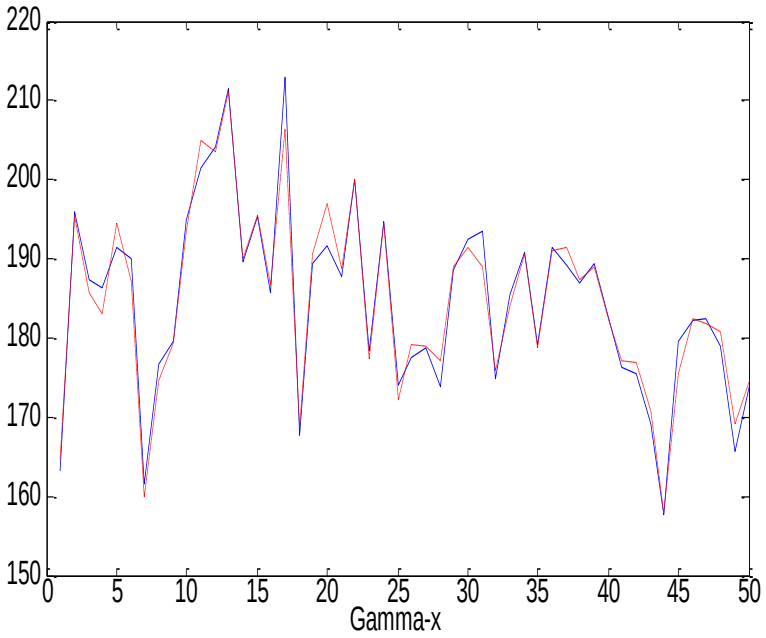


Fig. 6.9 σ_x **estimated separately (blue line) together with** σ_x **estimated jointly (red dashed line)**

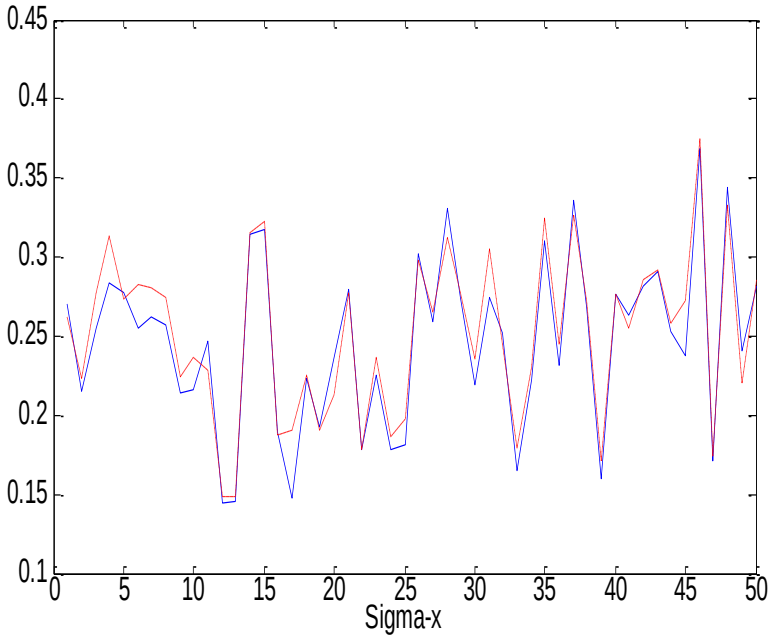


Fig. 6.10 γ_y **estimated separately (blue line) together with** γ_y **estimated jointly (red dashed line)**

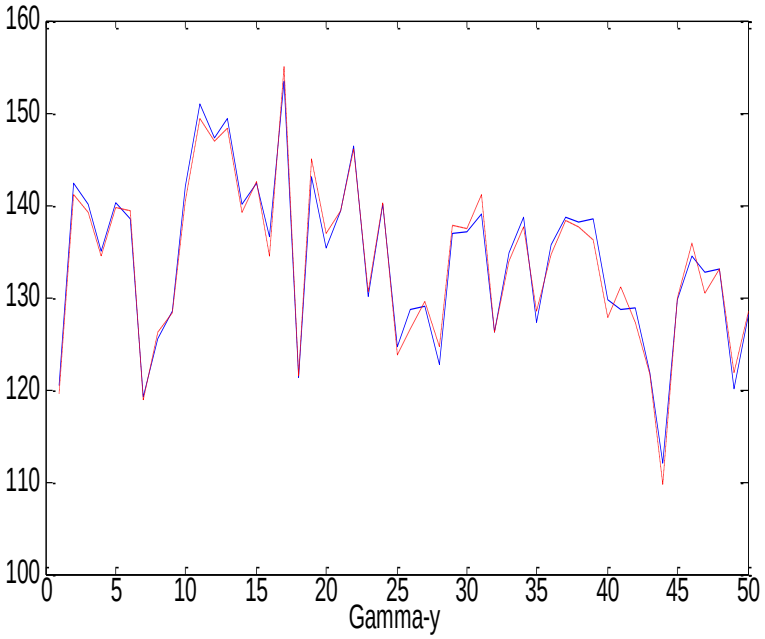
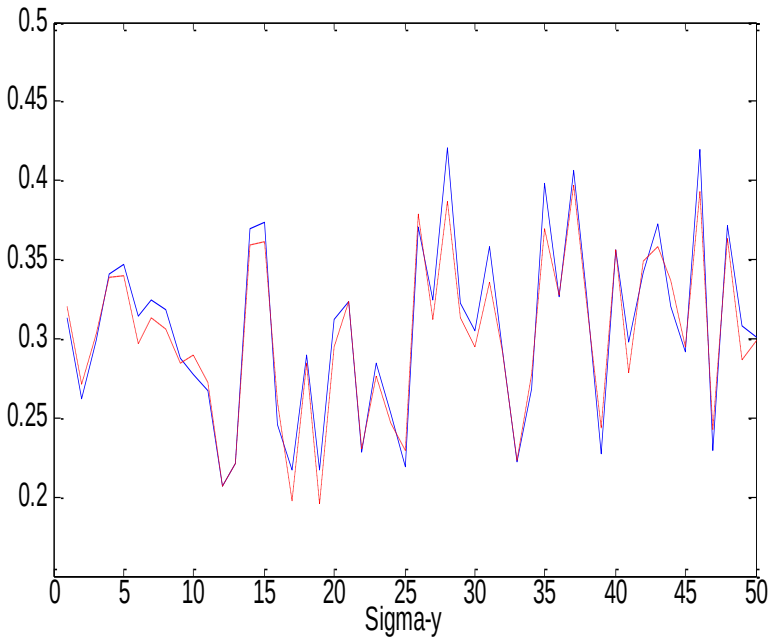


Fig. 6.11 σ_y **estimated separately (blue line) together with** σ_y **estimated jointly (red dashed line)**



Looking at Figures 6.8 – 6.11, it appears as though the ordering restrictions on the parameters have very little or no influence on the estimation of the parameters, because the separately and the jointed estimated parameter values are very close to one another.

6.6 Conclusion

In this chapter the Logistic copula is applied for estimating small tail probabilities in joint events of ordered multivariate extreme values. The marginal distributions in the practical application are assumed to be Strict Pareto distributed; in the simulated study the marginals are GPD. Other distributions might also be considered for the marginal distributions. For the practical application it seems that the Logistic copula is a good fit to the data. For further research other types of extreme value copulas may also be considered.

From the specific simulation study in Section 6.5 it appears that the ordering restrictions have little influence on the estimation of the parameters.

Chapter 7

Conclusion and further recommended research

The contribution of this thesis to the modelling of multivariate extreme values is investigated in this chapter.

7.1 Multivariate Generalized Burr-Gamma distribution

In Chapter 2 the modelling of multivariate extreme values through the MGBG distribution is investigated. Three methods are considered for estimating the parameters; the method of moments, the Bayesian-moments method and the pure Bayesian method. A comparison between the Bayesian-moments method and the pure Bayesian method are given in Table 2.2. The QQ-plot of the pure Bayesian method (Figure 2.11) is close to a straight line, indicating a good fit. In this thesis the pure Bayesian method is considered as the most favourable method. The MGBG distribution, together with the pure Bayesian estimation method, seems to give a good fit to the Gariep Dam data set when 12 variables are considered. This conclusion follows from the linearity of the QQ-plots in Figure 2.12. One advantage of this model is that it illustrates that quite a large dimensional case can be treated with reasonable success. This becomes problematic for the other models discussed in this thesis.

7.2 Multivariate regression

In Chapter 3 multivariate extreme values of the Gariep Dam data set are modelled through multivariate regression with SOI values as covariates. The parameters μ and Σ are estimated through the method of least squares while k and ξ are estimated through the pure Bayesian method. 18% of the variability in the data above the threshold is explained by the SOI covariates. Therefore, although multivariate regression can be used for modelling multivariate extreme data, covariates other than the SOI must also be considered. Further research on what covariates to include into the model

can be done. This regression method applied to the MGBG illustrates how covariates can be introduced into the model in high dimensions. Again this is a problem to introduce into the other models discussed here even in a small dimensional case.

7.3 Gumbel copula, Tawn copula and the Dirichlet mixture model

In Chapter 5 multivariate extreme values are modelled through extreme value copulas and through the Dirichlet mixture model. Two extreme value copulas are discussed, the Gumbel and the Tawn copula. For each copula the parameters are estimated separately and jointly through a Bayesian approach. The separate and the joint estimated parameters for each copula are fairly close to each other, shown in Tables 5.1 to 5.5. For each copula joint tail probabilities are calculated using the jointly estimated parameter values. The Tawn copula's tail probability is the closest to the actual tail probability; therefore the Tawn copula seems to give a better fit to the Gariep Dam data set than the Gumbel copula. Further research can be done on fitting other types of extreme value copulas to the Gariep Dam data.

The Dirichlet mixture model is also fitted to the Gariep Dam data. In the two dimensional case, the BIC indicates that a mixture of 2 Beta distributions must be fit to the Fréchet transformed data above the thresholds. The parameters are estimated jointly through a Bayesian approach. The joint tail probability is calculated and is very close to the actual tail probability, indicating that the mixture of two Beta distributions is a good fit. In the three dimensional case, the BIC indicates that a mixture of 2 Dirichlet distributions must be fit to the Fréchet transformed data above the thresholds. The parameters are again estimated jointly through a Bayesian approach. The estimation of joint probabilities becomes very difficult in more than 2 dimensions; therefore the probabilities are calculated through simulation, as shown in Section 5.7.4. The Dirchlet mixture model seems to be the most appropriate model for modelling multivariate extreme values.

7.4 Modelling of ordered multivariate extreme values

In Chapter 6 an ordering restriction is placed on the multivariate data. The data above the thresholds that obey the ordering restriction are modelled through the Logistic copula. The parameters are estimated jointly through a Bayesian approach. Due to the ordering restriction on the data certain ordering restrictions on the parameters need to be build into the estimation process (see Section 6.3). The Logistic copula seems to be a good fit to the Gariep Dam data set (refer to the contours shown in Figure 6.4). The influence of the ordering restrictions on the estimation of the parameters is discussed in Section 6.5. For the specific simulated example, the ordering restriction has little influence on the estimation of the parameters. Further research can be done on other ordering restrictions on the data and the parameters. For different data sets the ordering might have a greater influence.

7.5 Selection of a threshold

In Chapter 4 a crucial aspect of extreme value theory, how to select a threshold is discussed. Before the modelling of extreme events takes place a threshold is selected. Therefore, the research on how to choose the best threshold is an important part of this thesis. Four methods are considered and compared: the Guillou and Hall method, the method of minimizing the mean squared error, the method of the median of the optimal k-values and the method of the entropy of the Dirichlet process. A conclusion on the goodness of fit is made when a GPD is fitted to the data above each threshold and when a QQ-plot is drawn. By judging the linearity of the QQ-plots the best threshold method is chosen. In Chapter 4, Section 4.5, it is shown that the method of the entropy of the Dirichlet process is an appropriate and consistent method for choosing thresholds. This method is used most in this thesis. With the modelling of multivariate extreme values a threshold is chosen for each marginal distribution (by using one of the above mentioned methods) and only data that exceeds all the thresholds simultaneously are modelled.

7.6 Comparing different approaches

If we compare the different approaches with each other, we can conclude that:

(i) The advantage of the MGBG model is that it can be applied to high dimensional cases. The disadvantage of applying the model is the estimation of the large number of parameters, but we managed to provide methods to address it. The model has also the advantage that covariates can be introduced which is not addressed in the other methods discussed. There is in fact very little been done in the literature on introducing covariates in analyzing multivariate extremes.

(ii) The method of Copulas is popular for constructing joint distributions from the marginals. There is quite a large selection of copulas available in the literature. The choice of copulas is still not clear. Here we discussed two types of copulas with illustrations of their applications. Applying this method to a high dimensional case will be complicated.

(iii) The spectral approach where a mixture of Dirichlet's is used to define the spectral density is promising. The only real problem here is to define the number of mixtures. Although the BIC method is suggested as a measure to choose the number of mixtures this still remains an open problem. This method has an advantage over the copulas in that there is only the one method.

(iv) The ordered multivariate case is easily identified by examination of the data. The method of modelling order multivariate extremes can not be compared with the other methods because it is assumed that the data follows an ordered restriction.

Table 7.1 shows a comparison of the tail probabilities, $P\left(X_1 > 3000, X_2 > 11000\right)$, that were calculated for the Gariep data set, using

the MGBG, the two copulas and the Dirichlet mixture model, in the two dimensional case. The tail probabilities for the copulas and the Dirichlet mixture model were calculated in Sections 5.6 and 5.7.3.4 respectively.

The tail probability for the MGBG is calculated by simulating a 100 000 observations from a MGBG with $\hat{\underline{\mu}} = [-6,5378; -6,6496]$, $\hat{\Sigma} = \begin{bmatrix} 1,4787 & 1,0199 \\ 1,0199 & 1,8190 \end{bmatrix}$, $\hat{\underline{k}} = [1,3957; 1,4588]$ and $\hat{\underline{\xi}} = [0,1422; 0,1575]$. Dividing the number of observation that falls within the regions by 100 000 gives an estimate for the tail probability. $\hat{\underline{\mu}}$ and $\hat{\Sigma}$ are the estimates of Section 2.5, but here only the two dimensional case (January and February) is considered. $\hat{\underline{k}}$ and $\hat{\underline{\xi}}$ were estimated with the Pure Bayesian method for the two dimensional case (January and February). Tail probabilities can also be calculated for the multivariate regression model by following a similar procedure as in Section 3.5 where SOI values are incorporated as covariates.

Table 7.1 Comparing the tail probabilities

	MGBG model	**Gumbel copula**	**Tawn Copula**	**Dirichlet mixture model**
Tail probability	0,0177	0,0013	0,0098	0,0272

The empirical probability is 0,0344. This probability is calculated as the relative number of data in the above mentioned region. Table 7.1 shows that the tail probability of the Dirichlet mixture model is the closest to the empirical probability.

In this thesis we emphasize that the Dirichlet mixture model is the preferable method for modelling multivariate extreme values.

BIBLIOGRAPHY

[1] Amore, P 2005, *Asymptotic and exact series representations for the incomplete Gamma function*. Retrieved August 30, 2007, from www.edpsciences.org/articles/epl/abs/2005/13/epl8802/epl8802.html/

[2] Balkema, AA & de Haan, L 1974, 'Residual life at great age', *Ann. Math. Statist.*, vol. 31, pp. 1015-1027.

[3] Beirlant, J, De Waal, DJ & Teugels, JL 2000, 'A Multivariate Generalized Burr-Gamma Distribution', *South African Statist. J.,* vol. 34, pp. 111-133.

[4] Beirlant, J, De Waal, DJ & Teugels, JL 2002, 'The generalized Burr-Gamma family of distribution with applications in extreme value analysis', *Limit Theorems in Probability and Statistics I*, pp.113-132.

[5] Beirlant, J, Goedgebeur, Y, Segers, J & Teugels, J 2004, *Statistics of Extremes Theory and Applications*, Weily, England.

[6] Boldi, MO & Davison, AC 2006, 'A mixture for multivariate extremes', *J. R. Statist. Soc.,* vol. 69, no. B, pp. 217-229.

[7] Capéraà, P, Fougères, AL, & Genest, C 1997, 'A nonparametric estimation procedure for bivariate extreme value copulas', *Biometrika*, vol. 84, pp. 567-577.

[8] De Haan, L 1970, 'On Regular Variation and its Applications to the Weak Convergence of Sample Extremes', *Mathematical Centre Tract,* 32, Amsterdam.

[9] De Haan, L 1985, 'Extremes in higher dimensions: the model and some statistics', Bulletin of the International Institute, *Proceedings of the 45th Session*, Book 4.

[10] Dekkers, ALM, Einmahl, JHJ & de Haan, L 1989, 'A moment estimator for the index and large quantile estimation', *Annals of Statistics,* vol. 17, pp. 1833-1855.

[11] De Waal, DJ 2008, 'A Dirichlet mixture for modelling Bloemfonrtein extreme minimum and maximum temperatures', *to be published South African Statistical Journal,* 2008.

[12] De Waal, DJ, Van Gelder, PHAJM & Nel, A 2007, 'Estimating joint tail probabilities of river discharges through the logistic copula', *Environmetrics*, vol. 18, pp.621-631.

[13] De Waal, DJ & Van Wyk, A 1998, 'Eskom Progress Report', *Department of Mathematical Statistic, University of the Free State*.

[14] Dowd, K 2007, *An Informal Introduction to Copulas*. Retrieved May 18, 2007, from http://www.fenews.com/fen36/topics_act_analysis/topics_act_analysis.html/

[15] Ferguson, TS 1973, 'A Bayesian analysis of some nonparametric problems', *Aan. Statist.* vol. 1, pp. 209-230.

[16] Ferguson, TS 1983, *Bayesian density estimation by mixtures of normal distributions*, *Recent Advances in Statistics*, Rizvi H, Rustag J (eds), Academic Press, New York, pp. 287-302.

[17] Fisher, RA, Tippett, LHC 1928, 'On the estimation of the frequency distributions of the largest of smallest member of a sample', *Proceedings of the Cambridge Philosophical Society,* vol. 24, pp. 180-190.

[18] Gamerman, D & Lopes, HF 2006, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference Second Edition*, Chapman & Hall/CRC, New York.

[19] Gnedenko, BV 1943, 'Sur la distribution limited u terme maximum d'une série aléatoire', *Annals of Mathematics,* vol. 44, pp. 423-453.

[20] Guillou, A & Hall, P 2001, 'A diagnostic for selecting the threshold in extreme value analysis', *Journal of the Royal Statistical Society*, vol. 63, no. B, pp. 293-305.

[21] Honkela, A 2001, *Dirichlet distribution*. Retrieved September 04, 2007, from http://www.cis.hut.fi/ahonkela/dippa/node95.html/

[22] Kotz, S & Nadarajah, S 2000, *Extreme Value Distributions Theory and Applications*, Imperial College Press, London.

[23] Lo, AY 1984, 'On a class of Bayesian non-parametric estimates I Density estimates', *Ann. Statisti.,* vol. 12, pp. 351-357.

[24] Mason, DM 1982, 'Laws of large numbers for sums of extreme values', *Annals of Probability,* vol. 10, pp. 754-764.

[25] Mazzucchi, T 2002, *Bayesian Estimate and Inferences for Entropy and Information Index of Fit*. Retrieved October 02, 2007, from http://hydr.ct.tudelft.nl/wbk/public/gelder/sanpadworkshop2002.pdf/

[26] McQuarrie, ADR & Tsai, CL 1998. *Regression and Time Series Model Selection*. World Scientific Press, Singapore.

[27] Nadarajah, S, Anderson, CW & Tawn, JA 1998, 'Ordered multivariate extremes', *J. Roy. Statist. Soc.,* vol. 60, no. B, pp. 473-496.

[28] National Weather Service, *Climate Prediction Centre*. Retrieved September 04, 2007, from www.cpc.noaa.gov/data/indices/soi/

[29] Nelsen, RB 2006, *An Introduction to Copulas Second Edition*, Springer, New York.

[30] Obretenov, A 1991, 'On the dependence function of Sibuya in multivariate extreme value theory', *J. Multivariate Anal.,* vol. 36, pp. 35-43.

[31] Panjer, HH 2006, *Operational Risk Modeling Analytics*, Wiley-interscience, New Jersey.

[32] Pickands, J 1975, 'Statistical inference using extreme order statistics', *Annals of Statistics,* vol. 3, pp. 119-131.

[33] Pickands, J 1981, 'Multivariate extreme value distributions', *Proc. 43rd Session of the ISI Buenos Aires,* vol. 49, pp. 859-878.

[34] Rencher, A.C 2002, *Methods of Multivariate Analysis*, Wiley interscience, New York.

[35] Rice, JA 1995, *Mathematical Statistics and Data Analysis Second edition*, Duxbury Press, California.

[36] Romeo, JS, Tanaka, NI & Pedroso-de-Lima, AC 2006, 'Bivariate survival modeling: a Bayesian approach based on Copulas', *Lifetime Data Anal*, vol. 12, pp. 205-222.

[37] Salas, SL, Hille E & Etgen, GJ 1999, *Calculus one and several variables Eight Adition*, John Wily & Sons inc., USA.

[38] Segers, J 2004, '*Non-Parametric Inference for Bivariate Extreme-Value Copulas*'. Retrieved September 06, 2007, from

http://www.stat.ucl.ac.be/Samos2004/proceedings2004/Segers.pdf/

[39] Sklar, A 1959, 'Functions de réparation à *n* dimensions et leur marges', *Publications de l'Insitut de Statistique de l'Université de Paris*, vol. 8, pp. 229-231.

[40] Tawn, JA 1988, 'Bivariate Extreme Value Theory: Models and Estimation', *Biometrika*, vol. 75, no. 3, pp. 397-415.

[41] Tiago de Oliveira, J 1958, 'Extremal distributions', *Rev. Fac. Ciencias Lisboa,* vol. 7, no. A, pp. 215-227.

[42] Zellner, A 1996, *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons inc., New York.

APPENDIX

The Matlab programmes used in each chapter of this thesis are included on the accompanying CD; the names of the programmes and a description of each programme are listed below.

Chapter 2

1. MGBG.m – Simulates from a MGBG distribution with n = 700, p = 3, $\underline{k} = (3;2,1;2,6)$, $\underline{\mu} = (-3,9;-3,8;-3,6)$, $\underline{\xi} = (0,9;0,1;0,5)$ and $\sum = \begin{bmatrix} 0,014 & 0,004 & 0,007 \\ 0,004 & 0,007 & 0,004 \\ 0,007 & 0,004 & 0,013 \end{bmatrix}$. After simulation the upper quantiles are obtained as threshold values. μ and Σ are estimated from the data below the threshold using the method of moments.

2. Dirichent.m – Selects the threshold, of the simulated MGBG distribution, with the entropy of the Dirichlet process. The method of moments is used to estimate μ and Σ from the data below the threshold.

3. EstBayesMoments.m – Estimates k and ξ simultaneously with the Bayesian-Moments approach. The process is repeated a 100 times.

4. GPDest.m - Estimates k and ξ simultaneously with the pure Bayesian approach. The Metropolis Hastings algorithm is considered with 5000 iterations. The process is repeated a 100 times.

5. GPDcom.m – Compares the approximated GPD with the incomplete Gamma distribution.

6. GariepEst.m – Estimates the parameters of the Gariep dam data set. The Method of moments and the pure Bayesian method are considered for the estimation.

7. Variance2.m – Calculates the variance of the predictions.

Chapter 3

1. GariapMaxSOI.m – Gives the data set of the maximum monthly inflow of water into the Gariep dam and the data set of the SOI covariates. This programme also selects the threshold through the entropy of the Dirichlet process and obtains the transformed data below the threshold.

2. RegressionEst.m – Estimates β and Σ by using regression and estimates *k* and ξ by using the Pure Bayesian approach.

3. TailProb.m – Simulates a data set and calculates the tail probabilities for different SOI values.

Chapter 4

1. SimulateFrechetBurr.m – Simulates a 1000 observations from a Fréchet distribution with $\alpha = 1,9$ and from a Burr distribution with $\eta = 3,9, \tau = 1,2$ and $\lambda = 2,5$.

2. GandHmethod.m – Chooses thresholds for the simulated Fréchet and Burr data sets by using the Guillou and Hall method.

3. MSEmethod.m - Chooses thresholds for the simulated Fréchet and Burr data sets by using the method of minimizing the mean squared error plot.

4. MedianMethod.m - Chooses thresholds for the simulated Fréchet and Burr data sets by using the method of the median of the optimal *k*-values.

5. Dirichlet.m - Chooses thresholds for the simulated Fréchet and Burr data sets by using the method of the entropy of the Dirichlet process.

6. BayesEst.m – Estimates the parameters, γ and σ, of the GPD fitted to the simulated Fréchet and Burr data sets above the thresholds.

7. SimDirichlet.m – Simulates 50 data sets from a Burr distribution with $\tau = 0,9, \eta = 4,2$ and $\lambda = 1,5$. For each simulation a threshold is chosen through the method of the entropy of the Dirichlet distribution and a Strict Pareto distribution is fitted to the data above the threshold. This program investigates the consistency of the method of the entropy of the Dirichlet distribution.

Chapter 5

1. EstSeparately.m – Estimates the parameters of the GEV marginal distributions separately, using a Bayesian approach.

2. GumbelCopSeparately.m – Estimates the parameter of the Gumbel copula separately, using a Bayesian approach.

3. GumbelCopJoint.m - Estimates the parameters of the Gumbel copula and the GEV marginals jointly, using a Bayesian approach.

4. TawnCopSeparately.m - Estimates the parameters of the Tawn copula separately, using a Bayesian approach.

5. TawnCopJoint.m - Estimates the parameters of the Tawn copula and the GEV marginals jointly, using a Bayesian approach.

6. parestk2.m - Estimates the parameters of the mixture of 2 Beta distributions ,in the 2 dimensional case. The data is first transformed to Fréchet variables and then the pseudo angles are calculated.

7. parestk3.m - Estimates the parameters of the mixture of 3 Beta distributions, in the 2 dimensional case. The data is first transformed to Fréchet variables and then the pseudo angles are calculated. The BIC is compared with the BIC of the mixture of 2 Beta distributions.

8. parestk4.m - Estimates the parameters of the mixture of 4 Beta distributions in the 2 dimensional case. The data is first transformed to Fréchet variables and then the pseudo angles are calculated. The BIC is compared to the BIC of the mixture of 2 Beta distributions.

9. parest3dimk2.m - Estimates the parameters of the mixture of 2 Dirichlet distributions in the 3 dimensional case. The data is first transformed to Fréchet variables and then the pseudo angles are calculated.

10. parest3dimk3.m - Estimates the parameters of the mixture of 3 Dirichlet distributions in the 3 dimensional case. The data is first transformed to Fréchet variables and then the pseudo angles are calculated. The BIC is compared to the BIC of the mixture of 2 Dirichlet distributions.

11. parest3dimk4.m - Estimates the parameters of the mixture of 4 Dirichlet distributions in the 3 dimensional case. The data is first transformed to Fréchet variables and then the pseudo angles are calculated. The BIC is compared to the BIC of the mixture of 2 Dirichlet distributions.

12. SimDirichletMix.m – Simulates from a mixture of 2 Dirichlet distributions in the 3 dimensional case and draws the sample space. This program is also used to calculate probabilities.

Chapter 6

1. WholeDataset.m – Gives the data set of the daily inflow of water into the Gariep dam.

2. Groups20.m – Divides the daily inflow of water into the Gariep dam into blocks of size 20 and obtains the maximum of each block for *d* equal to 1 and *d* equal to the aggregate of 7 days.

3. DirichletSP.m – Chooses the threshold through the entropy of the Dirichlet process when the marginals are assumed to be Strict Pareto distributed.

4. LogistJointEst.m – Fits the Logistic copula to the data above the threshold and estimates the parameters jointly by taking into account the ordered restriction.

5. SimGPD.m – Simulates from 2 GPD distributions to investigate whether the ordering restriction has any influence on the estimation of the parameters. The parameters are estimated separately for each GPD and then jointly where the restriction is taking into account. The process is repeated 50 times.

SUMMARY

The aim of this thesis is to investigate the modelling of multivariate extreme values.

Extreme value theory is becoming very popular as a statistical discipline. Not only is extreme value theory emerging in the statistical field but also in other disciplines such as engineering, financial markets and energy markets. Extreme value analysis focuses on the probability of unusual or extreme events. Extreme value theory is particularly useful in environmental applications such as the modelling of high rainfall, strong wind, *etc*. A lot of literature is available on the modelling of univariate extreme values.

One topic of interest is the calculation of probabilities of related extreme events. To address this topic the modelling of multivariate extreme values is investigated.

Various models are considered for modelling multivariate extreme values. The data set used in the thesis is the daily inflow of water, in cubic meters per second, into the Gariep Dam over a period of 29 years, from 1976 to 2006, excluding 1980 and 1982 due to large data losses. The models considered are the Multivariate Generalized Burr-Gamma distribution (Chapter 2), multivariate regression (Chapter 3), the Gumbel, Tawn and Logistic copulas (Chapter 5 and 6), and the Dirichlet mixture model (Chapter 5). The emphasis of Extreme Value Theory lies in the extreme values or events. Therefore, a threshold is chosen and only data above the threshold is modelled. Hence, specific research is done in this thesis on how to choose a threshold (Chapter 4). The preferred threshold method in this thesis is the method of the negative differential entropy of the Dirichlet process (Chapter 4). The thesis mainly considers a Bayesian approach for the estimation of parameter, although other approaches are also discussed.

The last chapter, Chapter 7, gives a conclusion on the work that is covered in the thesis and recommendations for further research.

Keywords: Extreme value theory, QQ-plots, threshold, MGBG, GPD, Fréchet marginals, Bayesian estimates, MDI prior, joint estimating, Dirichlet process, tail probabilities, extreme value copula, Logistic copula, dependence structure, Dirichlet mixture model, spectral distribution function.

SAMEVATTING

Die doel van die tesis is om ondersoek is te stel na die modellering van meerveranderlike-ekstreemwaardes.

Die gewildheid van ekstreemwaardeteorie as ’n statistiese dissipline neem toe. Ekstreemwaardeteorie neem nie net toe in ’n statistiese veld nie maar ook in ander dissiplines bv. ingenieurswese, finansiële markte en energiemarkte. Ekstreemwaarde-analise fokus op waarskynlikhede van ongewone, ekstreme gebeurtenisse wat plaasvind. Ekstreemwaardeteorie is besonders bruikbaar in toepassings omtrent omgewingsaspakte bv. hoë reënval, sterk winde ens. Baie literatuur is beskikbaar omtrent modellering van eenveranderlike-ekstreemwaardes.

Die onderwerp van belang is die berekening van waarskynlikhede van verwante ekstreme gebeurtenisse. Hierdie onderwerp word aangespreek deur die modellering van meerveranderlike-ekstreemwaardes te ondersoek.

In hierdie tesis word verskeie modelle beskou vir die modellering van meerveranderlike-ekstreemwaardes. Die datastel wat in hierdie tesis gebruik word, is die daaglikse invloei van water in die Gariepdam, gemeet in kubieke meter per sekonde, oor ’n periode van 29 jaar,1976 tot en met 2006, met uitsluiting van 1980 en 1982 weens onvolledige data inskrywings. Die modelle wat beskou word is die Meerveranderlike Veralgemeende Burr Gammaverdeling (Hoofstuk 2), meerveranderlike regressie (Hoofstuk 3), die Gumbel, Tawn en Logistiese kopula (Hoofstuk 5 en 6), en die Dirichletmengselmodel (Hoofstuk 5). Ekstreemwaardeteorie plaas die klem op die ekstreme waardes of gebeurtenisse, daarom word ’n drumpelwaarde gekies en net data bokant die drumpelwaarde word gemodelleer. Spesifieke navorsing is in hierdie tesis gedoen oor hoe om ’n drumpelwaarde te kies (Hoofstuk 4). Die voorkeur metode van hoe om die drumpelwaarde te kies is die metode van die negatiewe differensiële entropie van die Dirichletproses (Hoofstuk 4).

Alhoewel ander benaderings ook bespreek word, word ’n Bayesbenadering hoofsaaklik vir die beraming van die parameters beskou.

Die laaste hoofstuk, Hoofstuk 7, gee ’n gevolgtrekking, oor die werk wat gedoen is in die tesis, en ook verdere navorsingsaanbevelings.

Sleutelwoorde: Ekstreemwaardeteorie, QQ-voorstellings, drumpelwaarde, MGBG, GPD, Fréchet-randverdelings, Bayesberaming, MDI-prior, gesamentlike beraming, Dirichletproses, stertwaarskynlikhede, ekstreemwaarde-kopula, Logistiese kopula, afhanklikheidstruktuur, Dirichletmengselmodel, spektraalverdelingsfunksie.