# Extreme Value Mixture Modelling with Simulation Study and Applications in Finance and Insurance

**By**

**Yang Hu**

**A thesis submitted in partial fulfillment of the requirements**
**for the degree of Master of Science**
**at University of Canterbury, New Zealand**
**July 2013.**

**Abstract**

Extreme value theory has been used to develop models for describing the distribution of rare events. The extreme value theory based models can be used for asymptotically approximating the behavior of the tail(s) of the distribution function. An important challenge in the application of such extreme value models is the choice of a threshold, beyond which point the asymptotically justified extreme value models can provide good extrapolation. One approach for determining the threshold is to fit the all available data by an extreme value mixture model.

This thesis will review most of the existing extreme value mixture models in the literature and implement them in a package for the statistical programming language `R` to make them more readily useable by practitioners as they are not commonly available in any software. There are many different forms of extreme value mixture models in the literature (e.g. parametric, semi-parametric and non-parametric), which provide an automated approach for estimating the threshold and taking into account the uncertainties with threshold selection.

However, it is not clear that how the proportion above the threshold or tail fraction should be treated as there is no consistency in the existing model derivations. This thesis will develop some new models by adaptation of the existing ones in the literature and placing them all within a more generalized framework for taking into account how the tail fraction is defined in the model. Various new models are proposed by extending some of the existing parametric form mixture models to have continuous density at the threshold, which has the advantage of using less model parameters and being more physically plausible. The generalised framework all the mixture models are placed within can be used for demonstrating the importance of the specification of the tail fraction. An `R` package called `evmix` has been created to enable these mixture models to be more easily applied and further developed. For every mixture model, the density, distribution, quantile, random number generation, likelihood and fitting function are presented (Bayesian inference via MCMC is also implemented for the non-parametric extreme value mixture models).

A simulation study investigates the performance of the various extreme value mixture models under different population distributions with a representative variety of lower and upper tail behaviors. The results show that the kernel density estimator based non-parametric form mixture model is able to provide good tail estimation in general, whilst the parametric and semi-parametric forms mixture models can give a reasonable fit if the distribution below the threshold is correctly specified. Somewhat surprisingly, it is found that including a constraint of continuity at the threshold does not substantially improve the model fit in the upper tail. The hybrid Pareto model performs poorly as it does not include the tail fraction term. The relevant mixture models are applied to insurance and financial applications which highlight the practical usefulness of these models.

# Acknowledgments

I would like to thank my supervisors Dr Carl Scarrott and Associate Professor Marco Reale for their support and guidance throughout my study.

I would like to thank all the members in the Department of Mathematics and Statistics at University of Canterbury. In particular, thanks to Steve Gourdie and Paul Brouwers for providing IT support.

Another thank you goes to all the postgraduate students in the department, who have made my time as a student very enjoyable. Special thanks must go to James Dent for reading my draft thesis and to Chitraka Wickramarachchi for discussing statistics questions.

Finally, I would like to thank my parents for their endless support and encouragement through my study.

# Contents

# Chapter 1

# Introduction

## 1.1 Motivation

Extreme value theory is used to describe the likelihood of unusual behavior or rare events occurring, and has been used to develop mathematically and statistically justifiable models which can be reliably extrapolated. Extremal data are inherently scarce, thus making inference challenging. Extreme value modeling has been widely used in financial, insurance, hydrology or environmental applications where the risk of extreme events is of interest. A good reference book that discusses different applications with extreme value theory is by Reiss and Thomas (2001). In this thesis, I will apply the relevant models to financial, insurance and hydrology applications.

It is well known that financial data often exhibit heavy (heavier than normal) tails with dependences, which makes the inference more challenging and problematic. However, extreme value modeling often assumes that the observations are independent and identically distributed (iid). One way of overcoming such issues is to use a volatility model to capture the dependence, which leads to clustering of extreme values. The autoregressive conditional heteroskedasticity (ARCH) and generalized autoregressive conditional heteroskedasticity (GARCH) volatility model have been extensively used in financial applications. One of the most influential papers is by McNeil and Frey (2000). They develop the two stage approach by fitting the GARCH model at the first stage to describe the volatility clustering and then the generalised Pareto distribution(GPD) fits to the tails of the residual. Zhao et al. (2010) further extend this method at the second stage by fitting GPDs on both lower and upper tails of the residuals, with a truncated normal distribution for the main mode in an extreme value mixture model.

Extreme value theory also plays an important role in insurance risk management, as the insurance company needs to consider the potential risk of large claims (e.g. earthquake, volcanic eruptions, landslides and other possibilities). The actuary will estimate the price of the insurance products (net premium) to cover the potential insurance risk. In particular, it is of great importance for re-insurance. McNeil (1997) studies the classic Danish insurance data by applying the peaks over threshold method. Resnick (1997) acknowledges that McNeil's work is of great importance in insurance claim application and an excellent example of adopting the extreme value model. Embrechts (1999) demonstrates that extreme value theory is a very useful tool for insurance risk management.

In the hydrology field, flood damage can cause a lot of losses. Researchers aim to predict the probability of the next flood event, which could be important for future plans. They often study the flood frequency and associated flood intensity. Rossi et al. (1984) and Cooley et al. (2007) use extreme value theory models for flood frequency analysis. Katz et al. (2002) investigate the extreme events in hydrology as well. Similarly, sea-level analysis is also a very hot research topic in hydrology field. For example, Smith (1989) and Tawn (1990) apply extreme value modeling to the sea level applications.

Based on the above, extreme value modeling has attracted great interest in many different fields. In particular, the GPD is a very popular extreme value model, which can give a good model for the upper tail providing reliable extrapolation for quantiles beyond a sufficiently high threshold. An important challenge in application of such extreme value models is the choice of threshold, beyond which the data can be considered as extreme data or more formally where the asymptotically justified extreme value models will provide an adequate approximation to the tail of the distribution. Various approaches have been developed to aid the selection of the threshold, all with their own advantages and disadvantages. Recently, there has been increasing interests in extreme value mixture models, which not only consider the threshold estimation but also the quantification of the uncertainties that result from the threshold choice.

## 1.2 Thesis Objectives

The main purpose of the thesis is to implement the existing extreme value mixture models in the statistical package `R` and create a library to make these more readily available for practitioners to use. This library will include both likelihood and computational Bayesian inference approaches where appropriate. A detailed simulation study is carried out which aims to compare different extreme value mixture model properties to provide some concrete advice about the performance of each mixture model and guidance on how their formulation influences their performance.

There is no generalized framework for defining these extreme value mixture models. In particular, there is no consistency in terms of how the proportion of the population $\phi_u$ above the threshold is defined in the model and subsequently estimated. The proportion in the upper tail is sometimes referred to as the tail fraction, which treated as a parameter the MLE is the sample proportion. However, it is common when defining extreme value mixture models to allow the tail fraction $\phi_u$ to be inferred by the proportion of the bulk model. The approach of treating the tail fraction as a parameter estimated using the MLE is consistent with traditional applications of the GPD. However, the definition using the bulk model is commonplace and has no formal justification and apart from convenience. This thesis derives a general framework for the extreme value mixture models which permits the tail fraction, $\phi_u$ to be specified in either of these forms, and will explore which approach performs well in general.

The parametric form of mixture models provide the simplest form of extreme value mixture models (e.g. Frigessi et al. (2002), Behrens et al. (2004) and Zhao et al. (2010)). These models typically treat threshold as a parameter to be estimated. The main concern is whether they can provide sufficient flexibility if they are mis-specified.

Continuity at the threshold is another common issue with some of these mixture models, as most of them do not require the density function to be continuous at the threshold. Carreau and

Bengio (2008) develop the hybrid Pareto model by combining a normal and a GPD distribution and set two continuity constraints at the threshold (i.e. continuity in zeroth and first derivative). They claim that this model works well for heavy tailed and asymmetric distribution. This thesis will evaluate that claim, and explore whether this mixture model can perform well or not in short or exponential tail distribution. The simulation study will also explore whether the extra continuity constraint at the threshold would improve the model performance.

## 1.3 Thesis Structure

Chapter 2 summaries the key background material on relevant statistical topics including extreme value theory, mixture models and kernel density estimator.

Chapter 3 reviews the existing extreme value mixture models approaches, with some references to alternative approaches in the literature. All of these existing models are placed within my new generalised framework, with a consistent notation used throughout. Some new models I have created by extensions from the original mixture models are also presented.

Chapter 4 demonstrates the usage of the `R` library I have developed implementing these mixture models. A very detailed simulation study is carried out in Chapter 5. The different extremal mixture models are divided into three groups in evaluating their performance based on the support of the distribution and whether a single tail or both tails are described. The simulation study will consider the following questions:

- How does each model perform in general for tail estimation? Is there a "best" mixture model, or "best type" of mixture model? Are there any which can perform poorly in certain situations?
- Is it better to specify the tail fraction in terms of a parameter to be estimated, or to use the bulk model tail fraction?
- Is it better to include a constraint that the density be continuous at the threshold?

In the chapter 6, all the relevant models are applied to the classic Danish insurance data, FTSE 100 index log daily return data and Fort Collins precipitation data.

# Chapter 2

# Background

## 2.1 Extreme Value Theory

This chapter will review the classic probabilistic extreme value theory and the classical models used for describing extreme events. Followed by an overview of kernel density estimation to support the presentation in the rest thesis.

### 2.1.1 Extreme Value Modeling

Classical extreme value theory is used to develop stochastic models towards solve real life problems related to unusual events. Classical theoretical results are concerned with the stochastic behavior of some maximum (minimum) of a sequence of random variables which are assumed to be independently and identically distributed.

Coles (2001) explains that let $X_1, ..., X_n$ be a sequence of independent random variable with population distribution function $F$ and $M_n$ represents the maxima or minima of the process over the block size $n$. The distribution of $M_n$ can be derived by:

$$Pr\{M_n \leq x\} = Pr\{X_1 \leq x, \ldots, X_n \leq x\} = Pr\{X_1 \leq x\} \times \cdots \times Pr\{X_n \leq x\} = \{F(x)\}^n.$$

If the population distribution is known, the distribution of $M_n$ can be determined exactly. However, in the more usual case of unknown population distributions, the distribution of $M_n$ can be approximated by modeling $F^n$ through asymptotic theory of $M_n$. As $n \to \infty$, the distribution of $M_n$ will degenerate to a point mass at the upper end point of $F$. For most population distributions this degeneracy problem can be eliminated by allowing some linear renormalization of the $M_n$, similar to that used in the Central Limit Theorem. Consider a linear renormalization:

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

for sequences of constants $a_n > 0$ and $b_n$. After choosing suitable $a_n$ and $b_n$, the distribution of $M_n$ may be stabilized and which lead to extremal types theorem.

**Extremal Types Theorem.** *If there exists sequences of constants* $\{a_n > 0\}$ *and* $\{b_n\}$*, as* $n \to \infty$*, such that*

$$Pr\{(M_n - b_n)/a_n \leq x\} \to G(x)$$

*where G is a non-degenerate distribution function, then G belongs to one of the following families:*

$$I : Gumbel : G(x) = \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\} \quad -\infty < x < \infty; \tag{2.1}$$

$$II : Fréchet : G(x) = \begin{cases} 0, & x \leq \mu; \\ \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)\right\}, & x > \mu; \end{cases} \tag{2.2}$$

$$III : Weibull : G(x) = \begin{cases} \exp\left\{-\left[-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}\right]\right\}, & x \leq \mu; \\ 1, & x > \mu; \end{cases} \tag{2.3}$$

*for parameters* $\sigma > 0$, $\mu \in \mathbb{R}$.

The proof of this theorem can be found in Leadbetter et al. (1983). The extremal types theorem suggests that if $M_n$ can be stabilized with choosing suitable $a_n$ and $b_n$, then the limit distribution $M_n^*$ will be one of the three extreme value distributions. No matter what the population distribution of $M_n$ is, if a non-degenerate limit can be obtained by linear renormalisation, then these three different extreme value distributions which have distinct behaviors are the only possible limit distribution of $M_n^*$. But there is no guarantee of a non-degenerate distribution.

For an unknown population distribution it is inappropriate to adopt one type of limiting distribution and ignore any uncertainties associated with family distribution. A better approach is to adopt a universal extreme value distribution which encompasses all three types as a special case.

#### 2.1.1.1 Generalized Extreme Value Distribution (GEV distribution)

von Mises (1954) and Jenkinson (1955) proposed a way of unifying the three different types of extreme value distributions, which lead to the generalized extreme value distribution GEV$(\mu, \sigma, \xi)$.

**Theorem 1.** *If there exists sequences of constants* $\{a_n > 0\}$ *and* $\{b_n\}$, *as* $n \to \infty$, *such that*

$$Pr\left\{(M_n - b_n)/a_n \leq x\right\} \to G(x)$$

*where G is a non-degenerate distribution function,then G is a member of the GEV family:*

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

*defined on* $\left\{x : 1 + \xi(\frac{x-\mu}{\sigma}) > 0\right\}$, *where* $\sigma > 0$, $\mu \in \mathbb{R}$ *and* $\xi \in \mathbb{R}$.

Now $\mu$ is the location parameter, $\sigma$ is the scale parameter and $\xi$ is the shape parameter. The GEV distribution simplifies the three different types of extreme value distributions to one family. The GEV distribution is used as an asymptotic approximation for modeling the maxima of the finite sequences.

The GEV distribution provides us a way of modeling the extremes of block maxima (or minima). If the iid observations $x_1, ..., x_i$ are divided into $m$ blocks of size $n$, where $n$ is large enough, then we can generate a series of sample maxima $M_{n,1},\ldots,M_{n,m}$ for which their distribution can be approximated by the asymptotically motivated GEV distribution. The block size is of importance in GEV model. If the block size is too small, this can lead to a poor asymptotic approximation.

The distribution function of GEV$(\mu, \sigma, \xi)$ is given by:

$$F(x|\mu,\sigma,\xi) = \begin{cases} \exp\left\{-\left[1+\xi\left(\dfrac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\} & \xi \neq 0, \\ \exp\left[-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right] & \xi = 0. \end{cases}$$

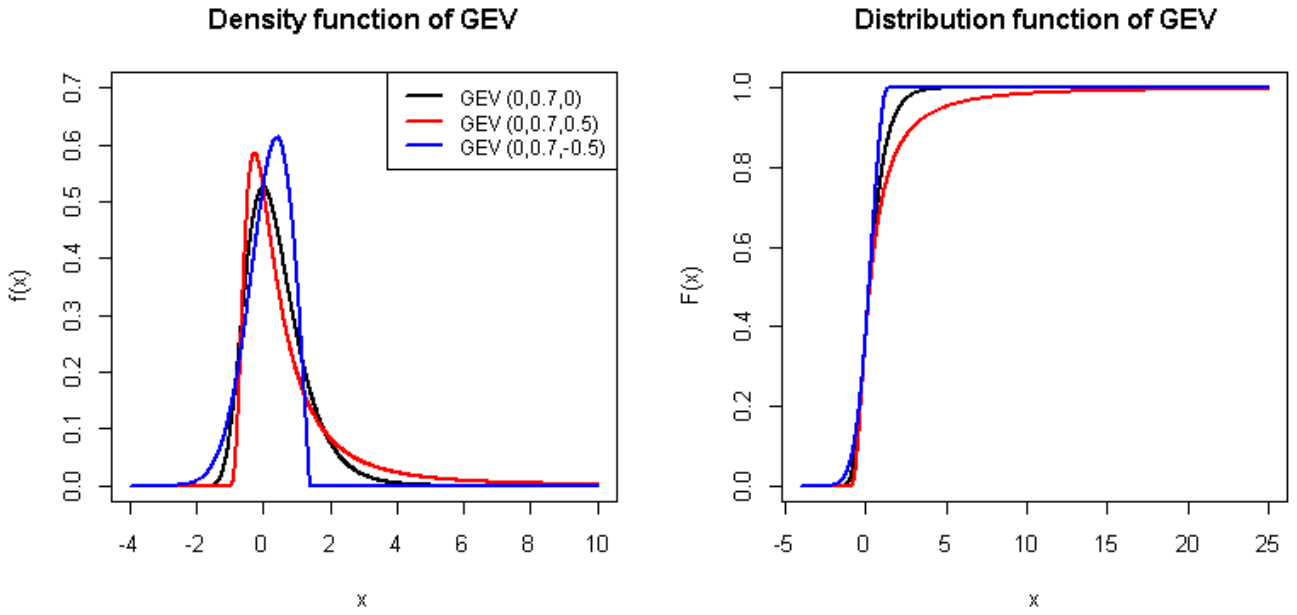



Figure 2.1: Examples of the density and corresponding distribution function for the GEV distribution with $\mu = 0$, $\sigma = 0.7$, and varying $\xi$ with (0, -0.5, 0.5)

From Figure 2.1, the example demonstrates how the shape parameter affects the tail behavior by keeping the location and scale parameters the same. If $\xi = 0$, the GEV distribution belongs to Gumbel distribution and it decays exponentially. If $\xi < 0$, the GEV belongs to the negative Weibull distribution with a finite short upper endpoint. Otherwise, the GEV distribution belongs to the Fréchet distribution, which has a heavy tail behavior.

The inverse of GEV distribution function represents the quantile function. Let $F(z_p) = 1 - p$, where $p$ is the upper tail probability.

$$z_p = \begin{cases} u - \frac{\sigma}{\xi}\left\{1 - [-\log(1-p)]^{-\xi}\right\} & \xi \neq 0, \\ u - \sigma \log\left[-\log(1-p)\right] & \xi = 0. \end{cases}$$

$z_p$ is referred to return level with return period of $1/p$. In general, $z_p$ is exceeded on average every $1/p$ period of time. Equivalently, the average time for a very rare event to exceed the $z_p$ is every $1/p$ period of time. The return level is often interpreted as the expected waiting time until another exceedance event.

#### 2.1.1.2 Generalized Pareto Distribution (GPD)

The block maxima based the GEV distribution is inefficient as it is wasteful of data when the complete dataset (or at least all extreme values) are available. One way of overcoming these difficulties is to model all the data above some sufficiently high thresholds. Such a model is commonly referred as the peaks over threshold or threshold excess model. The advantage of the threshold excess model is that it can make use of the all the "extreme" data. Under certain conditions, Coles (2001) explains that let $X_1, \ldots, X_n$ be a sequence of iid random variables, the excess $X - u$ over a suitable $u$ can be approximated by the generalized Pareto distribution $G(x|u, \sigma_u, \xi)$.

The distribution function of GPD is given by:

$$G(x|u,\sigma_u,\xi) = \begin{cases} 1 - \left[1 + \xi\left(\dfrac{x-u}{\sigma_u}\right)\right]_+^{-1/\xi} & \xi \neq 0, \\ 1 - \exp\left[-\left(\dfrac{x-u}{\sigma_u}\right)\right]_+ & \xi = 0. \end{cases}$$

where $x > u$, $\sigma_u > 0$, $\left[1 + \xi\left(\frac{x-u}{\sigma_u}\right)\right] > 0$ and $\sigma_u$ reminds us the dependence between scale parameter $\sigma_u$ and threshold $u$. The dependence of the GPD scale $\sigma_u$ on the threshold is made explicit by the subscript $u$, which is also retained in the functions in the `R` package described in Chapter 4.

The shape parameter $\xi$ is of importance in determining the tail distribution of the GPD:

- $\xi < 0$: short tail with finite upper end point of $u - \dfrac{\sigma_u}{\xi}$;
- $\xi = 0$ : exponential tail; and
- $\xi > 0$: heavy tail.

GPD is defined conditionally on being above threshold $u$. The unconditional survival probability is defined as:

$$\begin{aligned} Pr(X > x) &= Pr(X > u) * Pr(X > x | X > u) \\ &= Pr(X > u) * [1 - Pr(X < x | X > u)] \\ &= \phi_u * [1 - G(x|u, \sigma_u, \xi)], \end{aligned}$$

where $\phi_u = Pr(X > u)$ is the probability of being above the threshold. The value of $\phi_u$ is another parameter required when using the GPD. **It is an implicit parameter which is often forgotten about.**

It would be worth showing how GEV and GPD models are related with each other. Let Y follows GEV distribution over a high threshold $u$ and y is the excess $(x-u)$.

$$\begin{aligned} Pr\{Y > u+y|Y > u\} &= \frac{1-Pr(Y<y+u)}{1-Pr(Y<u)} \\ &\approx \frac{[1+\xi(u+y-\mu)/\sigma]^{-1/\xi}}{[1+\xi(u-\mu)/\sigma]^{-1/\xi}} \\ &= \left[1+\frac{\xi(u+y-\mu)/\sigma}{1+\xi(u-\mu)/\sigma}\right]^{-1/\xi} \\ &= \left[1+\frac{\xi y}{\hat{\sigma_u}}\right]^{-1/\xi}, \end{aligned}$$

where $\hat{\sigma_u} = \sigma + \xi(u-\mu)$. This derivation makes the dependence of the GPD scale $\sigma_u$ and threshold $u$ very clear. The GPD shape parameter is the same as the GEV shape parameter.

**Maximum Likelihood Estimation of GPD**

Let $x_1, ..., x_n$ be a sequence of observations of independent random variables which are the exceedances above a threshold $u$. The log-likelihood function is derived from taking logarithm of the joint density function. Let $nu$ be the number of observations above the threshold $u$.

$$\log L(\sigma_u, \xi|X) = -nu\log\sigma_u - (1+1/\xi)\sum_{i=1}^{nu}\log\left[1+\xi\left(\frac{x_i-u}{\sigma_u}\right)\right],$$

with $\left[1+\xi\left(\frac{x_i-u}{\sigma_u}\right)\right] > 0$ and $\xi \neq 0$. The constraint $\left[1+\xi\left(\frac{x_i-u}{\sigma_u}\right)\right]$ implies $x < \mu - \sigma_u/\xi$.

When $\xi = 0$, the log-likelihood function can be derived in similar way:

$$\log L(\sigma_u|X) = -nu\log\sigma_u - \sum_{i=1}^{nu}\left(\frac{x_i-u}{\sigma_u}\right)$$

with $\xi = 0$.

**Regularity condition**

There are certain regularity conditions for the maximum likelihood estimators to exhibit the usual asymptotic properties. It is worthwhile to notice that maximum likelihood estimators are not always valid and the regularity conditions do not always exist. The invalid regularity conditions are caused by upper endpoint of GEV distribution being dependent on the parameters where the endpoints are parameter value of $\mu - \sigma_u/\xi$ when $\xi < 0$.

Smith (1985) has outlined the following result:

- $\xi > -0.5$, maximum likelihood estimators have usual asymptotic properties;
- $-1 < \xi < -0.5$, maximum likelihood estimators generally exist, but do not have usual asymptotic properties;

- $\xi < -1$, maximum likelihood estimators generally do not exist.

In general, the GPD distribution has a very short bounded upper tail behavior if $\xi \leq -0.5$, so rarely comes up in applications.

## 2.2 Threshold Choice and Common Approach

This section will demonstrate the difficulty in choosing the threshold through the common diagnostic plots used in practice.

### 2.2.1 Threshold Choice

Threshold selection is still an area of ongoing research in the literature which can be of the critical importance. Coles (2001) states that the selection of the threshold process always is a trade-off between the bias and variance. If the threshold is too low the asymptotic arguments underlying the derivation of the GPD model are violated. By contrast, too high a threshold will generates fewer excesses $(x - u)$ to estimate the shape and scale parameter leading to high variance. Thus threshold selection requires considering whether the limiting model provides a sufficiently good approximation versus the variance of the parameter estimate.

In the traditional extreme value modeling approaches, the threshold is chosen and fixed before the model is fitted which is commonly regarded as the fixed threshold approach. Various diagnostics plots to evaluate the model fit are commonly used for choosing the threshold. Coles (2001) outlined three diagnostics for the choice of threshold including the mean residual life plot, parameter stability plot and more general model fit diagnostics plots.

#### 2.2.1.1 Parameter Stability Plot

Another plot that can be used for assessing the model fit to judge the threshold is usually referred as the parameter stability plot.

If the excesses over a high threshold $u$ follow a GPD with $\xi$ and $\sigma_u$, for any higher threshold $v > u$, the excesses still follow a GPD with $\xi$ but with scale parameter of :

$$\sigma_v = \sigma_u + \xi(v - u).$$

By re-parametrise the scale parameter $\sigma_v$:

$$\sigma^* = \sigma_v - \xi v.$$

The $\sigma^*$ now no longer depends on $v$, given that $u$ is a reasonably high threshold. This plot fits GPD over a certain values of thresholds against the shape and scale parameter. The threshold $u$ should be chosen where the shape and modified scale parameter remain constant after taking the sampling variability into account. Once a suitable threshold is reached, the excess over this threshold follows GPD.

#### 2.2.1.2 Mean Residual Life Plot

If the excess $X - u$ are approximated by a GPD, for a suitable $u$, the mean excess is:

$$E(X - u|X > u) = \frac{\sigma_u}{1 - \xi} \quad \text{for} \quad \xi < 1$$

For any higher threshold $v > u$, the mean excess is defined as:

$$E(X - v|X > v) = \frac{\sigma_v}{1 - \xi} = \frac{\sigma_u + \xi(v - u)}{1 - \xi} \quad \text{for} \quad \xi < 1$$

It shows that the mean excesses $E(X - u|X > u)$ is a linear function of $u$, once a suitable high threshold $u$ has been reached.

The sample mean residual life plot points, drawn using the

$$\left( u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right)$$

where $x_{(i)}$ is the observation above the $u$ and $n_u$ is the number of observation above the $u$.
For a sufficiently high threshold $u$, all the excesses $v > u$ in the mean residual life plot changes linearly with $u$. This property could potentially provide a way of choosing the the threshold. Once the sampling variability is included, the threshold should be chosen where the relationship for mean excess of all higher thresholds is linear. Coles (2001) concluded that interpretation of such plot is not an easy task.

Both the mean residual life and parameter stability plot are popular choices for graphically choose for the threshold.

#### 2.2.1.3 Model Fit Diagnostic Plot

Various standard statistical model diagnostic plots such as probability plot, quantile plot, return level plot and empirical versus fitted density comparison plot can be used for checking the model fit and the suitability of the threshold choice. These checks based on a chosen threshold, so they are post-calculation diagnostic plots. These plots provide an alternative way of assessing the model performance.

### 2.2.2 Discussion and example

A possible disadvantage with these diagnostics plots is that the threshold chosen from inspecting the graph could suffer from substantial subjectivity expertise. The drawback with these approaches would mean that some different thresholds can be suggested by different practitioners using the same diagnostics from their point of views. The following example will demonstrate that graphically inspecting data could result in substantial differences in such subjective threshold selection.

As seen in Figure 2.2, there is an example of mean residual life plot for the daily precipitation data for Fort Collins, Colorado from the `extRemes` package in `R`. With 95% confidence intervals,

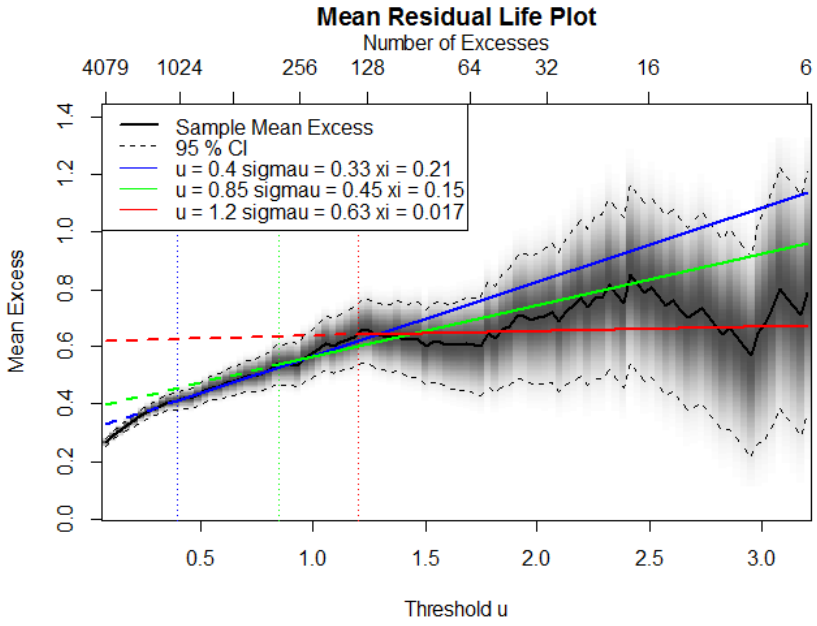



Figure 2.2: The mean residual life plot for Fort Collins precipitation data.

the ideal threshold $u$ is the one where the mean excess above the threshold $u$ is linear after considering the sample variability. In order to show the problematic diagnostics for the threshold selection, the non-stationarity and dependence of the series will be ignored. There are several possible choices for the thresholds. Katz et al. (2002) suggested the threshold $u$ to be 0.395. When threshold $u$ is 0.395, it results in a heavy tail behavior ($\hat{\xi} = 0.21$) while the threshold could lead to exponential tail behavior when the threshold $u$ is equal 1.2 ($\hat{\xi} = 0.003$). Another possible choice for $u$ is 0.85, which exhibits a heavy tail with the shape of $\hat{\xi} = 0.134$. From the plot, all three thresholds exhibit different tail behavior as the shape parameter varies a lot.

Figure 2.3 is the parameter stability plot for the Fort Collins daily precipitation data. When the thresholds are 0.395, 0.85 and 1.2, the modified scale parameter($\sigma^*$ in Section 2.2.1.1) seems reach a plateau. However, if the shape parameter is taken into account, it is difficult to tell which the lower threshold choice is. The parameter stability plot cannot provide a consistent conclusion once the sample variability of both parameters and the negative dependence between the modified scale and shape parameter are taken into account. Scarrott and MacDonald (2012) used this data to demonstrate the above threshold selection issue in detail.

These examples have shown that the diagnostics plots could result in subjective judgement. The threshold choice procedure needs a more objective approach. Not only does the traditional threshold selection approach face substantial subjectivity in threshold choice, but also it lacks the considerations of uncertainty of the threshold choice in the subsequent inferences. Once the threshold is chosen, the variability or uncertainty of the threshold will not be taken into account. Pre-fixed the threshold is at least not ideal in most cases unless there is a valid reason for doing that. In this case, threshold is always constrained by physical objectives.

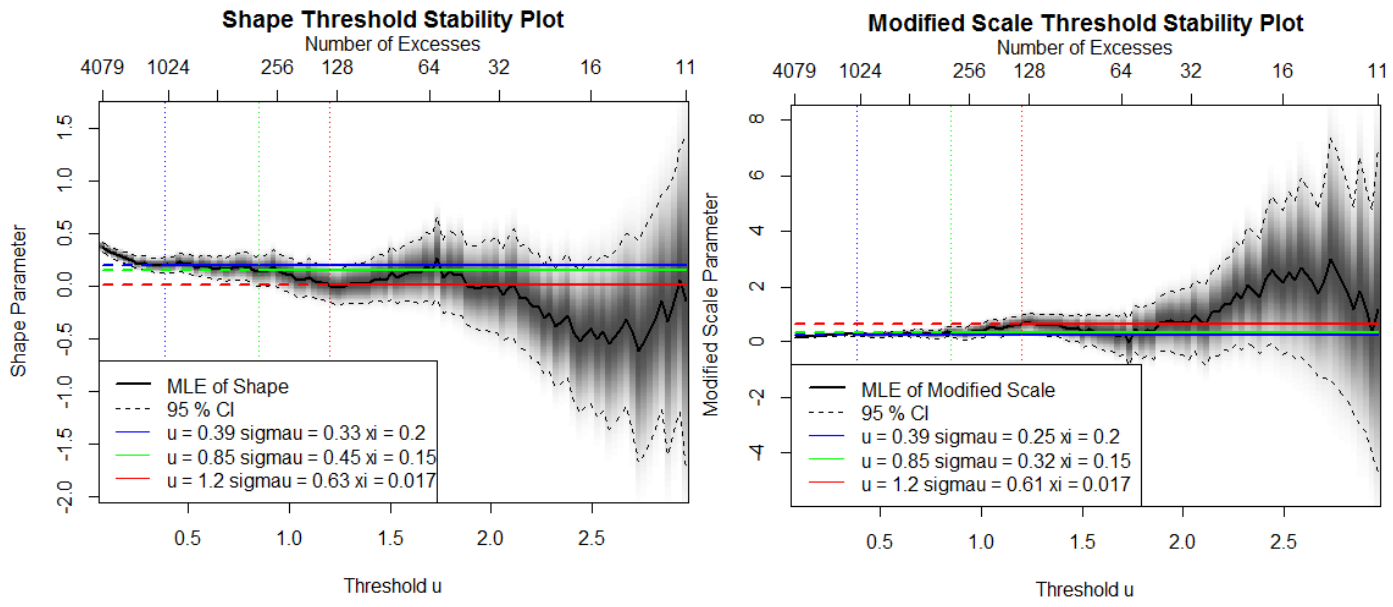



Figure 2.3: The parameter threshold stability plot for Fort Collins precipitation data.

## 2.3 Point Process Representation

The theory of point process in the extreme value literature can be dated back to Pickands (1971). Pickands (1971) presented a way of describing the classic block maximum and GPD based extreme value model through point processes point view. The mathematical theory behind the point process model can be found by Leadbetter et al. (1983), Resnick (1987) and Embrechts et al. (1997). Leadbetter et al. (1983) provided the comprehensive coverage of the classic extreme value theory including theoretical argument for the point process properties. Resnick (1987) presented the mathematically fundamental and detailed explanation of the point process characterization especially the Poisson process for the extreme value theory based on independent and identically distributed random variables.

Coles (2001) interpreted that the point process characterization can be used to measure the probability of a given event occurring during a particular time period or the expected time for the next event occurring. The point process representation unifies the classic extreme value models like the GEV and GPD into one model. It is popular to adopt the point process representation for GPD to solve the problem of the dependence between the scale parameter and threshold.

### 2.3.1 Point Process Theory

A point process describes the distribution of random points within a space. Coles (2001) explains, the most popular way of characterizing the point process is to define a set of positive integer, $N(A)$, for each subset A $\subset \hat{A}$. The set of $\hat{A}$ could be multi-dimensional, $\hat{A} \subset \mathbb{R}^k$. $N(A)$ simply measures the number of points in the subset A. The characteristics of the point process $N$ is determined by the distribution of every $N(A)$ in the $\hat{A}$.

The expected number of points in a set A is referred to as the intensity measure $\Lambda(A)$,

$$\Lambda(A) = \mathrm{E}(N(A)).$$

The intensity measure $\Lambda$ has a density function $\lambda$, which is referred to as the intensity function,

$$\lambda(A;\theta) = \int_A \lambda(X_i;\theta)\,\mathrm{d}x.$$

A Poisson process on an Euclidean space is a point process $N$ with intensity measure $\Lambda$ which satisfies with the following conditions.

1 For any set of A $\subset \hat{A}$, $N(A)$ is Poisson random variable with mean $\Lambda(A)$.

2 For all disjoint set $A_1, \ldots, A_n$, $N(A_1), \ldots, N(A_n)$ are independent.

In other words, the number of points has a Poisson distribution and the number of points falling in different subset are independent. One of the inherent property of a Poisson process is that the points occur independently of each other. If the intensity function of a Poisson process $\lambda$ is constant, then the corresponding Poisson process is referred to as a homogenous Poisson process.

Let $X_1, \ldots, X_n$ be a series of independent and identically distributed random variables with common distribution function F, and let

$$P_n = \left\{ \left( \frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right); i = 1, \ldots, n \right\}$$

Coles (2001) explains that the first ordinate is to restrict to map the time axis to (0,1) and the second ordinate with the nonnormalizing constant $a_n$ and $b_n$ is to stabilize the behavior of $X_i$ when $n \to \infty$. It is known that such distribution of $P_n$ can converge in the limit to a two-dimensional non-homogeneous Poisson point process. For a sufficiently high threshold, the distribution of the two-dimensional point process $P_n$ is approximated in the limit by a non-homogeneous Poisson point process over the set A of $(0, 1] \times [u, \infty)$.

The intensity measure on the subregion $B = [t_1, t_2] \times (u, \infty)$ for the non-homogeneous Poisson point process is given by:

$$\Lambda(B) = \begin{cases} (t_1 - t_2)\left[1 + \xi\left(\dfrac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi} & \xi \neq 0 \\ (t_1 - t_2)\exp\left\{-\left(\dfrac{x-\mu}{\sigma}\right)\right\}_+ & \xi = 0 \end{cases}$$

The corresponding intensity function is then given by:

$$\lambda(x) = \begin{cases} \dfrac{1}{\sigma}\left[1 + \dfrac{\xi}{\sigma}(x-u)\right]^{-1/\xi - 1} & \xi \neq 0 \\ \dfrac{1}{\sigma}\exp\left[-\left(\dfrac{x-u}{\sigma}\right)\right] & \xi = 0 \end{cases}$$

It is common to include the scaling constant variable $n_b$ which is the number of years of observation to the intensity function. By doing this, the estimated $(\mu, \sigma, \xi)$ refers to GEV parameters for the number of years maxima estimate. The intensity measure can be therefore given by:

$$\Lambda(B) = \begin{cases} (t_1 - t_2) n_b \left[1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi} & \xi \neq 0 \\ (t_1 - t_2) n_b \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)\right\}_+ & \xi = 0 \end{cases}$$

It is not difficult to show that GPD can be derived from the point process representation.

$$Pr\left\{\frac{(X_i - b_n)}{a_n} > x \Big| \frac{(X_i - b_n)}{a_n} > u\right\} = \left[\frac{1 + \xi(\frac{x-\mu}{\sigma})}{1 + \xi(\frac{u-\mu}{\sigma})}\right]^{-1/\xi} = \left[1 + \frac{\xi(x-u)}{\sigma_u}\right]^{-1/\xi}$$

### 2.3.2 Likelihood Function

The points of a Poisson process in a set $A$, $N(A)$, are independent and identically distributed with a density of $\frac{\lambda(X_i;\theta)}{\Lambda(A;\theta)}$. Suppose the point process $P_n$ with the intensity measure $\Lambda(A;\theta)$ is on the space of $A = (0,1] \times [u, \infty)$. For the points, $x_1, \ldots, x_{nu}$ within the space of $A$, the likelihood can be derived from:

$$l_{pp}(\theta, x) = \exp\{-\Lambda(A;\theta)\} \frac{\Lambda(A;\theta)^{n_u}}{n_u!} \prod_{i=1}^{n_u} \frac{\lambda(X_i;\theta)}{\Lambda(A;\theta)},$$

$$l_{pp}(\theta, x) \propto \exp\{-\Lambda(A;\theta)\} \prod_{i=1}^{n_u} \lambda(X_i;\theta),$$

where $\Lambda(A;\theta) = \int_A \lambda(X_i;\theta)\, \mathrm{d}x$.

The general from of Poisson process likelihood can be derived as:

$$l(u, \mu, \sigma, \xi | x) \propto \begin{cases} \exp\left\{-n_b \left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \prod_{i=1}^{n_u} \frac{1}{\sigma} \left[1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right]^{-1/\xi - 1} & \xi \neq 0, \\ \exp\left[-n_b \exp\left(\frac{u-\mu}{\sigma}\right)\right] \prod_{i=1}^{n_u} \frac{1}{\sigma} \exp\left[-\left(\frac{x-u}{\sigma}\right)\right] & \xi = 0. \end{cases}$$

Similarly with the peaks over threshold method, the estimate obtained from the point process likelihood are based on the data that are greater than the threshold.

## 2.4 Kernel Density Estimation

### 2.4.1 Introduction

Nonparametric density estimation plays an important role in analyzing data. Unlike parametric models, nonparametric estimation technique can be used for effectively describing the complex data structure such as bi-modality or multi-modality without making prior assumptions about that they exist and the form they take. Nonparametric density estimation doesn't assume a specific functional form for the population distribution, only that it is sufficiently smooth, which provides more flexibility in estimating the density of the underlying dataset. The subject of nonparametric density estimation was brought to prominence by Rosenblatt (1956). One of the most popular nonparametric techniques is kernel density estimation. There has been intensive discussion about kernel density estimation in the literature, for example Rosenblatt (1956), Parzen (1962), Watson and Leadbetter (1963) and Resnick (1987). A classic text book on the topic is Wand and Jones (1995).

### 2.4.2 The Univariate Kernel Density Estimator

The traditional kernel density estimator is defined as:

$$\hat{f}(x; \mathbf{x}, h) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - \mathbf{x}_i}{h}\right),$$

where $K(.)$ is the kernel function usually defined to be a symmetric and unimodal probability density function and $h$ is the bandwidth parameter. The kernel function usually meets the following conditions: $K(x) \geq 0$ and $\int K(x)\,\mathrm{d}x = 1$. Another formula for the kernel estimator can be obtained by including scale kernel $K_h(y) = h^{-1}K(y/h)$

$$\hat{f}(x; \mathbf{x}, h) = n^{-1} \sum_{i=1}^{n} K_h\left(x - \mathbf{x}_i\right)$$

Wand and Jones (1995) interpret that the value of the kernel estimate at the point $x$ is simply the average of the $n$ kernel ordinates at that point. Various kernel functions have been proposed in the literature such as uniform, biweight, normal and other functions. The choice of the kernel function is not limited to some probability density functions. In this case, the resulting density estimate will not be a density any more but can be renormalized to be a density. The normal probability density function is a popular choice for the kernel. When the normal distribution is suggested to be the kernel function, the bandwidth $h$ will play the role of the standard deviation in the normal distribution which controls the spread of the kernel.

Wand and Jones (1995) pointed out that the choice of the kernel function is less critical than the choice of the bandwidth. There is always a trade-off between the bias of the resulting estimator and corresponding variance. If the bandwidth is too low, the estimate is very spiky and thus increasing the variance in the estimate. On the other hand, if the bandwidth is too high, the kernel estimation is too smooth leading to bias and it fails to provide adequate estimation for the true structure of the underlying density. Thus a large bandwidth could lead to a high bias but low variance of the estimator.

### 2.4.3 Bandwidth Selection

There are several common approaches of choosing bandwidth $h$ in the literature. Silverman (1986) looks at varies method to choose the bandwidth parameter. A simple and natural way of choosing the smoothing parameter is to plot several possible parameter value and choose the best fit to the underlying the density. The process of examining the a range of possible smoothing parameter candidates can be inefficient, time-consuming and subjective. Another method of selecting the bandwidth is to minimize some error criteria such as the mean squared error(MSE), mean integrated squared error (MISE), mean integrated absolute error(MIAE) and other criteria. The detail of bandwidth selection is beyond the scope of this thesis. Simple rules of thumb for selecting the bandwidth are based on various measures of performance, for example by minimizing asymptotic mean integrated squared error (AMISE). When the normal distribution has been used as the kernel function, the optimal choice of bandwidth in terms of minimising the AMISE, for a normal population is given by Silverman (1986):

$$h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5},$$

where $\hat{\sigma}$ is the sample standard deviation and $n$ is the sample size.

### 2.4.4 Boundary Bias and Consistency

One inherent drawback of the kernel density estimator is that it has no prior knowledge about where the boundary should be. It is better to use kernel density estimator to fit some unbounded densities. When kernel density estimator fits to the bounded density, it doesn't achieve the same level of accuracy at the boundary compared to the bounded density. Jones (1993) finds that kernel density estimator causes more bias at the boundary than the interior point. Kernel density estimation simply assigns the probability mass outside the support and so it suffers from bias at or near the boundary.

Let $K$ be a symmetric kernel function with support [-1,1] and the bandwidth is denoted by $h$ as usual. Suppose $x = ph$, $p$ is a function of $x$. For any $x \geq h$, or $p \geq 1$, there is no overspill of the contributions to the boundary. For an interior point, or any $x \geq h$, if the first and second derivatives of $f$ are continuous, the usual mean and variance can be derived:

$$E\left\{\hat{f}(x)\right\} \simeq f(x) + \frac{1}{2}h^2 a_2(p) f''(x).$$

At or near the boundary, for any $x < h$,or $p < 1$, the usual mean and variance can be derived:

$$E\left\{\hat{f}(x)\right\} \simeq a_0(p) f(x) - h a_1(p) f'(x) + \frac{1}{2}h^2 a_2(p) f''(x),$$

$$V\left\{\hat{f}(x)\right\} \simeq (nh)^{-1} b(p) f(x),$$

where

$$a_l(p) = \int_{-1}^{\min\{p,1\}} u^l K(u)\,\mathrm{d}u \quad \text{and} \quad b(p) = \int_{-1}^{\min\{p,1\}} K^2(u)\,\mathrm{d}u.$$

The coefficient of the first term in the mean of the $\hat{f}(x)$ is $a_0(p)$, which can be interpreted as the probability mass outside the boundary. Thus kernel density estimation is not a consistent estimator at the boundary. If the coefficient of the first term in the mean of the $\hat{f}(x)$ is one, then the estimator is regarded as a consistent estimator. Jones (1993) points out that the boundary point has the bias of the order $O(h)$ while the interior point has the bias of the order $O(h^2)$.

Consequently, the kernel density estimator is more biased at or near the boundary compared with interior point. A boundary corrected approach is needed to achieve $O(h^2)$ bias everywhere. Jones (1993) presented two popular ways of achieving reduced bias at the boundary. The first approach is to use a local re-normalization technique to make the estimator become consistent at the boundary. By dividing $\hat{f}(x)$ by $a_0(p)$, the resulting estimator $\bar{f}_N(x)$ is defined as:

$$\bar{f}_N(x) = \hat{f}(x)/a_0(p),$$

where $a_0(p) = \int_{-1}^{\min\{p,1\}} K(u)\,\mathrm{d}u$. The resulting estimator successfully captures that the kernel probability mass within the boundary. $\bar{f}_N(x)$ has become a consistent estimator. This local re-normalization technique ($\bar{f}_N(x)$) still gives and $o(h)$ bias at the boundary. Another way of reducing the bias is to reflect the data at the boundary, with the estimator given by:

$$\bar{f}_R(x) = \hat{f}(x) + \hat{f}(-x)$$

It is proven that $\bar{f}_R(x)$ becomes a consistent estimator but it doesn't achieve $O(h^2)$ bias at the boundary. Through the theoretical comparison, both approaches can overcome the consistency issues. In general, the refection approach outperforms the re-normalization approach as the refection approach suffers a smaller bias compared with the re-normalization approach. One of the fundamental issues with these two estimators is that they all still suffer from the boundary bias with an order of $O(h)$.

### 2.4.5 Boundary Correction Estimator

Jones (1993) suggested a generalized jackknifing idea to obtain the $O(h^2)$ bias either in the interior or the boundary. Let $K$ be a general kernel function with the support of $[-S_K, S_K]$. The idea is to take a linear combination of $K$ with some other function $L$ which is closely related with function $K$. The corresponding kernel will have the following properties:

1. $a_o(p) = \int_{-1}^{\min\{p,S_k\}} K(u)\,\mathrm{d}u = 1$
2. $a_1(p) = \int_{-1}^{\min\{p,S_k\}} uK(u)\,\mathrm{d}u = 0$

where $p$ is a function of $x$, $p = \frac{x}{h}$. The first property is to make kernel mass within the boundary and the second property is to make the first moment of each kernel center about the data point. Let $L$ be a kernel function with support$[-S_L, S_L]$. The corresponding kernel density estimator is defined as:

$$\check{f}(x) = \frac{1}{n}\sum_{i=1}^{n} L_h\left(x - \mathbf{x}_i\right)$$

Like the standard kernel estimator, another formula for $\check{f}(x)$ can be obtained as $L_h(x - \mathbf{x}_i) = h^{-1}L((x - \mathbf{x}_i)/h)$. By dividing $c_0(p)$, $\check{f}(x)$ can be re-normalized as usual,

$$\tilde{f}(x) = \check{f}(x)/c_0(p) \quad \text{where} \quad c_l(p) = \int_{-S_L}^{\min\{p,S_L\}} u_l L(u)\,\mathrm{d}u.$$

Generalized jackknifing method undertakes such linear combination,

$$\dot{f}(x) \equiv \alpha_x \bar{f}_N(x) + \beta_x \check{f}(x).$$

The following linear combination is able to allow $O(h^2)$ bias either interior or boundary point, provided that $c_1(p)a_0(p) \neq a_1(p)c_0(p)$,

$$\alpha_x = c_1(p)a_0(p)/\left\{c_1(p)a_0(p) - a_1(p)c_0(p)\right\},$$
$$\beta_x = -a_1(p)c_0(p)/\left\{c_1(p)a_0(p) - a_1(p)c_0(p)\right\}.$$

It should be noticed that such boundary corrected kernel density estimator doesn't automatically integrate to unity, which requires further re-normalizing the estimator. A popular choice for $L(.)$ is to set $L(x) = xK(x)$. There is a linear relationship between the $L(.)$ and $K(.)$. Such choice results in a simple boundary corrected kernel $(l_x+m_x x)K(x)$,where $l_x = a_2(p)/\left\{a_2(p)a_0(p) - a_1^2(p)\right\}$ and $m_x = -a_1(p)/\left\{a_2(p)a_0(p) - a_1^2(p)\right\}$. Even though there are various boundary corrected estimators using generalized jackknifing technique, there are little differences in the performance. Jones (1993) suggests the linear boundary corrected estimator is a good choice.

The associated boundary corrected kernel density estimator is therefore defined as:

$$\dot{f}(x;\mathbf{x},h) = \frac{1}{nh}\sum_{i=1}^{n} K_L\left(\frac{x-x_i}{h}\right),$$

where $K_L$ is the kernel function,

$$K_L(x) = \frac{(a_2(p) - a_1(p)x)K(x)}{a_2(p)a_0(p) - a_1^2(p)}.$$

### 2.4.6 Non-Negative Boundary Correction Estimator

As mentioned earlier, there are several different choices for $L(.)$ kernel, which will lead to different boundary corrected kernel estimators. The major drawback with these generalized jackknifing estimators is the propensity to take the negative values at the boundary. Jones and Foster (1996) propose a simple non-negative boundary correction estimator, which is a combination of boundary corrected estimator and scaled standard kernel density estimator. The resulting estimator is defined as:

$$f_{BC}(x;\mathbf{x},h_{BC}) \equiv \bar{f}(x;\mathbf{x},h_{BC})\exp\left\{\frac{\dot{f}(x;\mathbf{x},h_{BC})}{\bar{f}(x;\mathbf{x},h_{BC})} - 1\right\}.$$

This is a combination of boundary corrected estimator $\dot{f}(x;\mathbf{x},h_{BC})$ and scaled standard kernel density estimator,

$$\bar{f}(x;\mathbf{x},h_{BC}) = \frac{\hat{f}(x;\mathbf{x},h)}{a_0(p)}.$$

Jones and Foster (1996) provide mathematical and theoretical justification of this estimator. They have shown that this estimator achieves the desirable property of obtaining the same order of bias between the interior point and near the boundary.

### 2.4.7 Likelihood Inference for the Bandwidth

The likelihood inference for kernel bandwidth was proposed by Habbema et al. (1974) and Duin (1976). The likelihood degenerates at zero bandwidth, leading to point mass at each unique datapoint. To avoid this degeneracy issue, they propose a cross-validation likelihood

$$L(\lambda|\mathbf{X}) = \prod_{i=1}^{n} \frac{1}{(n-1)} \sum_{\substack{j=1 \\ j\neq i}}^{n} K_\lambda(x_i - x_j)$$

This cross-validation likelihood is used in the Bayesian inference MacDonald et al. (2011) and MacDonald et al. (2013). Habbema et al. (1974) and Duin (1976) indicate that this cross-validation likelihood doesn't provide a good fit to heavy tail distribution. On the other hand, the standard kernel density estimator also suffers unavoidable boundary bias.

# Chapter 3

# Literature Review in Extremal Mixture Models

## 3.1 Background

Extreme value theory has been used to develop models for the behavior of the distribution function in the tail and has been used as reliable basis for extrapolation. In traditional GPD application, the non-extreme data below the threshold will be discarded and only those inherently scarce extremal data will be fitted by the GPD models.

The motivations for ignoring the non-extreme data can be summarized as follows:

1 Extreme and non-extreme data are generally caused by different processes. If the unusual and rare events are of concern, there is little reason that we should consider the non-extreme events.

2 The GPD model itself provides a mathematically and statistically justifiable model for examining tail excesses. It is sometimes hard to combine the distribution below the threshold with tail model as the choice of bulk model depends on application and inferences for mixture model can be challenging and problematic.

3 The bulk distribution has not contributed much in determining the tail behavior. It is arguable to consider bulk fit which may compromise examination of the tail fit.

For the last decade, there has been increasing interests in extreme value mixture models, which provide for an automated approach both for estimation of the threshold, quantification of the uncertainties that result from the threshold choice and use all the data for parameter estimation in the inference.

The extreme value mixture models typically have two components :

1 a model for describing all the non-extreme data below the threshold (as many data will be below the threshold, we refer to this model as bulk model);

2 a traditional extreme value threshold excess model for modeling data above the threshold to describe the tail of the population distribution (we refer to this model as the tail model).

The tail model could be used to describe the upper, lower or both tails dependent on which extremes are of interest in the application. Generally speaking, the tail component is a flexible threshold model (e.g. GPD) and the bulk model is a parametric, semi-parametric, non-parametric form as appropriate for the application. Figures 3.1 gives an example of a extreme value mixture model without continuity constraint at the threshold and the dotted line indicates the threshold.

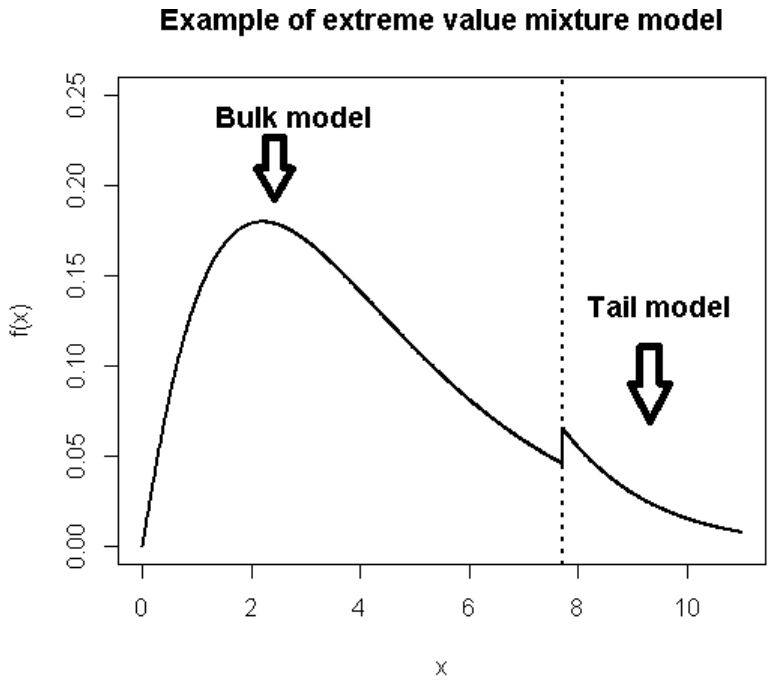



Figure 3.1: Example of an extreme value mixture model, where the dotted line is the threshold.

**Advantages**

1 Extreme value mixture models have the advantage of simultaneously capturing not only the bulk distribution below the threshold but the tail distribution above the threshold as well. It considers all the available data without wasting any information. One major goal of extreme value mixture models is to choose a flexible bulk model and tail model which fit to the non-extreme data as well as to the extreme data simultaneously.

2 The threshold selection is a challenging area in the extreme value literature. In some cases, it is likely that there are multiple suitable thresholds in some datasets. Different threshold choice may result in different tail behavior and corresponding with different return levels. Instead of applying the traditional fixed threshold approach, these mixture models treat the threshold as a parameter to be estimated. Furthermore, the uncertainty related with threshold estimation can be taken into account through the inference method. The major benefit of such models is not only providing a more natural way of estimating threshold, but also taking into account the uncertainties associated with threshold selection.

3 Great efforts have been made in developing different kinds of bulk models. Especially many different semi-parametric or non-parametric statistical techniques have been proposed for the

bulk model, which aims to describe the bulk model without assumptions on the model form. If the model is misspecified, the parametric bulk model doesn't offer much flexibilities compared with semi-parametric or non-parametric bulk model.

**Drawbacks**

Some common drawbacks of extreme value mixture model as follows:

1 The mixture model properties are not well understood and still need to be further investigated. Unfortunately, these mixture models are not commonly available in any existing software for practitioners to apply and develop further.

2 Another danger with applying these models is that they assume no strong influences between the bulk fit and tail fit. The bulk and tail components are not fully independent as they share the threshold with each other. If the bulk fit strongly affects the tail fit, the threshold will of course be subsequently affected. In this case, it is not clear that such a mixture model is still able to provide an reliable quantile estimates. Thus, the robustness between the bulk fit and tail fit are of great concern. There is a lack of sensitive analysis of robustness issue of the bulk and tail distribution in the literature although some discussions about robustness are available. A more detailed and comprehensive study is needed.

3 Some mixture models do not have continuous density function at the threshold. They normally have a discontinuity at the threshold and whether this is a problem for tail quantile estimation or whether the extra continuity constraints would improve the model performances are still not clear in the literature. In general, the advantage of a model without the continuity constraints is that it is more flexible compared to the model with the continuity constraints. On the other hand, it is physically sensible to have a continuous density function. However, by imposing the continuity constraints over the bulk and tail distribution, it has potential impact on threshold estimation. Thus the robustness of bulk fit and tail fit are of concern.

## 3.2 Extremal Mixture Models

This section will review some recent extreme value mixture models in the literature with some discussions of their properties.

### 3.2.1 Behrens et al. (2004)

Behrens et al. (2004) developed an extreme value mixture models including a parametric model to fit the bulk distribution and GPD for the tail distributions. The threshold is estimated as a parameter by splicing the two distributions at this point. The bulk model could be any parametric distributions, such as, gamma, Weibull or normal. In their paper, they use a (truncated) gamma for the bulk distribution.

They have mentioned to consider other parametric, semi parametric and nonparametric forms distribution to fit the bulk distribution and their model can be extended to a mixture distribution for any observations below the threshold. The distribution function of their model is defined as:

$$F(x|\eta, u, \sigma_u, \xi) = \begin{cases} H(x|\eta) & y \leq u; \\ H(u|\eta) + [1 - H(x|\eta)]\, G(x|u, \sigma_u, \xi) & y > u, \end{cases}$$

where $H(x|\eta)$ could be gamma, Weibull or normal distribution function and $G(.|u, \sigma_u, \xi)$ represents the GPD distribution function. The value of $1 - H(x|\eta)$ takes the place of the tail fraction $\phi_u = Pr(X > u)$ in the usual extreme tail modeling.

**Advantages: Behrens et al. (2004)**

1 This model is very straightforward and flexible and is the simplest of the extreme value mixture models. All available data can be used for estimating the parameters including the threshold which is treated as parameter. The most beneficial property of such model is to able to consider threshold estimation and corresponding uncertainty in the inference.

2 Bayesian inference is used throughout, which takes advantage of expect prior information to compensate the inherent sparse extreme data. Another benefit of using Bayesian inference is that entire parameter posterior distribution is available. For some certain applications from traditional graphic threshold selection method, different thresholds can be suggested by researchers. These multiple threshold choice could lead to a bi-modal or multi-modal posterior for the threshold. In this case, computational Bayesian approach is the ideal way to cope with difficulties with multiple local modes in the likelihood inference.

**Disadvantages: Behrens et al. (2004)**

1 One common drawback of such model is that the lack of a constraint of continuity at the threshold. MacDonald et al. (2011) indicated that uncertainty with threshold choice has strong localised effect on the estimates close to threshold. If there is a spurious peak in the tail, the threshold is very likely to be chosen close to this location. As a result, the threshold could be influenced by some spurious peaks in the tail due to natural sample variation and these features are common in many applications.

2 One disadvantage of this model is that the Bayesian posterior inference becomes less efficient by ignoring dependence between the threshold and GPD scale parameter and treating the prior for the threshold and GPD parameters as independent. One way of overcoming such issues is to use point process representation for the GPD model.

3 Another possible drawback with their method is that the parametric form used for bulk distribution could suffer if the model is misspecified particularly as the tail fraction is estimated using the bulk distribution function. The parameter estimation for the tail model of such mixture models could be sensitive to the choice of the bulk model. For instance, if the bulk distribution is poorly fit, then threshold may be effected. The tail fit could be subsequently influenced as well. There are no publications to discuss the model misspecification of such extreme value mixture model.

### 3.2.2 Carreau and Bengio (2008)

Carreau and Bengio (2008) proposed the hybrid Pareto model by splicing a normal distribution with generalized Pareto distribution (GPD) and set continuity constraint on the density and on its first derivative at the threshold.

The fundamental idea underlying these constraints are to make this hybrid Pareto model not only continuous but also smooth. Initially, this model has 5 parameters including the normal distribution mean parameter $\mu$, standard deviation parameter $\beta$, the threshold $u$, the GPD shape parameters $\xi$ and scale parameter $\sigma_u$. By setting two constraints on the threshold, these five parameters could be reduced to three. They use Lambert W function to obtain the scale $\sigma_u$ and threshold $u$ in terms of ($\mu$, $\beta$ and $\xi$), thus only these three parameters are needed for their model. This is an arbitrary choice for them to rewrite scale $\sigma_u$ and threshold $u$ in terms of the remaining parameters and this choice is for analytic tractability. It is not clear that it would make some differences to the inference by fixing any other pair of the parameters.

The density function of the hybrid Pareto model is defined as:

$$f(x|\mu,\beta,\xi) = \begin{cases} \dfrac{1}{\tau}h(x|\mu,\beta) & x \leq u, \\ \\ \dfrac{1}{\tau}g(x|u,\sigma_u,\xi) & x > u, \end{cases}$$

where $\tau$ normalizes to make it integrate to one and is given by,

$$\tau = 1 + \frac{1}{2}\left[1 + Erf\left(\sqrt{\frac{W(z)}{2}}\right)\right].$$

The $\tau$ is the usual normalising constant, where the 1 comes from the integration of the unscaled GPD and second term is from the usual normal component. $h(.|\mu,\beta)$ is the Gaussian density function with mean $\mu$ and standard deviation $\beta$ and GPD density is defined as $g(.|u,\sigma_u,\xi)$ with shape $\xi$ and scale $\sigma_u$.

The distribution function (CDF) of the hybrid Pareto model is defined as:

$$F(x|\mu,\beta,\xi) = \begin{cases} \dfrac{1}{\tau}H(x|\mu,\beta) & x \leq u, \\ \\ \dfrac{1}{\tau}\left[H(u|\mu,\beta) + G(x|u,\sigma_u,\xi)\right] & x > u, \end{cases}$$

where $H(.|\mu,\beta)$ is the Gaussian distribution function and GPD distribution function is defined as $G(.|u,\sigma_u,\xi)$

**Advantages: Carreau and Bengio (2008)**

1 This is the first approach in the literature which attempts to set continuity between the bulk and tail component at the threshold. It has advantage of using two less parameters. By imposing continuity at the threshold, this potentially reduces the local degree of freedom problem in the mixture model.

**Disadvantages: Carreau and Bengio (2008)**

1 The main drawback of hybrid Pareto model is that it performs very poorly in practice. Recalled from Section 1.1.1.2, the GPD is defined conditionally on being above the threshold. The value of tail fraction $\phi_u$ is another parameter required when using the GPD. Carreau and Bengio (2008) don't include this parameter in their model. The simulation study in Chapter 5 will show it only works well for some asymmetric heavy tailed distributions. The GPD is not treated as a conditional model for the exceedances. The unscaled GPD is then spliced with normal distribution and the model parameters have to adjust for the lack of tail fraction(A modified version of hybrid Pareto model which improves the performance of the original hybrid Pareto model will be discussed in Section 3.3.2).

2 A big danger with this approach is that parameters are highly dependent on each other. The model formulation has restricted where the threshold can go. The potential position for the threshold is limited by the three free parameters. Thus the bulk fit will likely have a greater impact on the tail fit compared to any mixture model. In short, there will likely be a lack of robustness between the bulk and tail model.

3 It is still unclear that whether such a model provides sufficient flexibility to cope with model misspecification. The bulk fit strongly influence the tail fit, and subsequently the threshold. This question is explored in the applications in Chapter 6.

**Carreau and Bengio (2009)**

Carreau and Bengio (2009) extended this model to a mixture of hybrid Pareto distributions to account for the unsatisfactory model performance. There are two different ways of model formulation. The mixture of hybrid Pareto distributions can be formulated as either a finite normal distribution with GPD or a finite number of hybrid Pareto distributions. They choose to use the second representation to overcome undesirable model fits in applications. The tail is therefore estimated by finite number of GPD's. They show that the mixture of hybrid Pareto distribution performs better when the tail is fat-tailed or heavy tailed. Kang and Serfozo (1999) proved that the tail behavior of the mixture model will be dominated by the component which has the heaviest tail.

The main attraction for this new mixture model is the ability to handle asymmetry, multi-modality and heavy tail distributions. Although this new approach improves performance, it results in several issues. The performance of this model when the underlying distribution is short tail or exponential tail is not clear. Some drawbacks are naturally very similar with the original hybrid Pareto model, so are not detailed again. The tail is asymptotically dominated by the component which has the strongest impact on the tail, but the rest of the components will sub-asymptotically influence the tail behavior as well. Apart from that, model misspecification is another crucial issue for this mixture model. The choice of optimal number of components in the mixture model is not clear in Carreau and Bengio (2009).

### 3.2.3 Frigessi et al. (2002)

Frigessi et al. (2002) suggested a dynamically weighted mixture model by combining the Weibull for the bulk model with GPD for the tail model. Instead of explicitly defining the threshold, they use a Cauchy cumulative distribution function for the mixing function to make the transition between the bulk distribution and tail distribution. The bulk model is assumed to be light-tailed, so the GPD should dominate in the upper tail.

Maximum likelihood estimation is used throughout. The mixing function allows bulk model to dominate the lower tail and GPD to dominate the upper tail. It should be noted that the upper tail is not purely determined by the GPD model. When GPD asymptotically dominates the upper tail, the bulk model sub-asymptotically dominates the upper tail as well. They mentioned that if Heaviside function is used as the mixing function, then this mixture model encompasses the Behrens et al. (2004) as a special case. The model is designed for positive data set only and the full data set is used for inference for GPD component. Hence, the threshold is assumed to be zero.

The density function of the dynamic mixture model is given by:

$$f(x|\theta,\beta,u,\sigma_u,\xi) = \frac{[1-p(x|\theta)]\,h(x|\beta) + p(x|\theta)g(x|u=0,\sigma_u,\xi)}{Z(\theta,\beta,u=0,\sigma_u,\xi)},$$

where $h(x|\beta)$ denotes the Weibull density and $Z(\theta,\beta,u=0,\sigma_u,\xi)$ denotes the normalizing constant to make the density integrate to one.

The mixing function $p(x|\theta)$ is a Cauchy distribution function with location parameter $\mu$ and scale parameter $\tau$ given by:

$$p(x|\theta) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-u}{\tau}\right), \theta = (\mu,\tau), \mu, \tau > 0.$$

**Advantages: Frigessi et al. (2002)**

1 This model bypasses the usual threshold choice as threshold is not explicitly defined as parameter. The intuition behind this model is to utilise the mixing function to overcome the difficulties associated with threshold selection as the threshold selection has great impact on parameter estimation.

2 This is the first approach in the literature which uses a continuous transition function to connect the bulk with tail model. The mixing function dynamically determines the weight for the bulk model and tail model and makes the mixture density continuous. The idea of such transition function attempt to overcome the lack of continuity at the threshold.

**Disadvantages: Frigessi et al. (2002)**

1 Even if the mixing function makes continuous transition between two components, two issues have been pointed out by Scarrott and MacDonald (2012). The Cauchy CDF has been used as a smooth transition function. The spread out of transition effect between the two components is measured by the scale parameter of Cauchy distribution. The larger scale parameter, the more spread out of the transition effect. In spite of the optimization difficulties, a solution based on maximum likelihood estimation is sometimes obtainable when scale parameter of the Cauchy distribution is close to zero. If the scale parameter of Cauchy distribution approaches to zero, the beneficial smooth transition effects could be lost in the inference procedure.

2 The lack of robustness between bulk and tail model is the fundamental issue. Not only does Cauchy CDF contribute the bulk fit, but it also has an impact on tail fit as well. It has a continuous transition effect between the two components. The mixing function is able to give high weight for bulk model in the lower range of the support and high weight for the GPD model in the upper tail. The GPD also contributes to the lower tail. As all three components in this model are likely to influence each other, this mixture model could suffer the lack of robustness of the bulk fit and tail fit.

### 3.2.4 Mendes and Lopes (2004)

Mendes and Lopes (2004) propose a data driven approach of an extreme value mixture model. The bulk distribution has been suggested to be normal distribution and GPD models are fitted to the two tails.

The distribution function is defined as:

$$F(x) = \phi_{u_l} G^{l}_{\xi_l,\sigma_l}(x - t_l) + (1 - \phi_{u_l} - \phi_{u_r}) H^{t_r}_{t_l}(x - t_r) + \phi_{u_r} G_{\xi_r,\sigma_r}(x - t_r),$$

where $\phi_{u_l}$ and $\phi_{u_r}$ are proportion of data in the lower tail and upper tail respectively. $G^{l}_{\xi_l,\sigma_l}(x)$ is the GPD distribution function for lower tail, provided that $G^{l}_{\xi_l,\sigma_l}(x) = 1 - G_{\xi_r,\sigma_r}(x)$. $H^{t_r}_{t_l}$ denotes the normal distribution function truncated at $t_l$ and $t_r$.

They have outlined following estimation procedure:

Step 1. Data should be standardized and put in order. They apply some shift-scale transformation for all the data using the median ($med$) and the median absolute deviation ($mad$).

Step 2. Select some certain value of proportions data in both tails and compute empirical threshold respectively. Two groups of empirical thresholds $t_l$ and $t_r$ are calculated due to proportion of data $\phi_{u_l}$ and $\phi_{u_r}$.

Step 3. The L-moments method is used to estimate the two tail GPD parameters. $(\xi_l, \sigma_l)$, $(\xi_r, \sigma_r)$ based on different thresholds $t_l$ and $t_r$.

Step 4. Calculate the $\phi_{u_l} = H(t_l)$ and $\phi_{u_r} = 1 - H(t_r)$, where $H(.)$ is the normal distribution.

Step 5. Computer the log-likelihood of the mixture density using the set of parameter estimates $\{\phi_{u_l}, \phi_{u_r}, t_l, t_r, \xi_l, \sigma_l, \xi_r, \sigma_r\}$ .

Step 6. Choose the pair $(\phi_{u_l}, \phi_{u_r})$ maximizes the log-likelihood of data.

**Advantages: Mendes and Lopes (2004)**

1 The main benefit of such a model is that it provides some intuitions to consider two tail models rather than a single tail model.

2 Another benefit of this mixture model is the flexibility from both tails makes choice of the bulk model less sensitive, but there are 4 more parameters to estimate.

**Disadvantages: Mendes and Lopes (2004)**

1 Unlike the usual approach, they use robust technique to standardize the data before the fitting procedure. They claim that this robust standardized procedure which pushes the bulk data more close to the center is beneficial to distinguish between extreme data and non-extreme data. However, a linear rescaling used so this argument is flawed.

2 Another drawback with their method is that the uncertainty associated with threshold is ignored in the inference.

### 3.2.5 Zhao et al. (2010)

Zhao et al. (2010) or Zhao et al. (2011) consider a mixture model with lower tail and upper tail using GPD, which is so called as two tail model. The bulk distribution has been suggested as normal distribution in this two tail mixture model. This model is similar with the work of Mendes and Lopes (2004) but the method of parameter estimation is different. Bayesian inference is used throughout.

Let $\Theta = (u_l, \sigma_{u_l}, \xi_l, \mu, \beta, u_r, \sigma_{u_r}, \xi_r)$ be the parameter vector. The distribution function is defined as:

$$F(x|\Theta) = \begin{cases} \phi_{u_l}\left\{1 - G(-x| - u_l, \sigma_{u_l}, \xi_l)\right\} & x < u_l, \\ H(x|\mu, \beta) & u_l \leq x \leq u_r, \\ (1 - \phi_{u_r}) + \phi_{u_r} G(x|u_r, \sigma_{u_r}, \xi_r) & x > u_r, \end{cases}$$

where $\phi_{u_l} = H(u_l|\mu, \beta)$ and $\phi_{u_r} = 1 - H(u_r|\mu, \beta)$ and $H(.|\mu, \beta)$ is the normal distribution function with mean $\mu$ and standard deviation $\beta$. $G(.| - u_l, \sigma_{u_l}, \xi_l)$ and $G(.|u_r, \sigma_{u_r}, \xi_r, )$ are GPD distribution functions for lower tail and upper tail. Bayesian inference is used throughout.

It is possible to define $\phi_{u_l} = P(X < u_l)$ and $\phi_{u_r} = P(X > u_r)$. The corresponding distribution function is given by:

$$F(x|\Theta) = \begin{cases} \phi_{u_l}\left\{1 - G(-x| - u_l, \sigma_{u_l}, \xi_l)\right\} & x < u_l, \\ \phi_{u_l} + (1 - \phi_{u_l} - \phi_{u_r})\dfrac{H(x|\mu, \beta) - H(u_l|\mu, \beta)}{H(u_r|\mu, \beta) - H(u_l|\mu, \beta)} & u_l \leq x \leq u_r, \\ (1 - \phi_{u_r}) + \phi_{u_r} G(x|u_r, \sigma_{u_r}, \xi_r) & x > u_r, \end{cases}$$

**Advantages: Zhao et al. (2010)**

1 A major improvement of this approach is that the choice of bulk distribution becomes less influential on both tail fit. The advantage of this mixture model is to provide more flexibility in fitting asymmetry data which is very common in many applications. The both tails are less sensitive with choice of bulk model, especially compared with single tail model.

**Disadvantages: Zhao et al. (2010)**

1 The main drawback with this method is that dependence between the threshold and GPD scale parameter is ignored, which makes the Bayesian inference less efficient. Such an issue can be simply solved by using a different representation of the GPD model (point process representation in Section 2.3).

2 Model misspecification is another issue which associated with this model. Although this model is likely less sensitive to model misspecification, it is not clear that this mixture model is able to handle misspecification issue.

3 This model has 8 parameters to be estimated.

### 3.2.6 Tancredi et al. (2006)

Tancredi et al. (2006) constructed a mixture model by combining non-parametric density estimation using unknown number of uniform distribution for the bulk, with GPD model using PP representation. In order to overcome the lack of a flexible bulk model, they adopted the uniform density estimator of Robert and Casella (2004). The dimension of parameter spaces change on account of unknown number of uniform densities. Consequently, the standard Markov chain Monte Carlo algorithm is not applicable in this case. Instead, they work in Bayesian framework and use the reversible jump Markov chain Monte Carlo algorithm of Green (1995) to deal with changing parameter dimension.

As $u_0$ is a very low threshold, any data beyond $u_0$ are i.i.d observations from the density function by:

$$f(x) = \begin{cases} (1-\phi_u)h(x|\phi_u^{(x)}, a^{(k)}, u) & u_0 < x < u, \\ \\ \phi_u g(x|u, \sigma_u, \xi) & u \leq x < \infty, \end{cases}$$

where $g(.|u, \sigma_u, \xi)$ is the GPD density. Let $\phi_u = Pr(X > u|u_0)$ and $\phi_u$ is the probability that an observation from the distribution is greater $u$ conditional on being greater than very low threshold $u_0$. $h(x|\phi_u^{(x)}, a^{(k)}, u)$ is piecewise density on the support of $[u_0,u)$ where the number of $k$ is not known. The $k$ step piecewise density is given by:

$$h(x|\phi_u^{(x)}, a^{(k)}, u) = \sum_{i=1}^{m} \phi_{u_i} I_{[a_{i+1}, a_i)}(x),$$

where $a_1 = u_0$, $a_{k+1} = u$, and $\sum_{i=1}^{m} \phi_{u_i}(a_{i+1} - a_i) = 1$.

If we let $\phi_{u_i} = \frac{p_i}{a_{i+1} - a_i}$ and $\sum_{i=1}^{m} p_i = 1$, the $k$ step piecewise density can be represented as:

$$h(x|\phi_u^{(x)}, a^{(k)}, u) = \sum_{i=1}^{m} p_i U[a_{i+1}, a_i).$$

The $h$ is essentially a mixture of $k$ uniform distributions where the $k$ is not known.

**Advantages: Tancredi et al. (2006)**

1 This is the first approach in the literature to use PP representation for the extreme value tail. There are several advantages of using point process likelihood for the threshold excess model. The main advantage of adopting point process representation is to remove the dependence between the threshold $u$ and scale parameter $\sigma$. The dependence in the GPD case makes posterior sampling less efficient and more challenging to achieve good chain convergence.

2 Not only do they firstly use PP representation to solve the dependence problem, but also they have attempted to use non-parametric density estimator for bulk model for the first time in the literature. The most beneficial property of such non-parametric technique is that tail model is more robust to bulk model compared with usual parametric bulk models.

**Disadvantages: Tancredi et al. (2006)**

1 The main drawback of this model is that the use of the reversible jump Markov chain Monte Carlo algorithm may cause computational challenges. Thompson et al. (2009) criticize that there is huge computational complexity of bulk model.

2 Another issue with this model is the required the hyper-parameter in the prior information. There are some hyper-parameters in the prior for the unknown uniform distribution. Tancredi et al. (2006) acknowledge that there is a lack of guidelines for selecting hyper-parameter in the prior. Thompson et al. (2009) also point out that the prior information which is needed to specify then is seldom available.

### 3.2.7 MacDonald et al. (2011)

MacDonald et al. (2011) constructed a standard kernel density estimator as the bulk with a GPD or PP tail model. The motivation of formulating such model is that they aim to provide a more flexible framework for extreme value analysis. The distribution function is defined as :

$$F(x|X, \lambda, u, \sigma_u, \xi, \phi_u) = \begin{cases} (1 - \phi_u)\frac{H(x|X,\lambda)}{H(u|X,\lambda)} & x \leq u, \\ (1 - \phi_u) + \phi_u \times G(x|u, \sigma_u, \xi) & x > u, \end{cases}$$

where $H(.|X,\lambda)$ is the distribution function of the kernel density estimator and $\lambda$ is the bandwidth.

MacDonald et al. (2011) also introduced a two tailed mixture model. This mixture model is constructed by splicing the standard kernel density estimator with two extreme value tail models. This two-tailed extreme value mixture model also overcomes the inconsistency in the cross-validation likelihood estimation of the bandwidth for heavy tailed distributions, see MacDonald et al. (2011) for details. The parameter vector is $\Theta = (X, \lambda, u_1, \sigma_{u_l}, \xi_l, \phi_{u_l}, u_2, \sigma_{u_r}, \xi_r, \phi_{u_r})$.

The distribution function of two tailed mixture model is given by:

$$F(x|\Theta) = \begin{cases} \phi_{u_l}\left\{1 - G(-x| - u_l, \sigma_{u_l}, \xi_l)\right\} & x < u_l, \\ \phi_{u_l} + (1 - \phi_{u_l} - \phi_{u_r})\dfrac{H(x|X,\lambda) - H(u_l|X,\lambda)}{H(u_r|X,\lambda) - H(u_l|X,\lambda)} & u_l \leq x \leq u_r, \\ (1 - \phi_{u_r}) + \phi_{u_r} G(x|u_r, \sigma_{u_r}, \xi_r) & x > u_r, \end{cases}$$

where $\phi_{u_l}$ and $\phi_{u_r}$ are estimated as the sample proportions less than threshold $u_l$ and above the threshold $u_r$ respectively. $G(-x| - u_l, \sigma_{u_l}, \xi_l)$ is the unconditional GPD function for $x < u_l$ and $G(x|u_r, \sigma_{u_r}, \xi_r)$ is the unconditional GPD function for $x > u_r$. These two unconditional GPD functions can be represented by corresponding PP representation to remove the dependence between the threshold and GPD scale parameter.

**Advantages: MacDonald et al. (2011)**

1 The main drawback of the cross validation likelihood based kernel bandwidth estimators is that is biased high in the face of heavy tail behavior. Bowman (1984) further mentions that the problem of poor performance is due to the inconsistent likelihood bandwidth parameter. They even address that the cross-validation likelihood inference also over smooth the bandwidth parameter when an outlier are observed from a distribution. This problem comes about due to the lack of separation of the highest (and lowest) order statistics for heavy tailed populations, thus no allowing the bandwidth to converge to zero as the sample size increases (see Duin (1976) for details). As these order statistics are in the tails, this problem is overcome by replacing the upper and/or lower tail with a GPD in these mixture models, so that the data in the tails (those highest and lowest order statistics) do not contribute to the cross-validation likelihood given by equation 8 in MacDonald et al. (2011). A key feature of this two tailed mixture model is that MacDonald et al. (2011) demonstrated that this model has managing to solve the inconsistency issue of kernel density bandwidth estimator. Form their simulation studies, this mixture model provide a good fit in bulk model, lower tail and upper tail model. By allowing both lower and upper tail, the likelihood based bandwidth estimation consistency issue is simply removed. The boundary bias is also effectively removed by including lower tail and upper tail model as these GPD models can automatically consider the finite bounded constraints.

2 If both lower tail and upper tail are of interest and the underlying process has a proper lower tail, this model is capable to provide a simple and flexible framework for most applications.

The bulk model is able to provide a very effective way of estimating structure of the data without assuming a particular parametric form. This model is unlike Mendes and Lopes (2004) and Zhao et al. (2010), it is not necessary to impose any parametric form restrictions for the bulk model. The choice of bulk model becomes less sensitive to the both tail models, which suggest that the bulk model has little impact on the tail model. Thus the three different components are fairly robust to each other.

3 MacDonald et al. (2011) utilise PP model representation for both tail models. The benefits of adopting PP model have been intensively discussed in Section 2.3.

**Disadvantages: MacDonald et al. (2011)**

1 One drawback of this two tailed mixture model is its assumption about the lower tail or the shape of the underlying distribution. In order to apply this model, the underlying distribution has to have a lower tail which decays to zero at the boundary. When there is a pole or shoulder at the lower tail, this model is obviously not appropriate to fit such distribution.

2 The major drawback of this model is the computational complexity, which results in a huge computational burden. Hence the MCMC takes a lot of the computational time.

### 3.2.8 MacDonald et al. (2013)

MacDonald et al. (2013) combine a boundary corrected kernel density estimator for bulk model sliced at the threshold with a point process representation of GPD tail model. The boundary correction kernel density estimator tries to reduce the inherent bias of the kernel density estimator. Many forms of boundary correction based kernel density estimators have been proposed in the literature. However, many of these estimators can take negative values at the boundary. MacDonald et al. (2013) utilise the non-negative boundary correction estimator which is developed by Jones and Foster (1996) to overcome the negative values near the boundary, which is detailed in Section 2.4.6.

The distribution function of boundary corrected mixture model is given by :

$$F(x|X,\lambda_{BC},u,\sigma_u,\xi,\phi_u) = \begin{cases} (1-\phi_u)\frac{H_{BC}(x|X,\lambda_{BC})}{H_{BC}(u|X,\lambda_{BC})} & x \leq u, \\ (1-\phi_u) + \phi_u \times G(x|u,\sigma_u,\xi) & x > u, \end{cases}$$

where $H_{BC}(x|X,\lambda_{BC})$ is the distribution function of the non-negative boundary corrected kernel density estimator. $\phi_u$ is estimated as the sample proportion of the data above the $u$.

**Advantages: MacDonald et al. (2013)**

1 The most beneficial property of adopting the non-parametric technique as the bulk model is that the bulk fit has no strong influence on tail fit. MacDonald et al. (2013) utilise the Turkey's sensitive curve to visualize the impact of the outlier to the kernel bandwidth parameter and this curve indicates that bandwidth parameter estimation is robust to the outlier. In other words, the bulk model is robust to the tail model.

2 The flexibility of this boundary corrected mixture model is a good feature. Unlike the two tailed kernel mixture model by MacDonald et al. (2011), there is no need to assume a proper lower tail with additional parameters. The boundary corrected mixture model can be applied when there is a proper tail, pole, and shoulder at the boundary.

**Disadvantages: Boundary Corrected Kernel GPD Model**

1 This model improves on the standard kernel density estimator used in MacDonald et al. (2011) for heavy tailed distributions by reducing boundary bias. However, this model is more complex compared to the standard kernel GPD model.

2 Boundary correction of the cross-validation likelihood causes a huge computational burden. This approach can be time consuming and computationally intensive when the sample size of the population distribution is large.

### 3.2.9 Nascimento et al. (2011)

Nascimento et al. (2011) considered a semi-parametric Bayesian approach by including a weighted mixture of gamma densities for the bulk distribution and GPD as the tail distribution. The number of gamma components in the bulk distribution is determined by BIC or DIC based statistics. As bulk model could include a number of gamma components, the bulk model of this mixture model is relatively flexible compared to Behrens et al. (2004) model. In order to overcome the identifiability problem due to a number of gamma densities in the model, they have adopted some parameter restrictions.

The density function of their model is defined as:

$$f(x|\theta, u, \sigma_u, \xi) = \begin{cases} h(x|\theta) & y \leq u, \\ [1 - H(u|\theta)]\, g(x|u, \sigma_u, \xi, ) & y > u, \end{cases}$$

where $H(.|\theta)$ is distribution function of the mixture of gamma and $g(x|u, \sigma_u, \xi)$ represents the GPD density function. $h(x|\theta)$ is the density function of the mixture of gamma and define by $h(x|\theta) = \sum_{i=1}^{k} p_i * h(x|\alpha_i, \eta_i)$. $\theta$ is denoted by the gamma shape parameter $\alpha = (\alpha_i, ..., \alpha_k)$ and gamma mean parameter $\eta = (\eta_i, ..., \eta_k)$. The shape parameter $\alpha$ and mean parameter $\eta$ can be any positive value. The weight component $p_i$'s can only take value between 0 and 1 with the sum of weight component $p_i$'s need to be 1 for obvious reasons.

**Advantages: Nascimento et al. (2011)**

1 The main advantage of such model is that it provides more flexibility for the bulk model compared to simple parametric forms. The bulk model includes a number of gamma densities and it has advantage of fitting some bi-modal or multi-modal distributions. The weighted mixture of gamma densities is to eliminate any constraint on the possible shape of the bulk model and eliminate the unimodal constraint. However, it uses up less degree of freedom compared to the non-parametric density estimator.

2 This mixture model is like Behrens et al. (2004) model, the density function may have a discontinuity at the threshold.

**Disadvantages: Nascimento et al. (2011)**

1 The concern of such model is whether it is still flexible enough when the model is misspecified. In particular, the robustness between the bulk fit and tail fit. It is clear that such mixture model is more flexible compared with parametric approach mixture model. However, it is still not clear that whether this mixture model can cope with model misspecification well than the parametric approach mixture model, which would be interesting to investigate such issues in the future.

### 3.2.10 Lee et al. (2012)

Lee et al. (2012) proposed a mixture mixture model which is made up of mixture of exponential distribution for the data below the threshold and GPD for the threshold excess. This mixture model is essentially an extension from Behrens et al. (2004), and similar with Nascimento et al. (2011). This mixture model is a special case of Nascimento et al. (2011) with different inference methods. Instead of directly applying the maximum likelihood estimation method, they adopt EM algorithm for parameter estimation. They adopt a mixture of two exponential distributions model to the usual Danish fire claim and a medical claim dataset.

The density function of their model is defined as:

$$f(x|p,\theta,u,\sigma_u,\xi) = \begin{cases} h(x|\theta) & y \leq u, \\ [1 - H(u|\theta)]\, g(x|u,\sigma_u,\xi,) & y > u, \end{cases}$$

where $H(.|\theta)$ is distribution function of the mixture of exponential and $g(x|u,\sigma_u,\xi)$ represents the GPD density function. $h(.|\theta)$ is the exponential density function and is given by:

$$h(x|\theta) = h(x|p_i,\lambda_i) = \sum_{i=1}^{m} p_i * \lambda_i \exp^{(-\lambda_i x)}$$

**Advantages: Lee et al. (2012)**

1 The benefit of this mixture model is that the bulk model may offer additional more flexibility compared with a single exponential distribution. Due to the limitation about the shape of exponential distribution, this mixture model is theoretically more flexible.

2 The main feature of this model is the multi-level procedure based EM algorithm. The fitting procedure can reduce the computational burden, which suggest that the EM algorithm estimator is more efficient than the ML estimator. Moreover, unlike the ML estimator, they also notice that the initial values are not sensitive with the EM algorithm. This could be an appealing feature of this inference approach.

**Disadvantages: Lee et al. (2012)**

1 The model may provide a somehow reasonable fit to some distribution which has a pole at the lower boundary and the upper tail is of interest. However, it seems that this model may not be appropriate for an underlying process which has a proper lower tail, a shoulder or there is zero at the boundary. Even though, a mixture of two exponential models can provide some flexibilities, it still has relatively restrictive shape.

## 3.3 New Model Extensions

This section will show two new mixture models that have been developed:

- hybrid Pareto with single continuity constraint,
- normal GPD with single continuity constraint model.

### 3.3.1 Hybrid Pareto Model with Single Continuity Constraint

Carreau and Bengio (2008) propose the hybrid Pareto model by splicing a normal distribution with GPD and set two continuity constraints at the threshold. They do not include usual scaling $\phi_u$ in their model. It is possible that the two continuity constraints are too strong assumptions for the hybrid Pareto model, which could be the reason why the hybrid Pareto model can perform poorly. However, I suspect the model formulation is also restrictive.

A similar model can be suggested by splicing a normal distribution with GPD and set a single continuity constraint on the density only at the threshold. The usual scaling $\phi_u$ is ignored, which is consistent with the original hybrid Pareto model. The 5 initial parameters have been reduced to 4 and the GPD scale parameter $\sigma_u$ is replaced by the other 4 parameters (normal mean parameter $\mu$, standard deviation parameter $\beta$, the threshold $u$, the GPD shape parameters $\xi$).

Let $h(.|\mu, \beta)$ and $g(.|u, \sigma_u, \xi)$ be the normal density function and GPD density function. The single continuous constraint on the density at the threshold implies:

$$h(u|\mu, \beta) = g(u|u, \sigma_u, \xi).$$

Rearrange the equation,

$$h(u|\mu, \beta) = \frac{1}{\sigma_u}.$$

The GPD scale parameter is replaced by,

$$\sigma_u = \frac{1}{h(u|\mu, \beta)}.$$

The density function of the hybrid Pareto model with single continuity constraint is therefore defined as:

$$f(x|\mu,\beta,u,\xi) = \begin{cases} \dfrac{1}{\tau}h(x|\mu,\beta) & x \leq u, \\ \\ \dfrac{1}{\tau}g(x|u,\sigma_u,\xi) & x > u, \end{cases}$$

The distribution function (CDF) of the hybrid Pareto model with single continuity constraint is defined as:

$$F(x|\mu,\beta,u,\xi) = \begin{cases} \dfrac{1}{\tau}H(x|\mu,\beta) & x \leq u, \\ \\ \dfrac{1}{\tau}\left[H(u|\mu,\beta) + G(x|u,\sigma_u,\xi)\right] & x > u, \end{cases}$$

where $H(.|\mu,\beta)$ is the Gaussian distribution function and GPD distribution function is defined as $G(.|u,\sigma_u,\xi)$. The $\tau$ is the usual normalising constant and $\tau = 1 + H(u|\mu,\beta)$, where the 1 comes from the integration of the unscaled GPD and second term is from the usual normal component.

The density and distribution function of the hybrid Pareto model with single continuity constraint are essentially same as the hybrid Pareto model except the single continuity constraint model has a one more parameter (threshold $u$). The advantages and disadvantages of this model are very similar with hybrid Pareto model, so they are not detailed again.

It will be interesting to see if this model provide sufficient flexibility to overcome the lack of the tail fraction $\phi_u$, which will be explored in the the simulation study in Chapter 5.

### 3.3.2 Normal GPD model with Single Continuity Constraint

Carreau and Bengio (2008) ignore the usual tail fraction term ($\phi_u = P(X > u)$) of the conditional GPD in their definition of the hybrid Pareto. This extension will take the extreme value mixture with any distribution for the bulk (e.g. normal) and GPD for the tail and require a continuity constraint at the threshold. This extension with a normal for the bulk is then directly comparable to that in Section 3.3.1, so the impact of the inclusion of the tail fraction can be explicitly examined..

The modified version of single continuity constraint model reduces the five parameters to four, by replacing scale parameter $\sigma_u$ from GPD model using the other four parameters by a single continuity constraint on the density at the threshold. We will also introduce $\phi_u = P(X > u)$, which is included to natural application. This is consistent with more usual extreme tail models. Let $h(.|\mu,\beta)$ and $g(.|u,\sigma_u,\xi)$ be the normal density function and GPD density function.

The single continuous constraint on the density at the threshold implies:

$$(1-\phi_u)\frac{h(u|\mu,\beta)}{H(u|\mu,\beta)} = \phi_u g(u|u,\sigma_u,\xi).$$

From the single constraint condition, we equate two components at $u$:

$$(1-\phi_u)\frac{h(u|\mu,\beta)}{H(u|\mu,\beta)} = \frac{\phi_u}{\sigma_u}.$$

Rearrange the above equation,

$$\sigma_u = \frac{\phi_u H(u|\mu,\beta)}{(1-\phi_u)h(u|\mu,\beta)}.$$

The single constraint model of hybrid Pareto density function is given by:

$$f(x|\mu,\beta,u,\xi) = \begin{cases} (1-\phi_u)\dfrac{h(x|\mu,\beta)}{H(u|\mu,\beta)} & x \leq u, \\ \\ \phi_u g(x|u,\sigma_u,\xi) & x > u. \end{cases}$$

The $\phi_u$ is estimated using the sample proportion being above the threshold which is an extra parameter in the model.

Otherwise, if $\phi_u = 1 - H(u|\mu,\beta)$, the density function can be simplified to:

$$f(x|\mu,\beta,u,\xi) = \begin{cases} h(x|\mu,\beta) & x \leq u, \\ \\ \phi_u g(x|u,\sigma_u,\xi) & x > u. \end{cases}$$

The $\sigma_u$ is given by:

$$\sigma_u = \frac{1-H(u|\mu,\beta)}{h(u|\mu,\beta)}.$$

In this case, only 4 parameters are needed. The bulk model can be any parametric form distribution, such as, gamma, Weibull or other distributions. The normal distribution is used in this example. It is possible to extend this single continuity constraint model to have different bulk model choices with GPD.

**Discussion About Tail Fraction $\phi_u$**

The normal GPD with single continuity constraint model is different from the hybrid Pareto and hybrid Pareto with single continuity constraint model. With a key difference that the usual tail fraction $\phi_u$ is included.

**Advantages Extension of Hybrid Pareto Model: Single Constraint Model**

1 The beneficial property of this model is the required continuity condition over the two components with using only four or five parameters.

2 Although this model faces the usual robustness issues between the bulk model and tail model, the threshold parameter would offer extra flexibility which may be able to offset some of the robustness issues. The bulk model doesn't affect the tail model as much as the original hybrid Pareto model due to the increased number of free parameters.

3 The inclusion of the tail fraction (compared to the hybrid Pareto) means this model formulation is more consistent with classical extremal tail applications.

**Disadvantages Extension of Hybrid Pareto Model: Single Constraint Model**

1 This single constraint mixture model faces some similar shortcomings with hybrid Pareto model, so they won't be repeated here again.

# Chapter 4

# User's Guide

## 4.1 Introduction

### 4.1.1 What is the `evmix` package for?

`evmix` is an add-on package for the statistical software (R Development Core Team, 2006). The main purpose of this package is to implement extreme value mixture models and other threshold diagnostics to make these new tools accessible for researchers and practitioners in this field. This packages includes the density, distribution, quantile and random number generation function of most of the extreme value mixture distributions defined in the current literature, including extensions of many of them to provide more flexibility. This package provides likelihood inference for these extreme value mixture models, and will later be extended to provide Bayesian inference via MCMC for some of the models which have likelihoods which are challenging to numerically maximise. A reasonably comprehensive review of extreme value mixture models can be found in Scarrott and MacDonald (2012).

**Extreme Value Mixture Distributions**

The following is a list of all the mixture models which are currently implemented and a reference to where I believe they were first proposed or applied. Very few extreme value mixture models in the literature are constrained to be continuous at the threshold. So to provided greater flexibility we have provided both the original version of each mixture model and where appropriate, an extended version which is constrained to continuous at the threshold is also provided (with the function names post-fixed with `con`).

The implemented extreme value mixture models with a **parametric** model for the bulk component are:

- Normal GPD mixture distribution (Behrens et al., 2004)
- Normal GPD with single continuity constraint mixture distribution
- Gamma GPD mixture distribution (Behrens et al., 2004)
- Gamma GPD with single continuity constraint mixture distribution
- Weibull GPD mixture distribution (Behrens et al., 2004)

- Weibull GPD with single continuity constraint mixture distribution
- Lognormal GPD mixture distribution (Behrens et al., 2004)
- Lognormal GPD with single continuity constraint mixture distribution
- Beta GPD mixture distribution (Behrens et al., 2004)
- Beta GPD with single continuity constraint mixture distribution
- GPD-Normal-GPD(GNG) mixture distribution (Zhao et al., 2010)
- GPD-normal-GPD (GNG) with single continuity constraint mixture distribution
- Dynamically weighted mixture distribution (Frigessi et al., 2002)
- Hybrid Pareto distribution (Carreau and Bengio, 2008)
- Hybrid Pareto with single continuity constraint distribution

The implemented extreme value mixture models with a **semi-parametric** model for the bulk component are:

- Mixture of gamma GPD distribution (Nascimento et al., 2011)

The implemented extreme value mixture models with a **non-parametric** model for the bulk component are:

- Kernel GPD distribution (MacDonald et al., 2011)
- Kernel GPD with single continuity constraint distribution
- Two tailed kernel GPD distribution (MacDonald et al., 2011)
- Boundary correction kernel GPD distribution (MacDonald et al., 2013)

### 4.1.2 Obtaining the package/guide

The package can be downloaded from CRAN (The Comprehensive R Archive Network) at http://cran.r-project.org/. The PDF version of this guide will be in package installation directory and available from the help system, type `help(evmix)` and click `browse directory`.

### 4.1.3 Citing the package/guide

The package citation is available from `citation(evmix)`. The current citation is:

Hu, Y. and Scarrott, C.J. (2013). evmix: `R` Package for Extreme Value Mixture Modelling and Threshold Estimation. Available on CRAN. URL: http://www.math.canterbury.ac.nz/ c.scarrott/evmix

### 4.1.4 Caveat

These functions are provided on a best attempt basis, use at your own risk. They are well tried and tested and have many error checking facilities to ensure users can't do silly things and useful error messages are reported back where needed. However, there will almost certainly be bugs. If you do find a possible bug in the code or documentation then do report it to Carl Scarrott, `carl.scarrott@canterbury.ac.nz`. For bugs, please provide a reproducible example.

### 4.1.5 Legalese

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version. This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details. You should have received a copy of the GNU General Public License along with this program. If not, see `http://www.gnu.org/licenses`.

## 4.2 Generalized Pareto distribution(GPD)

Coles (2001) states that if $X_1, \ldots, X_n$ be a sequence of independent random variables, for some large enough threshold $u$, the excess $X - u$ may be well approximated by a generalized Pareto distribution.

The distribution function of GPD is given by:

$$G(x|u, \sigma_u, \xi) = \begin{cases} 1 - \left[1 + \xi \left(\frac{x-u}{\sigma_u}\right)\right]_+^{-1/\xi} & \xi \neq 0, \\ 1 - \exp\left[-\left(\frac{x-u}{\sigma_u}\right)\right]_+ & \xi = 0. \end{cases}$$

The shape parameter $\xi$ determines the tail behavior of the GPD:

- $\xi < 0$: short tail with finite upper end point of $u - \frac{\sigma_u}{\xi}$;
- $\xi = 0$: exponential tail; and
- $\xi > 0$: heavy tail.

It is common to use letters `d`, `p`, `q`, `r` with a distribution name to represent the corresponding density, distribution, quantile and random number generation functions respectively. This package follows this guideline. The parameter definitions and ordering are explicitly designed to be similar to those used in the `evd` package, such that most code would be interchangeable between them.

```
# User can specify whether density or log density should be evaluated by setting the log=FALSE or log=TRUE.
dgpd(0.1, u = 0, sigmau = 1, xi = c(0.05, 0.1, 0.15), log = TRUE)
[1] -0.1047384 -0.1094536 -0.1141460

# User can specify whether cumulative or exceedance probability by setting option lower.tail=FALSE or TRUE.
pgpd(0.1,u = 0, sigmau = c(0.99, 1, 1.01), xi = 0.1, lower.tail = FALSE)
[1] 0.9043821 0.9052870 0.9061749

rgpd(5,u = 0, sigmau = 1, xi = 0)
[1] 2.1070377 0.1620293 0.3380177 2.6219259 0.2536274
```

## 4.3 Parametric Mixture Models

This section will cover several existing parametric form models in the literature and some extensions.

### 4.3.1 Gamma/Normal/Weibull/Log-Normal/Beta GPD Model

Behrens et al. (2004) developed a simple extreme value mixture model by combining a parametric bulk model below the threshold with GPD above the threshold. This has been implemented using gamma, Weibull, normal distribution as the bulk model. There is no standardized framework for defining the renormalizing constant of combining the two distributions in the literature. This package implements two approaches for specifying the tail fraction $\phi_u$ either defined using the parameters from the bulk component or as a separate parameter to be estimated, to be detailed below.

The distribution function of the parametric form of the of the bulk component of the extreme value mixture mixture model can be defined as:

**Bulk Model Based Tail Fraction Approach**

$$F(x|u,\theta,\sigma_u,\xi,\phi_u) = \begin{cases} H(x|\theta) & x \leq u, \\ H(u|\theta) + (1 - H(u|\theta)) \times G(x|u,\sigma_u,\xi) & x > u. \end{cases} \tag{4.1}$$

**Parameterised Tail Fraction Approach**

$$F(x|u,\theta,\sigma_u,\xi,\phi_u) = \begin{cases} (1-\phi_u) \times \frac{H(x|\theta)}{H(u|\theta)} & x \leq u, \\ (1-\phi_u) + \phi_u \times G(x|u,\sigma_u,\xi) & x > u, \end{cases} \tag{4.2}$$

where $\phi_u$ is the proportion of data above the threshold and $0 < \phi_u < 1$. $H(.|\theta)$ could be parametric distribution function such as gamma, Weibull or normal distribution function and $\theta$ is the parameter vector of the bulk distribution. $G(.|u,\sigma_u,\xi)$ represents the GPD distribution function where $u$ is the threshold, $\xi$ is the shape parameter and $\sigma_u$ is the scale parameter.

The latter approach in Equation (4.2) explicitly makes it clear that when conditional modeling the upper tail using the GPD also requires the proportion of excesses to obtain the unconditional quantities of interest. The maximum likelihood estimator of this tail fraction parameter is of

course the usual sample proportion. The former approach in Equation (4.1) is included in Equation (4.2) as a special case when $\phi_u$ is set to $1 - H(u|\theta)$.

### Base Function

The density function of the mixture model:

- `dgammagpd(x, gshape = 1, gscale = 1, u = qgamma(0.9, gshape, 1/gscale), sigmau = sqrt(gshape) * gscale, xi = 0, phiu = TRUE, log = FALSE)` ,

  where `gshape` and `grate` are represented the gamma shape and rate parameter.

- `dweibullgpd(x, wshape = 1, wscale = 1, u = qweibull(0.9, wshape, wscale), sigmau = sqrt(wscale^2 * gamma(1 + 2/wshape) - (wscale * gamma(1 + 1/wshape))^2), xi = 0, phiu = TRUE, log = FALSE)`,

  where `wshape` and `wscale` are represented the Weibull shape and scale parameter.

- `dnormgpd(x, nmean = 0, nsd = 1, u = qnorm(0.9, nmean, nsd), sigmau = nsd, xi = 0, phiu = TRUE, log = FALSE))`,

  where `nmean` and `nsd` are represented the normal mean and standard deviation parameter. The `xi` and `sigmau` are the usual GPD shape and scale parameter. The use of the name `sigmau` reminds us of the dependence between the threshold and GPD scale parameter. If `phiu = TRUE`, the bulk model based approach is adopted. If `phiu` is between 0 and 1 (exclusive), the parameterised approach is adopted.

```
dnormgpd(1.1, c(0.02, 0), 1, 1.28, 0.5173, -0.1489, phiu = TRUE)
[1] 0.2226535 0.2178522

pnormgpd(1.4, 0, 1,c(1.1,1.28), 0.5173, c(-0.1,-0.1489), phiu=0.1)
[1] 0.9449776 0.9210278

par(mfrow=c(2,2))
x = seq(0,10,0.01)
f = dgammagpd(x, 2.1, 2, 7.7, 1.52, -0.07, phiu = 0.1)
plot(x, f, type = 'l', lty = 1, lwd = 2, ylab = "f(x)", main = "Gamma GPD density function")
abline(v = 7.7, lty = 3, lwd = 2)

f = dgammagpd(x, 9, 0.5, 6.5, 0.62, -0.21, phiu = 0.1)
plot(x, f, type = 'l', lty = 1, lwd = 2, ylab = "f(x)", main = "Gamma GPD density function")
abline(v = 6.5, lty = 3, lwd = 2)

f = pgammagpd(x, 2.1, 2, 7.7, 1.52, -0.07, phiu = 0.1)
plot(x, f, type = 'l', lty = 1, lwd = 2, ylab = "F(x)", main = "Gamma GPD distribution function")
abline(v = 7.7, lty = 3, lwd = 2)

f = pgammagpd(x, 9, 0.5, 6.5, 0.62, -0.21, phiu = 0.1)
plot(x, f, type = 'l', lty = 1, lwd = 2, ylab = "F(x)", main = "Gamma GPD distribution function")
abline(v = 6.5, lty = 3, lwd = 2)
```

### Likelihood Function

The negative log likelihood function is pre-fixed by `nl` with a distribution name (e.g. `nlweibullgpd` and and the corresponding log likelihood functions prefixed by `l` (e.g.`lweibullgpd`). It is worthwhile to notice that these likelihood functions are implemented allowing the two approaches for the parameter $\phi_u$, where phiu=FALSE gives uses the sample proportion estimate and phiu=TRUE gives the bulk model estimate.

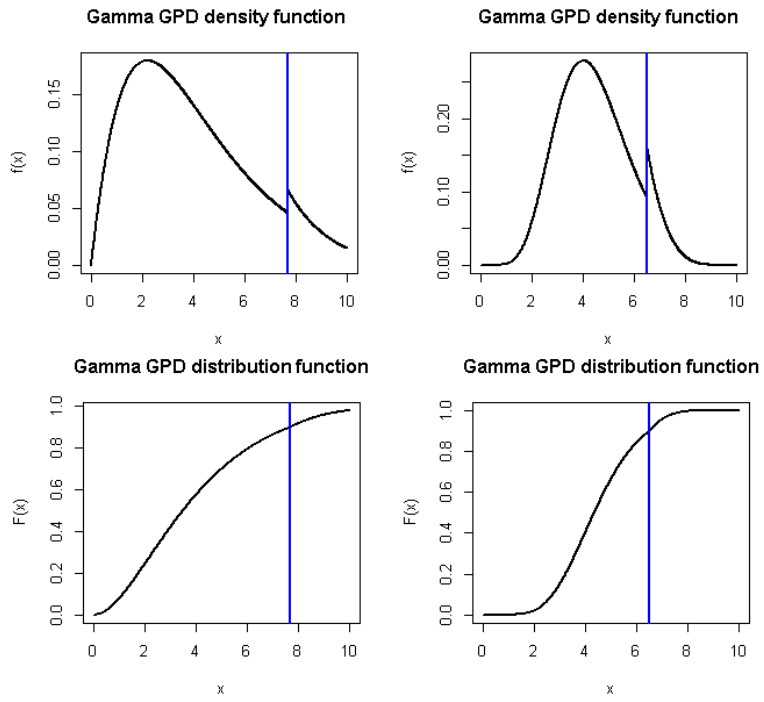



Figure 4.1: The density function of the parametric form GPD mixture model showing that it may have a discontinuity at the threshold. However, the corresponding distribution function (or CDF) is always continuous at the threshold. The vertical dashed line indicates the threshold.

An example of the syntax for the negative log likelihood function is explicitly designed for use in the optimisation, so all the parameters are input as a vector (except $\phi_u$) and is just a wrapper for the likelihood function itself::

- `nlgammagpd(pvector, x, phiu = TRUE, finitelik = FALSE)`, `pvector` has 5 arguments with gamma shape, gamma scale, threshold, GPD scale and GPD shape respectively.

The log likelihood function is pre-fixed by `l` with a distribution name (e.g. `lweibullgpd` and `lnormgpd`). The log likelihood function syntax is similar to that of the usual distribution functions::

- `lgammagpd(x, gshape = 1, gscale = 1, u = qgamma(0.9, gshape, 1/gscale), sigmau = sqrt(gshape) * gscale, xi = 0, phiu = TRUE, log = TRUE)`

Now `x` is the data and `phiu` is the logical input indicating either parameterised approach or bulk model based approach. If let `phiu = TRUE`, the likelihood function will apply the bulk model based approach. If `phiu = FALSE`, the likelihood function will adopt the parameterised approach, and estimate `phiu` using the sample proportion above the threshold.

### Fitting Function

The maximum likelihood based fitting function is pre-fixed by `f` with a distribution name (e.g. `fweibullgpd`). An example of the syntax for the MLE fitting function of the mixture model:

- `fgammagpd(x, phiu = TRUE, pvector = NULL, std.err = TRUE, method = "BFGS", control = list(maxit = 10000), finitelik = TRUE, ...)`,

The `pvector` is a vector of initial values of the mixture model and when `pvector = NULL` then the default initial values are used. `x` is a vector of sample data. The default value for `phiu = TRUE` so that the tail fraction is specified by gamma distribution $\phi_u = 1 - H(.|\theta)$, where $H(.|\theta)$ could be parametric distribution function. When `phiu = FALSE` then the tail fraction is treated as an extra parameter estimated using the MLE which is the sample proportion above the threshold. `std.err` is a logical variable, if `std.err = TRUE`, then the standard errors of the model parameters are calculated. `finitelik` is a logical variable, if `finitelik = TRUE`, then log-likelihood returns finite value for invalid parameters. The default optimisation algorithm is "BFGS", which requires a finite negative log-likelihood function evaluation `finitelik = TRUE`. For invalid parameters, a zero likelihood is replaced with `exp(-1e6)`. The "BFGS" optimisation algorithms require finite values for likelihood, so any user input for `finitelik` will be overridden and set to `finitelik = TRUE` if either of these optimisation methods is chosen. The `method`, `control` and `...` are optional inputs passed to `optim`.

**Examples**

An example program for fitting a parametric bulk model is now given in Figures 4.2. The fit around the threshold are rather different.

```
x = rgamma(1000, shape = 2)
xx = seq(-1, 10, 0.01)
y = dgamma(xx, shape = 2)
par(xaxs="i", yaxs="i")

# Bulk model base tail fraction
fit = fgammagpd(x, phiu = TRUE, std.err = FALSE)
hist(x, breaks = 100, freq = FALSE, xlim = c(0, 10))
lines(xx, y)
lines(xx, dgammagpd(xx, gshape = fit$gshape, gscale = fit$gscale, u = fit$u,
   sigmau = fit$sigmau, xi = fit$xi, phiu = TRUE), col="red")
abline(v = fit$u)

# Parameterised tail fraction
fit2 = fgammagpd(x, phiu = FALSE, std.err = FALSE)
plot(xx, y, type = "l")
lines(xx, dgammagpd(xx, gshape = fit$gshape, gscale = fit$gscale, u = fit$u,
  sigmau = fit$sigmau, xi = fit$xi, phiu = TRUE), col="red")
lines(xx, dgammagpd(xx, gshape = fit2$gshape, gscale = fit2$gscale, u = fit2$u,
   sigmau = fit2$sigmau, xi = fit2$xi, phiu = fit2$phiu), col="blue")
abline(v = fit$u, col = "red")
abline(v = fit2$u, col = "blue")
legend("topright", c("True Density","Bulk T. F","Parameterised T. F"),
   col=c("black", "red", "blue"), lty = 1, cex= 0.8)
```

### 4.3.2 Gamma/Normal/Weibull/Log-Normal/Beta GPD with Single Continuity Constraint Mixture Model

The continuity constraint is achieved by equating the bulk distribution and tail model at the threshold and replacing the GPD scale parameter in terms of other parameters. For the mixture extreme value models, the postfixes `con` means a continuity constraint at the threshold. For

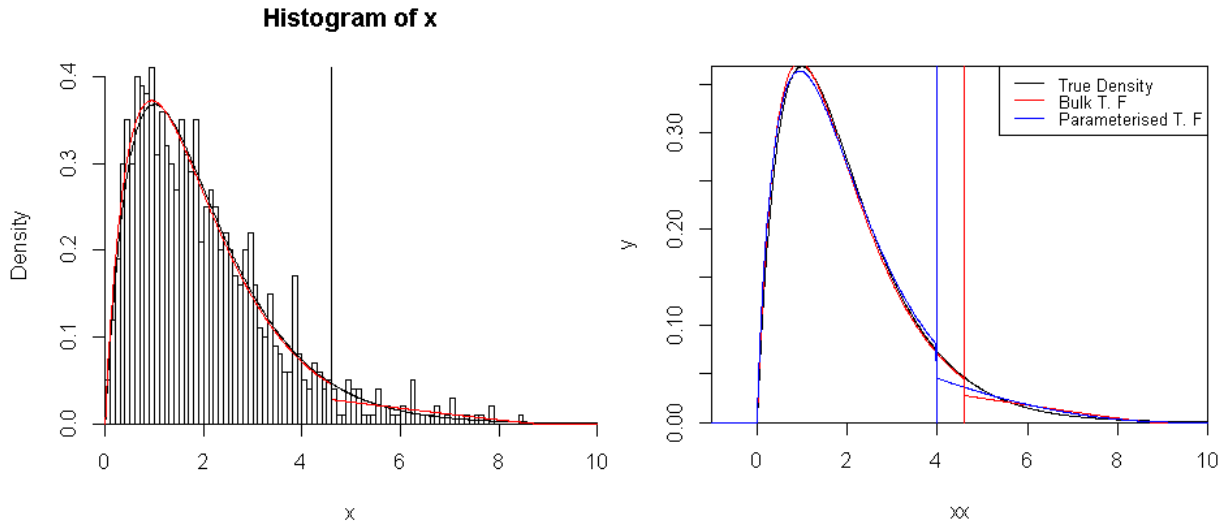



Figure 4.2: Density function of the parametric extreme value mixture model with gamma for bulk with bulk and parameterised tail fraction approach, where the vertical line indicates the threshold. A sample of size 1000 from a gamma distribution(2,1) is generated.

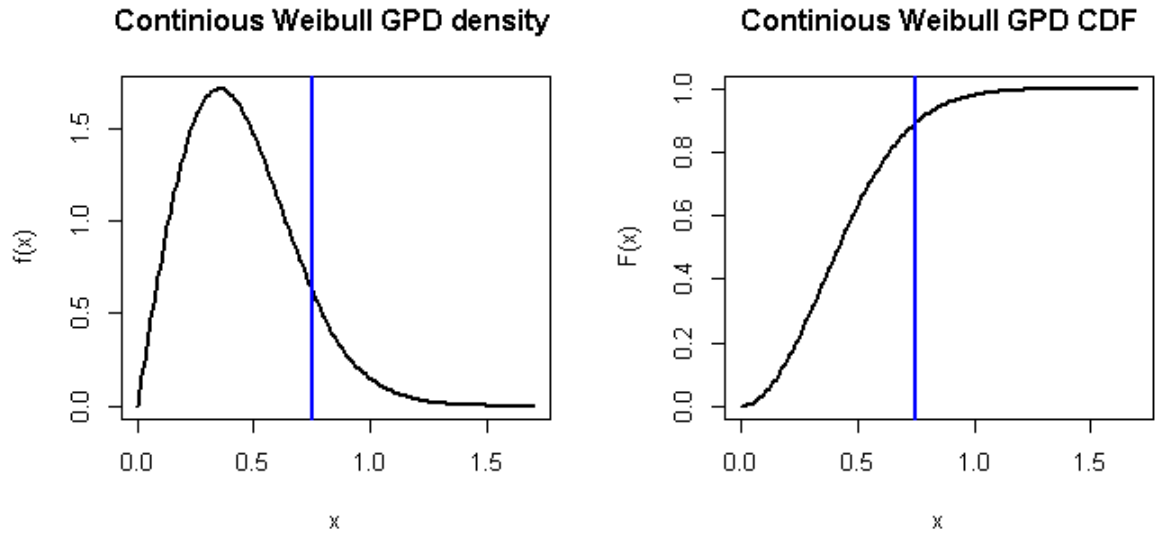



Figure 4.3: Example of density function and distribution function for continuous Weibull GPD distribution, where the Weibull shape parameter is 2, the scale parameter is 0.5, the threshold is fixed at 0.75 and GPD shape is -0.13. The blue line indicates the threshold.

example, the Weibull distribution is defined as `weibull` in `R`, so the quantile function for Weibull GPD model with continuity constraint is therefore abbreviated as `qweibullgpdcon`.

In the Figure 4.3, the Weibull shape parameter is 2, the scale parameter is 0.5, the threshold is fixed at 0.75 and GPD shape is -0.13. It can be seen that not only the density function for this mixture model is continuous at the threshold and distribution function will be continuous in first derivative.

**Base Function**

The base functions of the parametric form continuous mixture model have very similar set up with the functions of parametric form mixture model, so will not be demonstrated again.

Examples of the syntax of the density function of parametric form continuous mixture model:

- `dgammagpdcon(x, gshape = 1, gscale = 1, u = qgamma(0.9, gshape, 1/gscale), xi = 0, phiu = TRUE, log = FALSE)`, As usual, `phiu` can represent either bulk model based approach or parameterised approach.

**Fitting and Likelihood Function**

The fitting and likelihood functions of the continuity constraint mixture model has very similar set up with the fitting function of parametric form mixture model. Hence, the basic use of these functions will not be demonstrated again.

### 4.3.3 GPD-Normal-GPD (GNG) Model

Zhao et al. (2010) proposed a two tail mixture model, where the normal distribution describes the bulk and the GPD has been fitted to both tails separately. The parameter vector is $\Theta = (\mu, \beta, u_l, \sigma_{u_l}, \xi_l, u_r, \sigma_{u_r}, \xi_r)$.

**Bulk Model Based Tail Fraction Approach**

The distribution function is defined as:

$$F(x|\Theta) = \begin{cases} \phi_{u_l}\left\{1 - G(-x| - u_l, \sigma_{u_l}, \xi_l)\right\} & x < u_l, \\ H(x|\mu, \beta) & u_l \leq x \leq u_r, \\ (1 - \phi_{u_r}) + \phi_{u_r} G(x|u_r, \sigma_{u_r}, \xi_r) & x > u_r, \end{cases}$$

where $\phi_{u_l} = H(u_l|\mu, \beta)$ and $\phi_{u_r} = 1 - H(u_r|\mu, \beta)$ and $H(.|\mu, \beta)$ is the normal distribution function with mean $\mu$ and standard deviation $\beta$. $G(.| - u_l, \sigma_{u_l}, \xi_l)$ and $G(.|u_r, \sigma_{u_r}, \xi_r, )$ are GPD distribution functions for lower tail and upper tail.

**Parameterised Tail Fraction Approach**

It is possible to define $\phi_{u_l} = P(X < u_l)$ and $\phi_{u_r} = P(X > u_r)$. The corresponding distribution function is given by:

$$F(x|\Theta) = \begin{cases} \phi_{u_l}\left\{1 - G(-x| - u_l, \sigma_{u_l}, \xi_l)\right\} & x < u_l, \\ \phi_{u_l} + (1 - \phi_{u_l} - \phi_{u_r})\dfrac{H(x|\mu, \beta) - H(u_l|\mu, \beta)}{H(u_r|\mu, \beta) - H(u_l|\mu, \beta)} & u_l \leq x \leq u_r, \\ (1 - \phi_{u_r}) + \phi_{u_r} G(x|u_r, \sigma_{u_r}, \xi_r) & x > u_r, \end{cases}$$

**Base function**

The example of the density function of GNG model.

- `dgng(x, nmean = 0, nsd = 1,`
  `ul = qnorm(0.1, nmean, nsd), sigmaul = nsd, xil = 0, phiul = TRUE,`
  `ur = qnorm(0.9, nmean, nsd), sigmaur = nsd, xir = 0, phiur = TRUE, log = FALSE),`

The `xil` and `sigmaul` are GPD shape and scale parameter for the lower tail and the corresponding value `xir` and `sigmaur` for the upper tail.

```
f = dgng(x, 0, 1, -1.28, 0.42, 0.2, phiul = TRUE, 1.28, 0.42, 0.2, phiur = TRUE)
plot(x, f, ylim = c(0, 0.5), type = 'l', lty = 1, ylab = "Density", main = "GNG density function")
abline(v = -1.28, col = "blue")
abline(v =1.28, col = "blue")

f1 = dgng(x, 0, 1, -1.28, 0.6, -0.11, phiul = TRUE, 1.28, 0.6, -0.12, phiur = TRUE)
plot(x, f1, ylim = c(0, 0.5), type = 'l',lty = 1, ylab = "Density", main = "GNG density function")
abline(v = -1.28, col = "blue")
abline(v = 1.28, col = "blue")
```

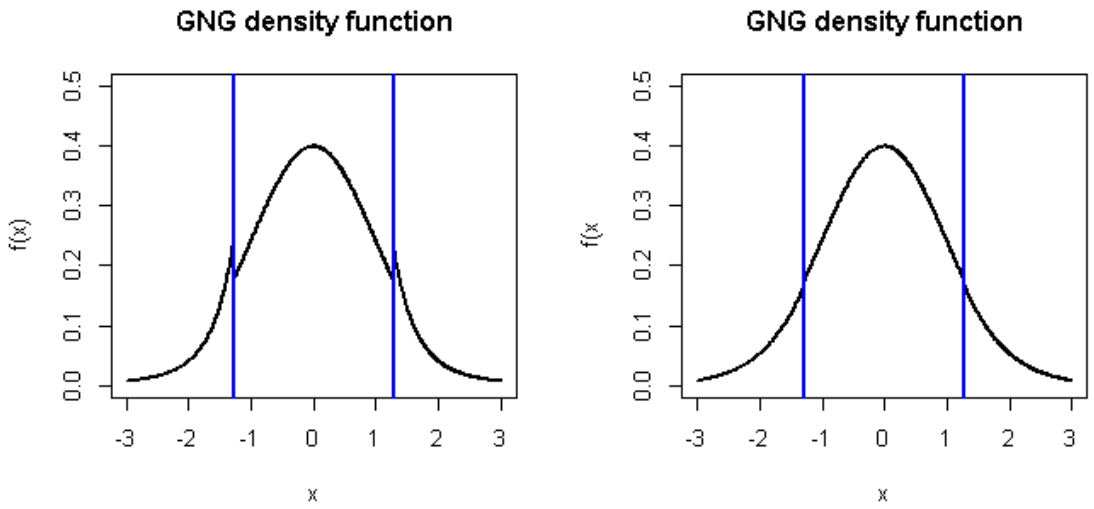



Figure 4.4: Example of the density function of GNG model. The parameter vector is $\Theta = (\mu = 0, \beta = 1, u_l = -1.28, \sigma_{u_l} = 0.42, \xi_l = 0.2, u_r = 1.28, \sigma_{u_r} = 0.42, \xi_r = 0.2)$ on left and $\Theta = (\mu = 0, \beta = 1, u_l = -1.28, \sigma_{u_l} = 0.6, \xi_l = -0.11, u_r = 1.28, \sigma_{u_r} = 0.6, \xi_r = 0.12)$ on right. This example shows the density function of GNG could have continuity / discontinuity constraints at either lower or upper threshold with different choices of parameter values.

An example of the density function of GNG model is given in Figure 4.4. The density function may have a discontinuity at either the lower threshold $u_l$ or the upper threshold $u_r$.

### Likelihood and Fitting Function

The fitting and likelihood functions are similar to those of the single tailed models, with the only difference that they now take parameters for both tails and the tail fraction can take either parameterisation for both tails.

### Examples

Figure 4.5 provides an example of simulated data with overlaid fitted `gng` and `normgpd` model. It shows that two tailed model provides more flexibility in fitting asymmetry data.

```
x = rnorm(1000)
xx = seq(-5, 5, 0.01)
y = dnorm(xx)

# Bulk model base tail fraction
fit = fgng(x, phiul = TRUE, phiur = TRUE, std.err = FALSE)
hist(x, breaks = 100, freq = FALSE, xlim = c(-5, 5), main = "N(0, 1)")
lines(xx, dgng(xx, nmean = fit$nmean, nsd = fit$nsd,
   ul = fit$ul, sigmaul = fit$sigmaul, xil = fit$xil, phiul = TRUE,
   ur = fit$ur, sigmaur = fit$sigmaur, xir = fit$xir, phiur = TRUE), col="blue", lwd = 2)
abline(v = c(fit$ul, fit$ur))
legend('topright', 'gng',
       col = "blue", lty =1, lwd = 2)
# Two tail model is safest if bulk has lower tail which is not normal tail
x = rt(3000, df = 3)
xx = seq(-10, 10, 0.01)
y = dt(xx, df = 3)
hist(x, breaks = 100, freq = FALSE, xlim = c(-10, 10), ylim= c(0, 0.5), main = "T (df=3)")

fit = fnormgpd(x, phiu = FALSE, std.err = FALSE)
fit2 = fgng(x, phiul = FALSE, phiur = FALSE, std.err = FALSE)
lines(xx, dnormgpd(xx, nmean = fit$nmean, nsd = fit$nsd,
   u = fit$u, sigmau = fit$sigmau, xi = fit$xi, phiu = fit$phiu), col="red", lwd = 2)
lines(xx, dgng(xx, nmean = fit2$nmean, nsd = fit2$nsd,
              ul = fit2$ul, sigmaul = fit2$sigmaul, xil = fit2$xil, phiul = fit2$phiul,
              ur = fit2$ur, sigmaur = fit2$sigmaur, xir = fit2$xir, phiur = fit2$phiur), col="blue", lwd = 2)
legend('topright', c("gng", "normgpd"), col = c("blue", "red"), lty =1, lwd = 2)
```

### 4.3.4 GNG with Single Continuity Constraint Model

The continuity constraints are achieved by equating the normal distribution with the GPD at the lower and upper threshold. The GPD scale parameters of the both tails are replaced by the other parameters. The function names are post-fixed by `dgngcon`, `pgngcon`, `qgngcon`, `rgngcon` to represent corresponding density, distribution, quantile and random number generation function respectively.

Those functions from the GNG continuous mixture model have very similar set up with the functions of GNG model. Hence, the details of these functions will not repeat again.

### 4.3.5 Dynamically Weighted Mixture Model

Frigessi et al. (2002) suggested a dynamically weighted mixture model by combining the Weibull for the bulk model with GPD for the tail model. Instead of explicitly defining the threshold, they use a Cauchy cumulative distribution function for the mixing function to make the transition between the bulk distribution and tail distribution.
The density function of the dynamic mixture model is given by:

$$l(x) = \frac{[1 - p(x|\theta)]\, f(x|\beta) + p(x|\theta) g(x|0, \sigma_u, \xi)}{Z(\theta, \beta, \sigma_u, \xi)},$$

where $f(x|\beta)$ denotes the Weibull density and $Z(\theta, \beta, \sigma_u, \xi)$ denotes the normalizing constant to make the density integrate to one. The mixing function $p(.|\theta)$ is a Cauchy distribution function with location parameter $\mu$ and scale parameter $\tau$.

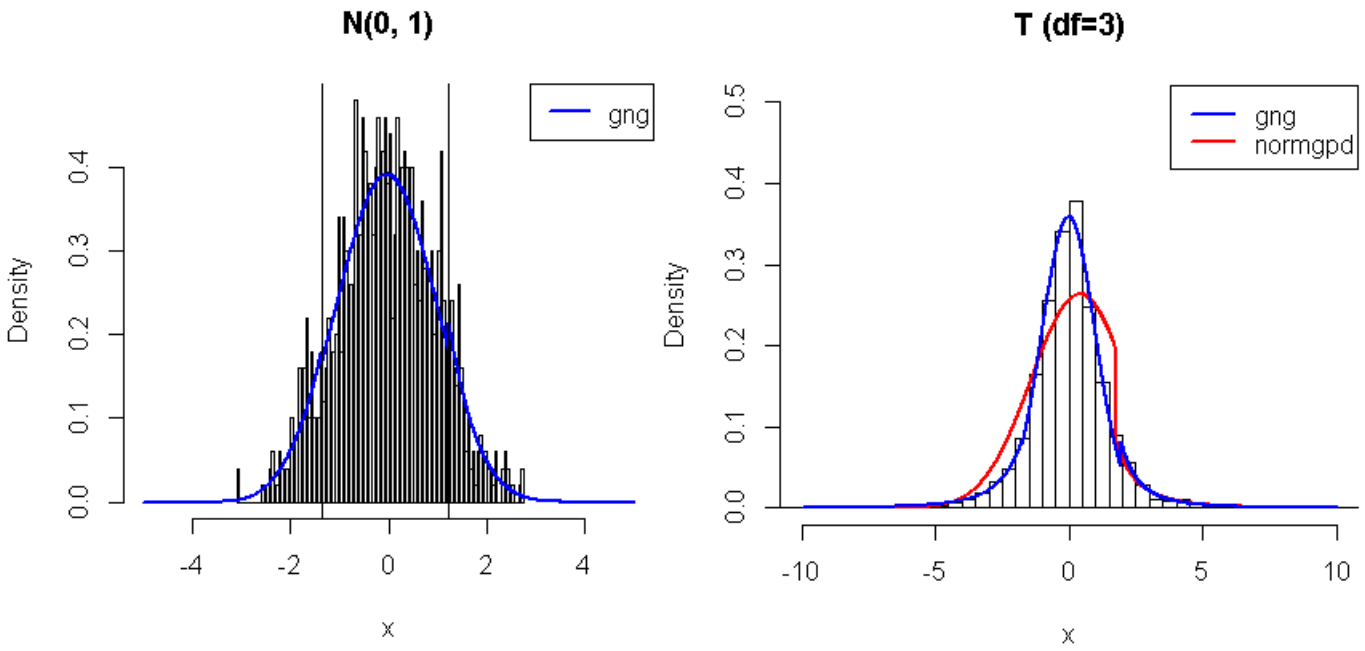



Figure 4.5: Left: The example of 1000 simulated data from a standard normal distribution with overlaid fitted GNG model with parameter vector ($\mu = -0.049, \beta = 0.971, u_l = -1.347, \sigma_{u_l} = 0.519, \xi_l = -0.170, u_r = 1.191, \sigma_{u_r} = 0.637, \xi_r = -0.223$). Right: The example of 3000 simulated data from Student-t(3) distribution with overlaid fitted normgpd and GNG model with parameter vector ($\mu = 0.455, \beta = 1.792, u = 1.729, \sigma_u = 0.940, \xi = 0.394$) and ($\mu = -0.005, \beta = 0.994, u_l = -1.778, \sigma_{u_l} = 0.888, \xi_l = 0.318, u_r = 1.687, \sigma_{u_r} = 0.913, \xi_r = 0.396$), respectively.

### Base Function

This dynamic weighted mixture distribution is abbreviated as dwm and an example of the syntax is given by:

- `ddwm(x, wshape = 1, wscale = 1, cmu = 1, ctau = 1,`
  `sigmau = sqrt(wscale^2 * gamma(1 + 2/wshape) - (wscale * gamma(1 + 1/wshape))^2),`
  `xi = 0, log = FALSE),`

As before, the `wshape` and `wscale` are the Weibull shape and scale parameters. `cmu` and `ctau` are Cauchy distribution location and scale parameter. The quantile function `qdwm` requires the numerical integration of the distribution function `pdwm`, which very rarely fails due to integration difficulties. Further, the quantile function is not available in close form so require numerical inversion, which can also fail. The argument `qinit` is the initial quantile estimate, which can be used in case of failure. Hence, the parameter `qinit` is set up for user to choose some reasonable initial values for the quantile function.

An example of the three components and the resultant density estimate are given in Figure 4.6, which reproduces the example given in Frigessi et al. (2002).

```
xx = seq(0.001, 5, 0.01)
f = ddwm(xx, wshape = 2, wscale = 1/gamma(1.5), cmu = 1, ctau = 1, sigmau = 1, xi = 0.5)
plot(xx, f, ylim = c(0, 1), xlim = c(0, 5), type = 'l', lwd = 2,
```

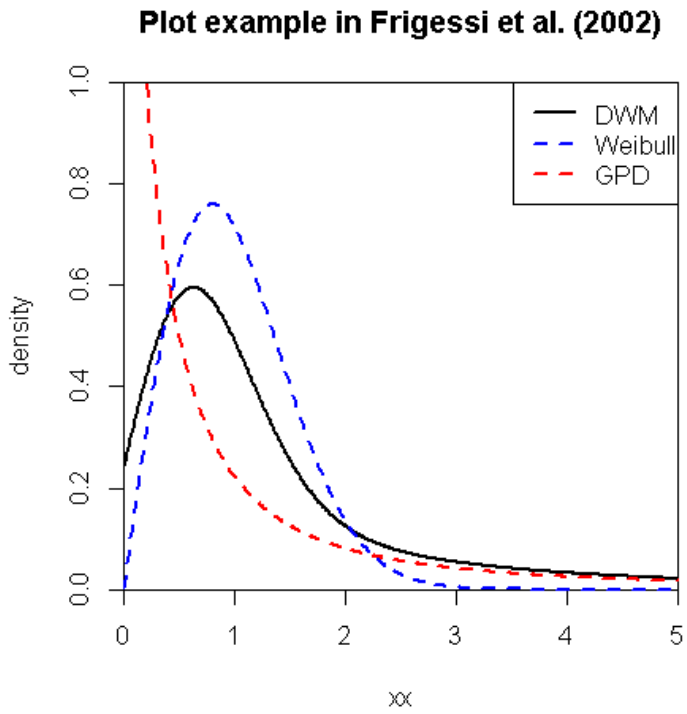



Figure 4.6: Reproduce the plot examples of the dynamically weighted mixture model from Frigessi et al. (2002) in page 226, where the parameter vector $\Theta = (\texttt{wshape} = 2, \texttt{wscale} = 1.128, \texttt{cmu} = 1, \texttt{ctau} = 1, \texttt{sigmau} = 1, \texttt{xi} = 0.5)$.

```
    ylab = "density", main = "Plot example in Frigessi et al. (2002)")
lines(xx, dgpd(xx, xi = 1, sigmau = 0.5), col = "red", lty = 2, lwd = 2)
lines(xx, dweibull(xx, shape = 2, scale = 1/gamma(1.5)), col = "blue", lty = 2, lwd = 2)
legend('topright', c('DWM', 'Weibull', 'GPD'),
       col = c("black", "blue", "red"), lty = c(1, 2, 2), lwd = 2)
```

### Fitting and Likelihood Function

The fitting and likelihood functions are similar with other models so is not discussed here.

### Examples

Figure 4.7 provides an example of simulated Weibull data with overlaid fitted dynamic weighted mixture model.

```
x = rweibull(1000, shape = 2, scale = 1.5)
xx = seq(0.1, 5, 0.01)
fit = fdwm(x, std.err = FALSE)
hist(x, 100, freq = FALSE, ylim = c(0, 1.3), main =" dwm example")

lines(xx, ddwm(xx, wshape = fit$wshape, wscale = fit$wscale, cmu = fit$cmu, ctau = fit$ctau,
              sigmau = fit$sigmau, xi = fit$xi),lwd=2)
lines(xx, dgpd(xx, sigmau = fit$sigmau, xi = fit$xi), col = "red", type = 'l', lwd = 2)
lines(xx, dweibull(xx, shape = fit$wshape, scale = fit$wscale), col = "blue", lwd = 2)
lines(xx, pcauchy(xx, location = fit$cmu, scale = fit$ctau), col = "green", lty = 2, lwd = 2)
legend("topleft", c("dwm", "GPD", "Weibull", "Cauchy"), col = c("black", "red", "blue", "green")
```

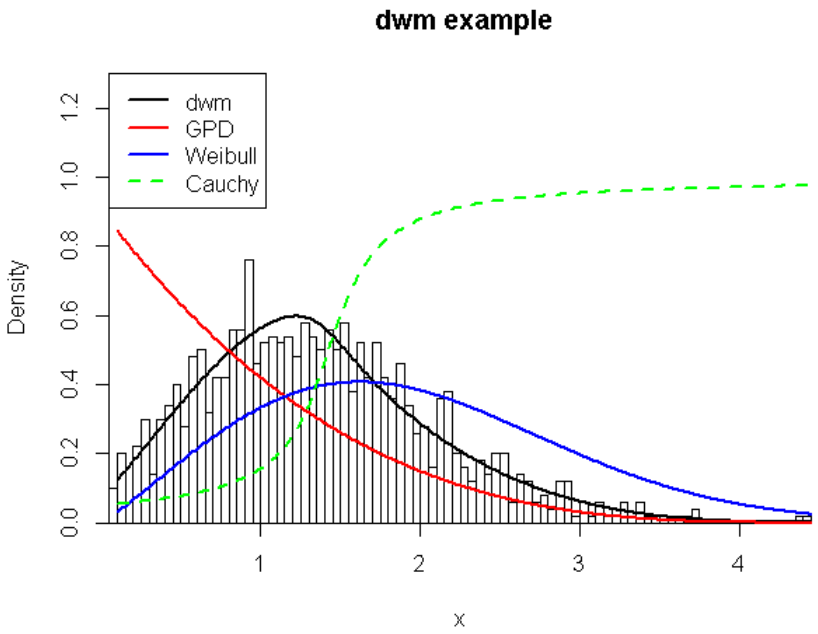



Figure 4.7: Example of 1000 simulated data from a Weibull(2,1.5) distribution with overlaid fitted dynamic weighted mixture model with parameter vector $\Theta = (\texttt{wshape} = 1.993, \texttt{wscale} = 1.530, \texttt{cmu} = 2.901, \texttt{ctau} = 0.001, \texttt{sigmau} = 1.027, \texttt{xi} = -0.145)$.

### 4.3.6 Hybrid Pareto Model

Carreau and Bengio (2008) proposed a hybrid Pareto model by splicing a normal distribution with GPD and set continuity constraint on the density and on its first derivative at the threshold. However, with a key difference to the usual mixture from Section 4.3.1 to Section 4.3.5 that it does not treat the GPD a conditional model, so the tail fraction scaling of the GPD $\phi_u = P(x > u)$ is not included. The unscaled GPD is simply spliced with the normal truncated at the threshold, so the density needs to be renormalized as it gets the contribution from the truncated normal upto the threshold and the entire unit integration of the GPD above the threshold.. The parameters have to adjust for the lack of tail fraction scaling. The GPD scale $\sigma_u$ and threshold $u$ are parameterised the other parameters to achieve two continuous constraints. Hence, they have reduced the initial 5 parameters to the normal mean $\mu$ and standard deviation $\beta$ with the GPD shape $\xi$. The parameter vector is $\theta = (\mu, \beta, \xi)$. The hybrid Pareto mixture density is defined as:

$$f(x|\theta) = \begin{cases} \dfrac{1}{\tau} h(x|\mu, \beta) & x \leq u, \\ \\ \dfrac{1}{\tau} g(x|u, \sigma_u, \xi) & x > u. \end{cases}$$

$\mu$ is the normal mean and $\beta$ is the standard deviation parameter. $h(.|\mu, \beta)$ is the normal density function and $g(.|u, \sigma_u, \xi)$ is the GPD density function. The $\tau$ is the usual normalising constant and $\tau = 1 + H(u|\mu, \beta)$, where the 1 comes from the integration of the unscaled GPD and second term is from the usual normal component.

The `condmixt` package written by one of the original authors of the hybrid Pareto model (Carreau

and Bengio (2008)) also has similar functions for the hybrid Pareto `hpareto` and mixture of hybrid Paretos `hparetomixt`, which are more flexible as they also permit the model to be truncated at zero.

### Base Function

The hybrid Pareto distribution is abbreviated as hpd. Hence, the corresponding functions are defined as:

- `dhpd(x, nmean = 0, nsd = 1, xi = 0, log = FALSE)`,

### Likelihood and Fitting Function

### Examples

Figure 4.8 demonstrates that the hybrid Pareto cannot provide a reasonable fit for the normal distribution but `normgpdcon` model could give a better fit.

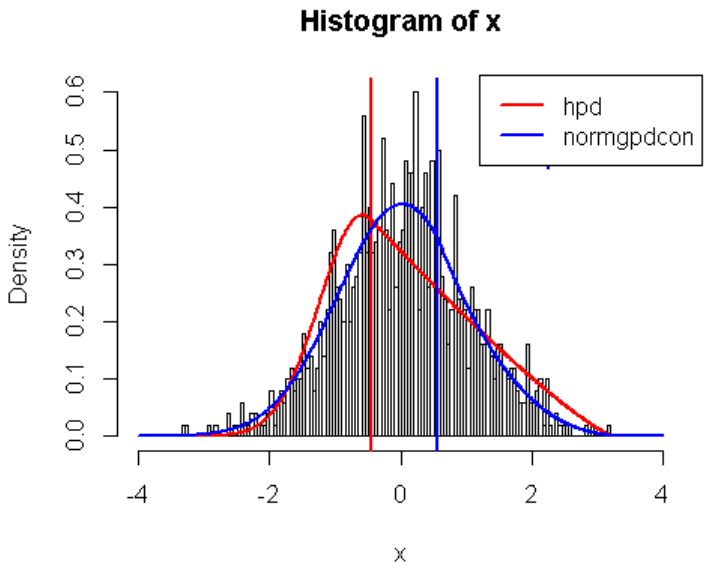



Figure 4.8: Example of 1000 simulated data from a standard normal distribution with overlaid fitted `hpd` model with parameter vector of $(\mu = -0.606, \beta = 0.614, \xi = -0.372)$ and `normgpdcon` model with parameter vector of $(\mu = -0.015, \beta = 1.001, u = 1.987, \xi = -0.089)$.

The negative log likelihood function of hpd is named as `nlhpd` and log likelihood function is named as `lhpd`. Similarly, the fitting function of hpd is named as `fhpd`. These functions have similar set up with other mixture models, so they are not detailed again.

```
x = rnorm(1000)
xx = seq(-4, 4, 0.01)
y = dnorm(xx)

# Hybrid Pareto provides reasonable fit for asymmetric heavy tailed distribution
# but not for cases such as the normal distribution
fit = fhpd(x, std.err = FALSE)
hist(x, breaks = 100, freq = FALSE, xlim = c(-4, 4))
lines(xx, dhpd(xx, nmean = fit$nmean, nsd = fit$nsd,
```

```
xi = fit$xi), col="red", lwd = 2)
abline(v = fit$u, col="red", lwd = 2)

# Notice that if tail fraction is included a better fit is obtained
fit2 = fnormgpdcon(x, std.err = FALSE)
lines(xx, dnormgpdcon(xx, nmean = fit2$nmean, nsd = fit2$nsd, u = fit2$u,
xi = fit2$xi), col="blue", lwd = 2)
abline(v = fit2$u, col="blue", lwd = 2)
legend("topright", c("hpd", "normgpdcon" ), col = c("red", "blue")
```

### 4.3.7 Hybrid Pareto with Single Continuous Constraint Model

A similar model can be suggested by splicing a normal distribution with GPD and set a continuity constraint on the density only not constraining it to also be continuous in first derivative at the threshold. The initial 5 parameters haven been reduced to threshold $u$, normal mean $\mu$, standard deviation $\beta$ and GPD shape $\xi$, which is consistent with previous continuity model. Again, note that the a key difference that $\phi_u$ is not included in this model, so there is no scaling of upper tail and bulk model. The parameter vector is $\theta = (u, \mu, \beta, \xi)$. The single continuous constraint hybrid Pareto mixture density is defined as:

$$f(x|\theta) = \begin{cases} \frac{1}{\tau} h(x|\mu, \beta) & x \leq u, \\ \\ \frac{1}{\tau} g(x|u, \sigma_u, \xi) & x > u, \end{cases}$$

where $h(.|\mu, \beta)$ is the normal density function and GPD density function is defined as $g(.|u, \sigma_u, \xi)$. The $\tau$ is the usual normalising constant and $\tau = 1 + H(u|\mu, \beta)$, where the 1 comes from the integration of the unscaled GPD and second term is from the usual normal component.

#### Base Function

The standard density, distribution , quantile, random number generation functions of the single continuous constraint hybrid Pareto model are similar with the original hybrid Pareto. The example of density function of single continuous constraint hybrid Pareto model:

- `dhpdcon(x, nmean = 0, nsd = 1, u = qnorm(0.9, nmean, nsd), xi = 0, log = FALSE)`

#### Likelihood and Fitting Function

An example of fitting the hybrid Pareto (with and without the constraint of continuous first derivative) and the normal with GPD tail with single continuity constraint is given below.

#### Examples

Figure 4.9 demonstrates that two hybrid Pareto models cannot provide a reasonable fit for the normal distribution but `normgpdcon` model could give a better fit.

```
x = rnorm(1000)
xx = seq(-4, 4, 0.01)
y = dnorm(xx)

# Two Hybrid Pareto models provide reasonable fit for asymmetric heavy tailed distribution
```

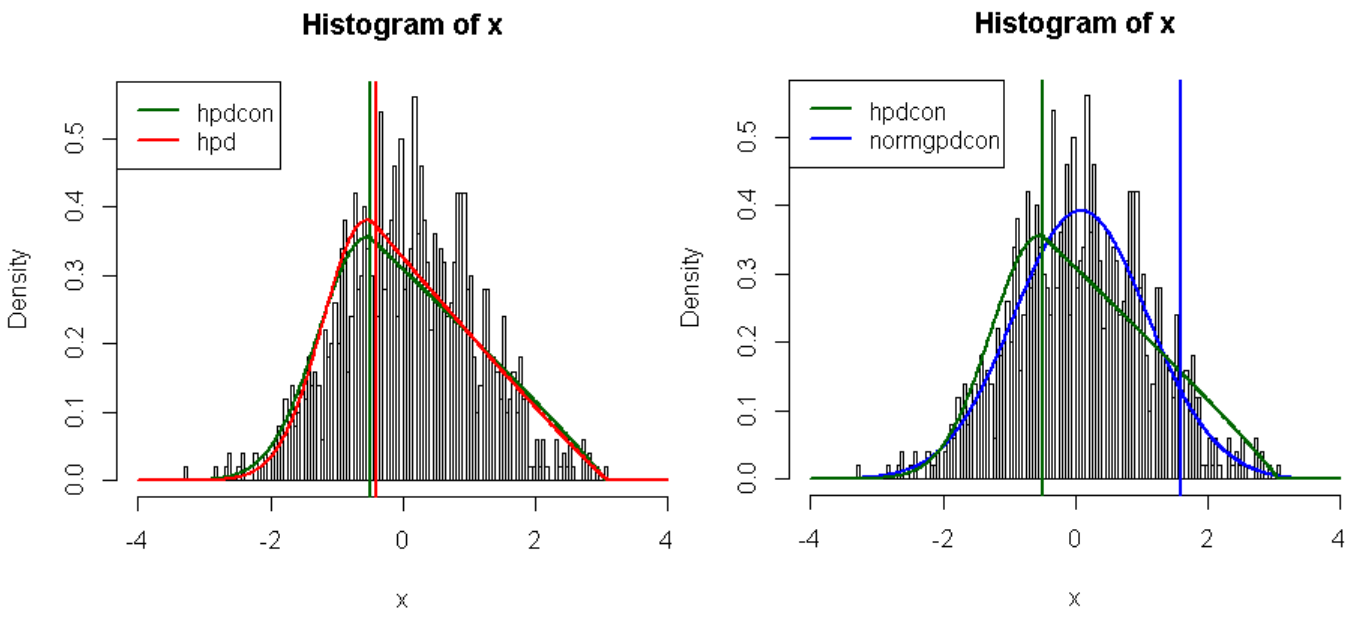



Figure 4.9: Left: Example of 1000 simulated data from a normal distribution with overlaid fitted `hpd` model with parameter vector of $(\mu = -0.606, \beta = 0.614, \xi = -0.372)$, and `hpdcon` model with parameter vector of $(\mu = -0.547, \beta = 0.651, u = -0.558, \xi = -0.472)$. Right: Example of 1000 simulated data from a normal distribution with overlaid fitted `hpdcon` model with parameter vector of $(\mu = -0.547, \beta = 0.651, u = -0.558, \xi = -0.472)$ and `normgpdcon` model with parameter vector of $(\mu = 0.007, \beta = 0.970, u = 1.315, \xi = -0.255)$.

```
# but not for cases such as the normal distribution
fitcon = fhpdcon(x, std.err = FALSE)
fit = fhpd(x, std.err = FALSE)
hist(x, breaks = 100, freq = FALSE, xlim = c(-4, 4))
lines(xx, dhpdcon(xx, nmean = fitcon$nmean, nsd = fitcon$nsd, u = fitcon$u,
    xi = fitcon$xi), col="darkgreen", lwd = 2)
abline(v = fitcon$u, col="darkgreen", lwd = 2)
lines(xx, dhpd(xx, nmean = fit$nmean, nsd = fit$nsd, xi = fit$xi), col="red", lwd = 2)
abline(v = fit$u, col="red", lwd = 2)
legend("topleft", c("hpdcon", "hpd" ), col = c("darkgreen", "red")
       ,lty = c(1, 1), lwd = c(2, 2))

# Notice that if tail fraction is included a better fit is obtained
fit2 = fnormgpdcon(x, std.err = FALSE)
hist(x, breaks = 100, freq = FALSE, xlim = c(-4, 4))
lines(xx, dnormgpdcon(xx, nmean = fit2$nmean, nsd = fit2$nsd, u = fit2$u,
   xi = fit2$xi), col="blue", lwd = 2)
lines(xx, dhpdcon(xx, nmean = fitcon$nmean, nsd = fitcon$nsd, u = fitcon$u,
   xi = fitcon$xi), col="darkgreen", lwd = 2)
abline(v = fit2$u, col="blue", lwd = 2)
abline(v = fitcon$u, col="darkgreen", lwd = 2)
legend("topleft", c("hpdcon", "normgpdcon" ), col = c("darkgreen", "blue")
```

## 4.4 Non-Parametric Mixture Models

MacDonald et al. (2011) and MacDonald et al. (2013) developed a non-parametric kernel density estimator based extreme value mixture models extending on that developed by Tancredi et al.

(2006). The former has been implemented in the first version of this package.

### 4.4.1 Kernel GPD Model

MacDonald et al. (2011) constructed a standard kernel density estimator as the bulk model below the threshold with GPD for the tail model. The distribution function of standard kernel GPD mixture model is given by:

**Bulk Model Based Tail Fraction Approach**

$$F(x|X,\lambda,u,\sigma_u,\xi,\phi_u) = \begin{cases} H(x|X,\lambda) & x \leq u, \\ (1-\phi_u) + \phi_u \times G(x|u,\sigma_u,\xi) & x > u, \end{cases}$$

where $\phi_u = 1 - H(u|X,\lambda)$.

**Parameterised Tail Fraction Approach**

$$F(x|X,\lambda,u,\sigma_u,\xi,\phi_u) = \begin{cases} (1-\phi_u)\frac{H(x|X,\lambda)}{H(u|X,\lambda)} & x \leq u, \\ (1-\phi_u) + \phi_u \times G(x|u,\sigma_u,\xi) & x > u, \end{cases}$$

where $H(.|X,\lambda)$ is the distribution function of the kernel density estimator and $G(.|u,\sigma_u,\xi)$ is the distribution function of GPD. The traditional kernel density estimator is defined as:

$$h(x;\mathbf{X},\lambda) = \frac{1}{n\lambda}\sum_{i=1}^{n} K\left(\frac{x-x_i}{\lambda}\right),$$

where $K(.)$ is the kernel function and usually to be a symmetric and unimodal probability density function and lambda is the bandwidth parameter. The kernel function usually meets the following conditions: $K(x) \geq 0$ and $\int K(x)\,\mathrm{d}x = 1$.

**Base Function**

The density function of kernel GPD model is given by:

- `dkdengpd(x, kerncentres, lambda = NULL, u = as.vector(quantile(kerncentres, 0.9)), sigmau = sqrt(6*var(kerncentres))/pi, xi = 0, phiu = TRUE, log = FALSE)`

where `lambda` is the bandwidth parameter of the kernel density estimator and `kerncentres` is the sample of data. The `phiu` can be used for determining either bulk model based or parameterised approach.

**Likelihood Function**

The negative log likelihood and log likelihood functions are pre-fixed as usual:

- `nlkdengpd(pvector, x, phiu = TRUE, finitelik = FALSE)`
- `lkdengpd(x, lambda = NULL, u = 0, sigmau = 1, xi = 0, phiu = TRUE, log = TRUE)`

### Fitting Function

Notice that the kernel centres are not needed in the likelihood functions as these are the data.

Two important practical issues arise with MLE for the kernel bandwidth:

- Cross-validation likelihood is needed for the kernel density estimator bandwidth parameter as the usual likelihood degenerates by Habbema et al. (1974) and Duin (1976). If the data has been heavily rounded, the bandwidth can be zero even if the using cross-validation likelihood. To overcome this issue an option to add a small jitter to the data has been included in the fitting inputs, using the `jitter` function, to remove the ties.
- For heavy tailed populations, the bandwidth is positively biased, giving a larger bandwidth. One solution to this problem is to splice the GPD to both the upper and lower tails (using the `gkg` function discussed below), as the bias comes from the lack of separation of the upper and lower order statistics in the two tails

The fitting function is given by:

- `fkdengpd(x, phiu = TRUE, pvector = NULL, add.jitter = FALSE, factor = 0.1, amount = NULL, std.err = TRUE, method = "BFGS", control = list(maxit = 10000), finitelik = TRUE, ...)`

`x` is a vector of sample data. `add.jitter` is a logical variable, if `add.jitter = TRUE`, then the jitter is needed for rounded data. `factor` and `amount` are the arguments passed to the `add.jitter`.

### Examples

An example program for fitting the kernel GPD model is now given in Figures 4.10, which highlights the usefulness of this model.

```
x = rnorm(1000, 0, 1)
fit = fkdengpd(x, phiu = FALSE, std.err = FALSE)
hist(x, 100, freq = FALSE, xlim = c(-4, 4))
xx = seq(-4, 4, 0.01)
lines(xx, dkdengpd(xx, x, fit$lambda, fit$u, fit$sigmau, fit$xi, fit$phiu), col="blue", lwd = 2)
abline(v = fit$u, col="blue", lwd = 2)
legend("topright", "kdengpd", col = "blue",
       ,lty = 1, lwd = 2)

# Try a bimodal data
x <- c(rnorm(1000,-1,0.7),rnorm(1000,3,1))

fit = fkdengpd(x, phiu = FALSE, std.err = FALSE)
hist(x, 100, freq = FALSE, xlim = c(-4, 6.5), main="Bimodal data example")
xx = seq(-5, 6, 0.001)
lines(xx, dkdengpd(xx, x, fit$lambda, fit$u, fit$sigmau, fit$xi, fit$phiu), col="blue", lwd = 2)
abline(v = fit$u, col="blue", lwd = 2)
legend("topright", "kdengpd", col = "blue",
       ,lty = 1, lwd = 2)
```

## 4.4.2 Two Tailed Kernel GPD Model

MacDonald et al. (2011) introduced a two tailed mixture model by splicing the standard kernel density estimator with two extreme value tail models. This two-tailed extreme value mixture

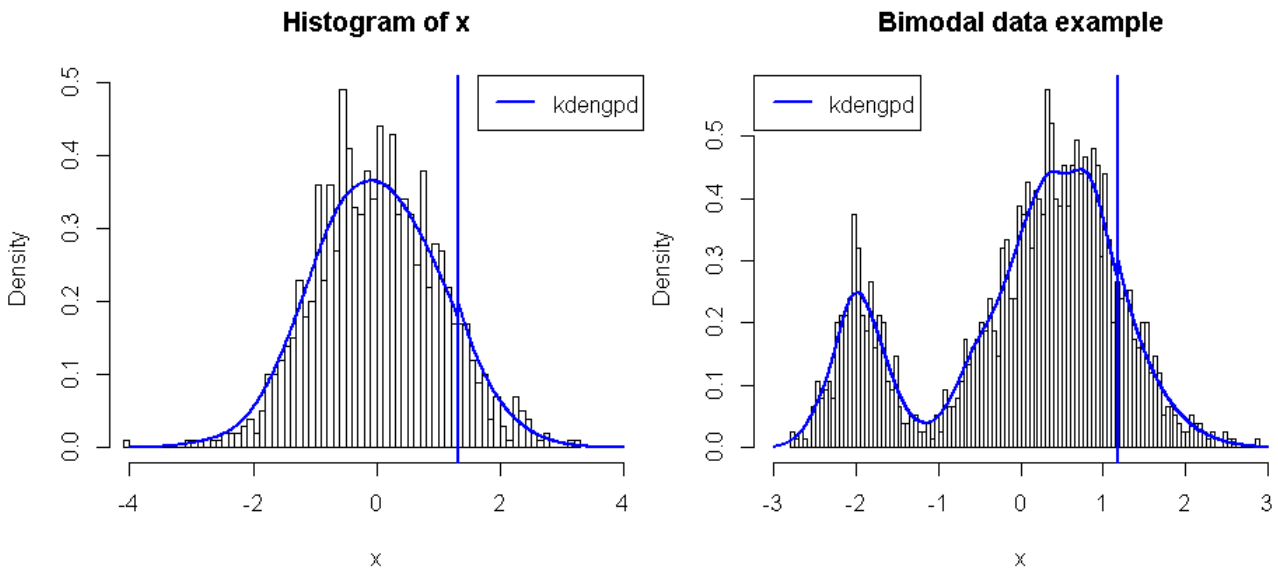



Figure 4.10: Left: Example of 1000 simulated data from a standard normal distribution with overlaid fitted `kdengpd` model with parameter vector of $(\lambda = 0.305, u = 1.048, \sigma_u = 0.522, \xi = -0.117)$. Right: Example of 2000 simulated data from a mixture of standard normals with overlaid fitted `kdengpd` model with parameter vector of $(\lambda = 0.201, u = 3.536, \sigma_u = 0.759, \xi = -0.260)$.

model also overcomes the inconsistency in the cross-validation likelihood estimation of the bandwidth for heavy tailed distributions, see MacDonald et al. (2011) for details.

The parameter vector is $\Theta = (X, \lambda, u_1, \sigma_{u_l}, \xi_l, \phi_{u_l}, u_2, \sigma_{u_r}, \xi_r, \phi_{u_r})$. The distribution function of two tailed mixture model is given by:

**Parameterised Tail Fraction Approach**

$$F(x|\Theta) = \begin{cases} \phi_{u_l}\left\{1 - G(-x| - u_l, \sigma_{u_l}, \xi_l)\right\} & x < u_l, \\ \phi_{u_l} + (1 - \phi_{u_l} - \phi_{u_r})\dfrac{H(x|X,\lambda) - H(u_l|X,\lambda)}{H(u_r|X,\lambda) - H(u_l|X,\lambda)} & u_l \le x \le u_r, \\ (1 - \phi_{u_r}) + \phi_{u_r} G(x|u_r, \sigma_{u_r}, \xi_r) & x > u_r, \end{cases}$$

where $\phi_{u_l}$ and $\phi_{u_r}$ are estimated as the sample proportions less than threshold $u_l$ and above the threshold $u_r$ respectively. $G(-x| - u_l, \sigma_{u_l}, \xi_l)$ is the unconditional GPD function for $x < u_l$ and $G(x|u_r, \sigma_{u_r}, \xi_r)$ is the unconditional GPD function for $x > u_r$.

**Bulk Model Based Tail Fraction Approach**

$$F(x|\Theta) = \begin{cases} \phi_{u_l}\left\{1 - G(-x| - u_l, \sigma_{u_l}, \xi_l)\right\} & x < u_l, \\ H(x|X,\lambda) & u_l \le x \le u_r, \\ (1 - \phi_{u_r}) + \phi_{u_r} G(x|u_r, \sigma_{u_r}, \xi_r) & x > u_r, \end{cases}$$

where $\phi_{u_l}$=$H(u_l, X, \lambda)$ and $\phi_{u_r} = 1 - H(u_r, X, \lambda)$.

### Base Function

The density function of two tailed kernel GPD model is defined as:

- `dgkg(x, kerncentres, lambda = NULL,`
  `ul = as.vector(quantile(kerncentres, 0.1)), sigmaul = sqrt(6*var(kerncentres))/pi,`
  `xil = 0, phiul = TRUE,`
  `ur = as.vector(quantile(kerncentres, 0.9)), sigmaur = sqrt(6*var(kerncentres))/pi,`
  `xir = 0, phiur = TRUE, log = FALSE)`,

The basic use of the standard function will not repeated again.

### Likelihood and Fitting Function

The likelihood and fitting functions are defined as usual:

- `nlgkg(pvector, x, phiul = TRUE, phiur = TRUE, finitelik = FALSE)`
- `lgkg(x, lambda = NULL,`
  `ul = as.vector(quantile(x, 0.1)), sigmaul = 1, xil = 0, phiul = TRUE,`
  `ur = as.vector(quantile(x, 0.9)), sigmaur = 1, xir = 0, phiur = TRUE, log = TRUE)`
- `fgkg(x, phiul = TRUE, phiur = TRUE, pvector = NULL, add.jitter = FALSE,`
  `factor = 0.1, amount = NULL, std.err = TRUE, method = "BFGS",`
  `control = list(maxit = 10000), finitelik = TRUE, ...)`

### Examples

An example program for fitting the two tailed kernel GPD model is now given in Figures 4.11, which highlights the flexibility of this model.

```
x = rt(1000, 3, 1)
fit = fgkg(x, phiul = FALSE, phiur = FALSE, std.err = FALSE)
hist(x, 100, freq = FALSE, xlim = c(-8, 10))
xx = seq(-8, 10, 0.01)
lines(xx, dgkg(xx, x, fit$lambda, fit$ul, fit$sigmaul, fit$xil, fit$phiul, fit$ur, fit$sigmaur, fit$xir, fit$phiur)
     , col="blue", lwd = 2)
abline(v = fit$ul, col="blue", lwd = 2)
abline(v = fit$ur, col="blue", lwd = 2)
legend("topright", "gkdeng", col = "blue",
       ,lty = 1, lwd = 2)

# Try a bimodal data
x <- c(rnorm(1000,-1,0.7),rnorm(1000,3,1))

fit = fgkg(x, phiul = FALSE, phiur = FALSE, std.err = FALSE)
hist(x, 100, freq = FALSE, xlim = c(-4, 6.5), main="Bimodal data example")
xx = seq(-4, 6.5, 0.01)
lines(xx, dgkg(xx, x, fit$lambda, fit$ul, fit$sigmaul, fit$xil, fit$phiul, fit$ur, fit$sigmaur, fit$xir, fit$phiur)
     , col="blue", lwd = 2)
abline(v = fit$ul, col="blue", lwd = 2)
abline(v = fit$ur, col="blue", lwd = 2)
legend("topright", "gkdeng", col = "blue",
       ,lty = 1, lwd = 2)
```

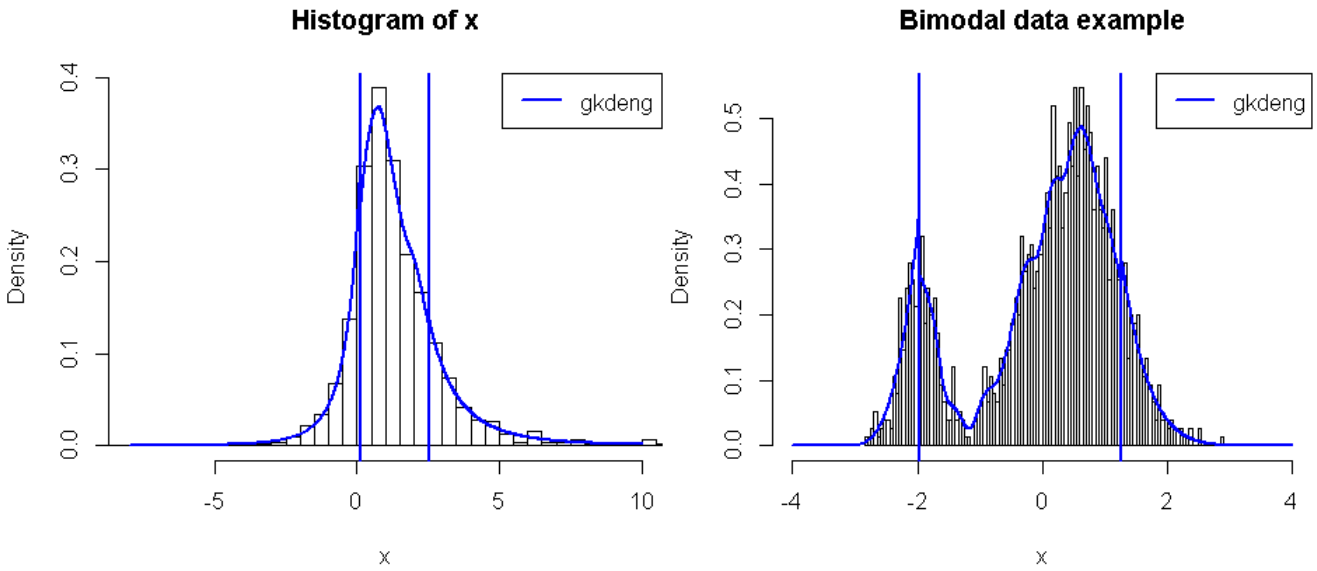



Figure 4.11: Left: Example of 1000 simulated data from Student-t(3,1) distribution with with overlaid fitted `gkg` model with parameter vector of $(\lambda = 0.347, u_l = -0.230, \sigma_{u_l} = 0.494, \xi_l = 0.334, u_r = 3.617, \sigma_{u_r} = 1.727, \xi_r = 0.171)$. Right: Example of 2000 simulated data from a mixture of standard normal distributions with with overlaid fitted `gkg` model with parameter vector of $(\lambda = 0.233, u_l = -1.565, \sigma_{u_l} = 0.472, \xi_l = -0.218, u_r = 3.686, \sigma_{u_r} = 0.669, \xi_r = -0.225)$.

### 4.4.3 Boundary Corrected Kernel GPD Model

MacDonald et al. (2013) combine a boundary corrected kernel density estimator for bulk model sliced at the threshold with a point process representation of GPD tail model. The boundary correction kernel density estimator tries to reduce the inherent bias of the kernel density estimator. The distribution function of boundary corrected mixture model is given by:

**Bulk Model Based Tail Fraction Approach**

$$F(x|X, h_{BC}, u, \sigma_u, \xi, \phi_u) = \begin{cases} H_{BC}(x|X, \lambda_{BC}) & x \leq u, \\ (1 - \phi_u) + \phi_u \times G(x|u, \sigma_u, \xi) & x > u, \end{cases}$$

where $\phi_u = 1 - H_{BC}(u|X, h)$.

**Parameterised Tail Fraction Approach**

$$F(x|X, h_{BC}, u, \sigma_u, \xi, \phi_u) = \begin{cases} (1 - \phi_u)\frac{H_{BC}(x|X,\lambda_{BC})}{H_{BC}(u|X,\lambda_{BC})} & x \leq u, \\ (1 - \phi_u) + \phi_u \times G(x|u, \sigma_u, \xi) & x > u, \end{cases}$$

where $H_{BC}(x|X, \lambda_{BC})$ is the distribution function of the boundary correction kernel density estimator.

**Base Function**

The density functionis given by:

- `dbckdengpd(x, kerncentres, lambda = NULL, u = as.vector(quantile(kerncentres, 0.9)), sigmau = sqrt(6*var(kerncentres))/pi, xi = 0, phiu = TRUE, bcmethod = "simple", proper = TRUE, nn = "jf96", offset = 0, xmax = Inf, log = FALSE)`

The density, distribution, quantile function of boundary corrected kernel GPD model have very similar set up with the standard kernel GPD model. Hence, the basic use of these functions will not repeat here. See help(dbckdengpd) for details of boundary correction methods that have been implemented.

### Likelihood and Fitting Function

The likelihood functions and fitting function are very similar with other mixture models.

- `nlbckdengpd(pvector, x, phiu = TRUE, finitelik = FALSE)`
- `lbckdengpd(x, lambda = NULL, u = 0, sigmau = 1, xi = 0, phiu = TRUE, log = TRUE)`
- `fbckdengpd(x, phiu = TRUE, pvector = NULL, add.jitter = FALSE, factor = 0.1, amount = NULL, bcmethod = "simple", proper = TRUE, nn = "jf96", offset = 0, xmax = Inf, std.err = TRUE, method = "BFGS", control = list(maxit = 10000), finitelik = TRUE, ...)`

The basic use of these functions will not be detailed again.

# Chapter 5

# Simulation Study

## 5.1 Introduction

The simulation study is to assess the performance of the different extreme value mixture model by approximating some standard parametric distributions. In particular, to focus on their performance for estimating tail quantiles. These population distributions are chosen to represent examples of symmetric and asymmetric behaviors with 3 types of lower and upper tail behaviors. Various mixture models can be classified as a single tail model with the full support, a single tail model with bounded positive support or two tail model with the full support. This study will examine only the relevant population distributions for the suitable mixture model.

It is common practice that the population distribution can be referred to one of the three types of distribution due to the limiting tail behaviors defined by the shape parameter $\xi$ as type I ($\xi = 0$), type II ($\xi > 0$) and type III ($\xi < 0$). These three types of distribution have been mentioned in chapter 2.

In the simulation study, 100 replicates with sample size 1000 or 5000 are simulated for each distribution. The estimated 90%, 95%, 99%, 99.9% quantiles of the single tailed extremal mixture models will be calculated to compare with true quantiles under different population distributions while the estimated 0.1%, 1%, 5%, 10%, 90%, 95%, 99% and 99.9% quantiles will be calculated for the two tailed extremal mixture models. The RMSE (the root of mean squared error) is then used for measuring the performance of the estimated quantile with true quantile at each population distribution. A 95% bootstrap percentile interval using 100,000 bootstrap samples of the RMSE is obtained to explore evidence for the difference in performance.

It is expected that the estimated quantile tends to have higher uncertainty (higher RMSE) as the quantile gets close to the tail of the distribution. The estimated quantile also tends to have larger RMSE if the population distribution is heavy tailed. All population distributions in this sensitive analysis have a single mode and it is worth mentioning that non-parametric kernel density estimator based models and the mixture of gamma GPD model can cope with more complex data structure (e.g. bimodal or multimodal).

## 5.2 Simulation Parameters for Inference

The section will explain how the initial value of the parameters for likelihood optimisation or Bayesian Inference MCMC were chosen.

### 5.2.1 Initial Value For The Likelihood Inference Based Model

The initial value of the parameter sometimes has a great impact on the likelihood inference as the mixture models often have multiple local modes. In order to compare the model performance differences which are not caused by the sensitive initial values, it is reasonable to make a few assumptions about the initial value of parameters. The initial threshold $u$ is assumed as the 90% quantile of the data. According to the gpd.fit function from `ismev` package in `R`, the initial values for the GPD shape $\xi$ and scale $\sigma_u$ parameter have been chosen as 0.1 and $\sqrt{(6 \times var(data))}/\pi$ which is the standard deviation for the Gumbel case of the GEV distribution. Thus, this simulation study follows this rule. The initial values for the bulk model are essentially selected as the true value, assuming the bulk model extends over the entire range of support.

#### 5.2.1.1 Optimization Method And Quantile Estimation

There are several different numerical methods to search for the MLE in `R`. The `optim` function which includes various optimization routines has been adopted. The quasi-Newton based "BFGS" method is used in the simulation study.

The estimated parameters are plugged into the relevant quantile function for quantile inference. In order to make sure the likelihood function converges, the maximum number of iterations has been changed to 20000. It is important to note that the parameters are only be used for quantile estimation when the likelihood function converges.

### 5.2.2 Initial Value for the Bayesian Inference Based Model

For the Bayesian inference is also evaluated for the kernel GPD and boundary corrected kernel GPD based extreme value mixture models, there are 5 parameters in the model including kernel bandwidth $h$, threshold $u$, point process shape $\xi_{pp}$, scale $\sigma$ and location parameter $\mu$. The point process shape $\xi_{pp}$ and GPD shape $\xi$ are the same. The point process scale $\sigma$ and GPD scale parameters are related by $\sigma = \sigma_u - \xi(u - \mu)$.

The initial value of the threshold $u$ is the 90% quantile for MCMC. The shape, scale and location for point process model are found by maximizing the point process log-likelihood function. The initial value for point process location parameter $\mu$ is set to equal 0.9*$u$. Moreover, the scaling variable $n_b$ from the point process likelihood function is equal to number of data greater than the threshold, following the advice given by Wadsworth and Tawn (2012).

The initial value of the bandwidth is found by minimizing the asymptotic mean integrated squared error (AMISE). When the normal distribution has been used as the kernel function, the optimal choice of bandwidth for a normal population is given by:

$$h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5},$$

where $n$ is refer to the number of data below the threshold $u$ and $\hat{\sigma}$ is refer to the standard deviation of the data below the threshold $u$. This rule refers as the Silverman's Rule of Thumb or normal reference rule. Consequently, this formula gives the initial value of $h$ for the MCMC.

#### 5.2.2.1 Prior Information in Bayesian Inference

For Bayesian inference, we also need to provide the prior information about each parameter. In order to let the data speak for themselves, very diffuse priors have been used in the simulation study. The details of the prior structure of the kernel GPD model and boundary corrected kernel GPD model can be found in MacDonald et al. (2011) and MacDonald et al. (2013). It is reasonable to assume that the prior for PP parameters $(\mu, \sigma, \xi)$ are independent with the threshold $u$ and the bandwidth parameter $h$:

$$\pi(h, u, \mu, \sigma, \xi) = \pi(h)\pi(u)\pi(\mu, \sigma, \xi).$$

The prior for the point process parameter, $\pi(\mu, \sigma, \xi)$, is a trivariate normal distribution with a large variances to indicate little prior information. Each point process parameter has been suggested a mean of zero and a standard deviation of 100 except the PP scale parameter is on a log scale.

The prior for the bandwidth is based on an inverse gamma distribution $(d_1, d_2)$ to specify the precision parameter $1/h^2$, where $d_1$ and $d_2$ are the hyperparameters. Brewer (2000) suggests to specify $d_1 > 0.5$ and $d_2 = \left[\dfrac{\Gamma(d_1)}{\Gamma(d_1 - 0.5)}\right]^2$.

The prior for the threshold is suggested by Behrens et al. (2004), which is a truncated normal distribution truncate at the minimum of the data. The mean of the truncated normal distribution is the 90th quantile with standard deviation of $\sqrt{12}$.

#### 5.2.2.2 Quantile Calculation

The MCMC algorithm runs for 10000 iterations with a burn-in period of 1000, which resulting a 9000 posterior draws for each simulation. The posterior mean of the each parameter from the 9000 posterior draws is used to estimate the corresponding quantile. The quantile functions for the kernel GPD and boundary corrected GPD model require integrating the distribution function of kernel GPD or boundary corrected GPD, respectively. It is possible that for some posterior mean values, the integration could fail. In this case, another simulation will be carried out.

## 5.3 A Single Tail Model With Full Support

This section will investigate different extremal mixture models including normal GPD mixture, continuous normal GPD mixture, standard kernel GPD, hybrid Pareto and single continuous constraint hybrid Pareto model. The first three models can be implemented by either bulk model based tail fraction approach or parameterized tail fraction approach. Hence, these extremal mixture models can be classified into two groups:

- bulk model based tail fraction approach,
- parameterized tail fraction approach.

Five different populations will be considered in this case including normal($\mu = 0, \sigma = 3$), Student-t($v = 3, \mu = 0$), Gumbel($\sigma = 1$), negative Weibull($l = 10, k = 5$) and Student-t($v = 4, \mu = 1$) shown in Figure 5.1. These provide a wide range of upper tail and bulk features. Both normal (type I) and Student-t($v = 3, \mu = 0$) (type II) are symmetric distributions while Gumbel($\sigma = 1$) is an asymmetric distribution with a type III lower tail and a type I upper tail and negative Weibull($l = 10, k = 5$) is an asymmetric distribution with a short lower and upper tail (type III). Student-t($v = 4, \mu = 1$) is an asymmetric distribution with a heavy lower and upper tail (type II) and the lower tail is not as heavy as the upper tail.

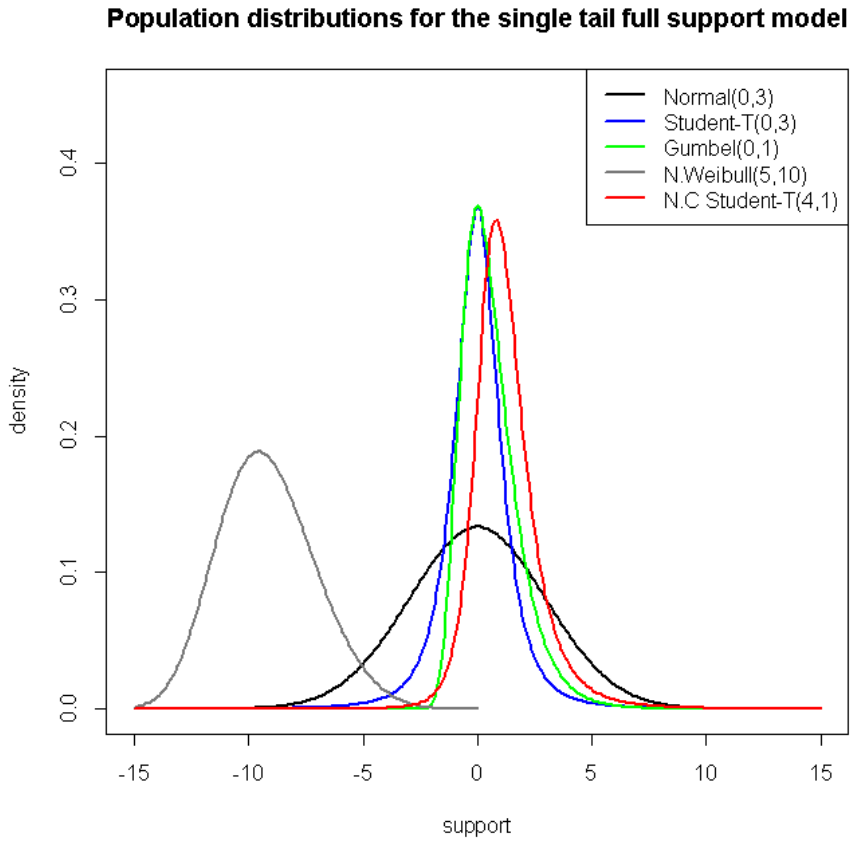



Figure 5.1: Population distributions used for the single tail full support model in the simulation study.

The 5 mixture models will be abbreviated as:

- CONnormGPD $\Rightarrow$ the continuous constraint normal GPD mixture model;
- normGPD $\Rightarrow$ the normal GPD mixture model;
- hPD $\Rightarrow$ hybrid Pareto model;

- shPD ⇒ the single continuous constraint hybrid Pareto model;
- kernelGPD ⇒ the kernel density estimator GPD model.

### 5.3.1 Parameterised Tail Fraction Approach

The overall results show that the RMSE tends to get larger when the tail probability decreases and the RMSE also become larger if the population distribution is heavy tail. It should be noticed that the model performance is only comparable under the same population distribution, as the RMSE generally increases with quantile level. High quantile estimates (99.9%) are fairly robust to the bulk model while the lower quantile estimates (90%, 95%, 99%) are more sensitive to the choice of the bulk model. In fact, all 5 models have a similar bootstrap percentile intervals and the associated intervals overlap with each other, which suggests that there are no significant model performance differences. This seems to suggest that for a sufficiently high quantile, the GPD will asymptotically dominate the upper tail while the distribution below the threshold no longer contributes much at the level of extremity.

From Table 5.1, both hPD and shPD perform poorly in most population distributions at lower quantile (90%, 95%, 99%), except the hPD works reasonably well for a Gumbel distribution at 99% quantile. In particular hPD and shPD have very poor fit for symmetrical distributions. There is no overlap between the 95% bootstrap percentile confidence interval of the best fit model and the associated interval of the hPD or shPD, which confirms they are significant worse in terms of RMSE compared to the best mixture model for a given quantile (90%, 95%, 99%). When the sample size increase to 5000, from Table 5.2, the performance of the two hybrid Pareto models (hPD and shPD) doesn't improve. There is no significant improvement for two hybrid Pareto models when the population distribution sample size is 5000 as the RMSE gets larger in most cases of the lower quantile. Initially, it is surprising that the performance is so poor for a normal distribution. Even though Carreau and Bengio (2008) propose the hPD model by combing a normal and GPD and set two continuous constraints to make the the density and first derivative continuous at the threshold, this model is different from the continuous constraint normal GPD model (CONnormGPD). In particular with the key difference is that they ignore the usual tail fraction $\phi_u$ scaling of the conditional GPD, which means that they don't treat the GPD as a conditional model. Essentially the hPD seems to only perform well for certain asymmetric distributions, but even this it does not work for the asymmetric non-central t distribution considered here.

There are some interesting results from normal GPD model (normGPD) and continuous normal GPD model (CONnormGPD). For the Student-t($v = 3, \mu = 0$) and Student-t($v = 4, \mu = 1$) distribution, it seems that CONnormGPD performs better than normGPD as the RMSE of the CONnormGPD is smaller in general. No such conclusions can be drawn for the normal distribution. It should be noted that the differences of the RMSE at different quantiles from two models are not likely to reach the level of a statistical significance. However, if the underlying distribution is Gumbel, it appears that normGPD outperforms CONnormGPD. There are no significant performance differences of these two models for a negative Weibull distribution ($l = 10, k = 5$). When the sample size increases to 5000, the normGPD has achieved a significant reduction of RMSE in the 90%, 95%, 99% quantile of Gumbel distribution and 95%, 99% quantile at the negative Weibull compared with corresponding RMSE of CONnormGPD. These

suggest that the normGPD is more flexible than the CONnormGPD, which is obvious as the CONnormGPD restricts GPD scale parameter in terms of other parameters. Hence the normal bulk model has influenced the tail fit.

It is apparent that in the majority of cases, the kernel GPD model (kernelGPD) has the best performance in the lower quantiles (90%, 95%, 99%) in all 5 population distribution in terms of the smallest RMSE. What is also apparent from the result is that kernelGPD still achieves reasonably good estimates at the 99.9% quantile. Hence, excellent results can be obtained in the lower quantile and reasonable estimate at very extreme quantile (99.9%) can also be achieved. KernelGPD has overwhelming advantage of estimating population distribution quantiles compared with using other models as it doesn't assume a particular function form. However, it suffers the computational burden. For the Bayesian inference based kernelGPD model, the MCMC takes about 1 hour and 15 hours for a population distribution with a sample size of 1000 and 5000. The parametric bulk mixture models (normGPD and CONnormGPD) only take a few seconds to get MLEs. Section 5.6 will introduce 2 likelihood inference based kernel GPD models, which will significantly reduce the computational time compared with using the Bayesian inference but it is still higher than the other models.

In most cases, normGPD, CONnormGPD and kernelGPD model have outperformed hPD and shPD model in the lower quantile (90%, 95%, 99% quantile) and all normGPD, CONnormGPD and kernelGPD model have achieved significant reduction in RMSE shown by the non-overlapping 95% bootstrap intervals. When the sample size increases to 5000, normGPD, CONnormGPD and kernelGPD have achieved significant reduction in RMSE compared with two hybrid Pareto models not only at lower quantile, but also at the high quantile (99%). This could highlight that the model formulations of the two hybrid Pareto based models are not ideal as the two models ignore the usual tail fraction scaling.

### 5.3.2 Bulk Model Based Tail Fraction Approach

This section aims to illustrate the model performance differences of the two approaches for the normGPD, CONnormGPD and kernelGPD. In general, the results are in line with the those for the parameterized approach. Therefore I will focus here on the key differences between the two approaches.

According to Table 5.3, there are no significant performance differences of the bulk approach kernelGPD compared with using the parameterized approach kernelGPD except the Student-t($v = 3, \mu = 0$) distribution. If the lower tail is heavy, the kernelGPD doesn't work well as the bandwidth is bias towards higher values and leading to over smoothing. However, the bulk approach based kernelGPD give a similar fit to Student-t($v = 4, \mu = 1$) distribution, as the lower tail is not as heavy as the upper tail. The bulk approach based kernelGPD is able to provide sufficient flexibility in this case. On the other hand, the Student-t($v = 3, \mu = 0$) has heavy lower and upper tails and the bulk approach based kernelGPD fails to provide a reasonable fit due to this bias in the bandwidth estimation.

The performances of the normGPD model is significantly worse than the performances of the parameterized approach normGPD at 90%, 95% and 99% quantile for the Student-t($v = 3, \mu = 0$)

and Gumbel($\sigma = 1$) distributions and 90%, 95% quantile of Student-t($v = 4, \mu = 1$) distribution. Similar conclusions can also be drawn for the CONnormGPD at 90%, 95% quantile for the Student-t($v = 3, \mu = 0$) distribution.

Table 5.4 provides further evidence that the parameterised approach based model is preferred if the bulk model is misspecified. There is evidence to indicate that normGPD doesn't achieve a similar RMSE a lower quantile for the Student-t($v = 3, \mu = 0$), Gumbel($\sigma = 1$), negative Weibull($l = 10, k = 5$) and Student-t($v = 4, \mu = 1$) distributions compared with the RMSE obtained from parameterised approach normGPD. Moreover, the CONnormGPD model doesn't perform well for the Student-t($v = 3, \mu = 0$) at the lower quantile.

All 5 models have very similar model fit at the 99.99% quantile once the sampling variability is taken into account in Table 5.3. When the sample size is 1000, normGPD, CONnormGPD and kernelGPD model still outperform the hPD and shPD model in the lower quantile (90%, 95%, 99% quantile) and in the high quantile as the sample size increases to 5000.

The take home message is that the parameterised approach based models provide a better model fit in terms of a smaller RMSE in many cases as it is more flexible than the bulk approach. The parameterised approach encompasses the bulk approach as a special case. If the bulk model is mis-specified, the parameterised approach based model is more flexible, such that even if the bulk model fit is poor the threshold is less effected and GPD is able to still provide a good fit. Whereas the lack of flexibility of the bulk approach means the threshold is effected and the GPD has to worker harder to fit well. This become more evident when the sample size of the simulation distribution is large. However, if the bulk model is correct, the bulk model parametrization of the tail fraction is preferred, and further the continuity constraint version of the model also has slight performance advantage.

## 5.4 A Single Tail Model With Bounded Positive Support

This section will examine several extremal mixture models including gamma GPD, continuous gamma GPD, Weibull GPD, continuous Weibull GPD, boundary corrected kernel GPD, mixture of gamma GPD and the dynamically weighed mixture model. The first 4 models can be implemented by the two tail fraction parameterization approaches.

The 7 mixture models will be abbreviated as:

- gammaGPD $\Rightarrow$ the gamma GPD mixture model;
- CONgammaGPD $\Rightarrow$ the continuous constraint gamma GPD mixture model;
- WeibullGPD $\Rightarrow$ Weibull GPD model;
- CONWeibullGPD $\Rightarrow$ the continuous constraint Weibull GPD model;
- BCkernelGPD $\Rightarrow$ the boundary corrected kernel density estimator GPD model;
- gamMixGPD $\Rightarrow$ the mixture of gamma GPD model with two gamma component;
- DWM $\Rightarrow$ the dynamically weighted mixture model.

The mixture of gamma GPD model could have many gamma components. In order to differ this model with the gamma GPD mixture model (gammaGPD), I have adopted a just two components for the bulk model as relatively simple population distributions are considered.

Seven different simulated population distributions with either an exponential upper tail (type I) or heavy upper tail (type II) are used to explore the performance of these 7 extremal mixture models. These population distributions are shown in Figure 5.2. They also exhibit a range of behaviors in the bulk and at the boundary. The seven population distributions are given by:

- gamma($\alpha = 1, \beta = 2$) - pole at zero and exponential tail;
- gamma($\alpha = 2, \beta = 2$) - sharp decay to zero at boundary and exponential tail;
- gamma($\alpha = 5, \beta = 1$) - proper lower tail decay to zero and exponential tail;
- non-central chi-squared($v = 2, \lambda = 2$) - shoulder at the origin and exponential tail;
- non-central chi-squared($v = 2, \lambda = 6$) - non-zero lower tail at boundary and exponential tail;
- inverse-gamma($\alpha = 4, \beta = 4$) - proper lower tail with heavy upper tail;
- inverse-gamma($\alpha = 4, \beta = 8$) - proper lower tail with heavy upper tail.

### 5.4.1 Parameterised Tail Fraction Approach

Generally speaking, the performance of the different models is mixed and no model can be described as the best model under all the different tail behaviors and population distributions based on Table 5.5, Table 5.6, Table 5.7 and Table 5.8. The results can be divided into two parts as follows: lower quantile (90%, 95%, 99%) and higher quantile (99.9%). It is important to reiterate that the RMSE tends to become larger for higher quantiles, this is consistent with the results obtained in Section 5.3 for the mixture models defined on the real line.

The 4 parametric bulk form mixture models can give very similar fit in the lower quantile shown by the overlapping bootstrap intervals. The main concern for the 4 parametric bulk form mixture models is whether they can provide good estimations if they are mis-specified. Generally speaking, the bulk model is able to provide reasonable fit if the bulk model is correctly specified. For example, the gammaGPD model can give a good tail inference for the gamma population distribution. Further, in the case of a correctly specified bulk model, the continuity constraint general model performs very slightly better.

On the other hand, if the bulk model is mis-specified, the bulk model based mixture model suffers. Moreover, the corresponding mixture model with a continuous constraint would perform worse in general as it is less flexible.

The gammaGPD outperfms the CONWeribullGPD model for the gamma($\alpha = 5, \beta = 1$) and gamma($\alpha = 2, \beta = 2$) distributions although the 95% bootstrap pencetile intervals just slightly

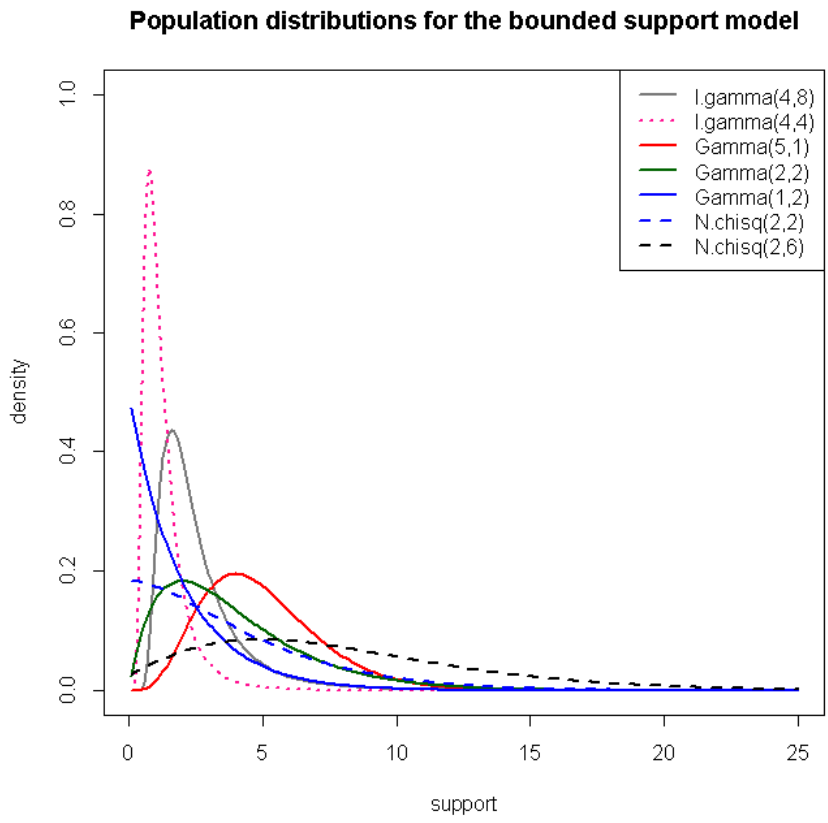



Figure 5.2: Population distributions used in the simulation study.

overlap. The gammaGPD and WeibullGPD also achieve significantly better performance compared to CONgammGPD and CONWeibullGPD at 95% quantile of the inverse-gamma($\alpha = 4, \beta = 4$) distribution. More evidence can be found that when the sample size of the distribution increase to 5000. CONWeibullGPD gives a poor fit compared to WeibullGPD at the gamma($\alpha = 5, \beta = 1$), gamma($\alpha = 2, \beta = 2$), inverse-gamma($\alpha = 4, \beta = 4$) and inverse-gamma($\alpha = 4, \beta = 8$). The gammaGPD performs significantly better to the CONgammaGPD at the 95% quantile of the non-central chi-squared($v = 2, \lambda = 6$), inverse-gamma($\alpha = 4, \beta = 4$) and inverse-gamma($\alpha = 4, \beta = 8$).

There is no evidence that mixture of gamma GPD model with 2 components performs better than other models in the 90%, 95% and 99% quantile due to the overlapping bootstrap percentile intervals. This model also faces the bulk model mis-specification issue.

The dynamically weighed mixture model achieves reasonable lower quantile estimate in terms of a similar RMSE compared to the best model fit. Not only does this model give reasonable fit to the exponential tail distributions, but it gives satisfactory tail estimation for the heavy tail distributions as well. However, this model is not as flexible as the 4 parametric form GPD models. In some cases, these 4 parametric form GPD models outperform the dynamically weighted mixture model. Although the sample size increases to 5000, the dynamically weighed mixture model is still slightly poorer than for the gamma($\alpha = 5, \beta = 1$) distribution compared with using the other models.

It should be noticed that it is sometimes difficult to judge whether the bulk model is mis-specified or not for a real applications before fitting the data. Hence, a more flexible approach is highly sought after. It can be seen that boundary corrected kernel GPD model gives an excellent quantile estimate in the majority of the lower quantiles under different distributions. Boundary corrected kernel GPD model is best for at least one case for tail estimation in the lower quantile (90%, 95%, 99%) in the following 5 distributions: gamma($\alpha = 2, \beta = 2$), gamma($\alpha = 1, \beta = 2$), non-central chi-squared($v = 2, \lambda = 2$), non-central chi-squared($v = 2, \lambda = 6$), inverse-gamma($\alpha = 4, \beta = 4$) and inverse-gamma($\alpha = 4, \beta = 8$) distributions. Although it fails to achieve the best tail estimation in non-central chi-squared($v = 2, \lambda = 2$) and gamma($\alpha = 5, \beta = 1$) distribution, the results obtained from boundary corrected kernel GPD is very close to the best fit model. As mentioned earlier, the differences between the RMSE obtained from boundary corrected kernel GPD and the RMSE from the best model won't necessarily reach the level of significance due to the overlapping intervals. However, the Bayesian inference based BCkernelGPD model is time consuming and computationally intensive. The MCMC takes about 3.5 hours and 60 hours for a sample size of 1000 and 5000, respectively, but the likelihood inference reduces this to 90 seconds and 15 minutes respectively.

Due to the computational burden of the boundary corrected kernel GPD model, this model will only apply to 3 population distributions (gamma($\alpha = 1, \beta = 2$), non-central chi-squared($v = 2, \lambda = 2$) and inverse-gamma($\alpha = 4, \beta = 8$)) with a sample size of 5000, which contains various lower and upper tails. According to Table 5.7 and Table 5.8, the boundary corrected kernel GPD model is able to provide very good quantile estimates across a range of population distributions than the other parametric and semi-parametric models.

For the 99.9% quantile, all 7 models achieve very similar performance based on the RMSE in most cases. It is likely that for very high quantiles (e.g. 99.9% quantile), the bulk model doesn't influence the GPD fit to the tail.

### 5.4.2 Bulk Model Based Tail Fraction Approach

As can be seen from Table 5.9 and Table 5.10, the overall result is that the RMSE of the 4 parametric bulk mixture models are almost identical compared to using the parameterized approach. However, the RMSE of the gammaGPD, CONgammaGPD and weibullGPD for the non-central chi-squared($v = 2, \lambda = 6$) distribution is not consistent with the RMSE from parameterised approach, which seems to suggest that the bulk model based approach is not always reliable especially at the lower quantile, this is due to the model mis-specification issue. As expected, there is no significant model fit differences at the 99.9% quantile.

There is evidence to suggest that bulk model based approach is not as flexible as the parameterised approach in Table 5.11 and Table 5.12. The RMSE of WeibullGPD for the gamma($\alpha = 5, \beta = 1$), inverse-gamma($\alpha = 4, \beta = 4$) and inverse-gamma($\alpha = 4, \beta = 8$) are higher than the RMSE obtained from the parameterised approach. Similar conclusions can be drawn for the gammGPD at the non-central chi-squared($v = 2, \lambda = 2$), non-central chi-squared($v = 2, \lambda = 6$) and inverse-gamma($\alpha = 4, \beta = 8$) distribution. Moreover, the CONgammGPD also suffers with the non-central chi-squared($v = 2, \lambda = 2$) and non-central chi-squared($v = 2, \lambda = 6$).

Based on the above, it seems that the parameterised approach based model is able to provide more flexibility in quantile estimation for the various population distributions as the bulk ap-

proach is a special case of the parameterised approach. If the bulk model is mis-specified, the parameterised approach based model can still give a reasonable tail quantile estimates but this doesn't always hold. On the other hand, if the bulk model is correctly specified, there is little differences between the parameterised tail fraction approach and bulk model based tail fraction approach.

## 5.5 Two Tail Model With Full Support

This section will compare the model fit of 3 mixture models including the two tail normal GPD (GNG), single continuous constraint normal GPD (CONGNG) and two tail kernel GPD model (GkerG). The GNG and CONGNG are implemented by the two approaches for tail fraction specification. All 5 population distributions from Section 5.3 will be used. The GkerG using likelihood inference takes less than 1 minute for a sample size of 1000 and 15 minutes for a sample size of 5000. The other models only take a second to get likelihood estimates.

### 5.5.1 Parameterised Tail Fraction Approach

From Table 5.13, all 3 models have very similar model fit under all population distributions due to the overlapping bootstrap percentile intervals. Table 5.14 shows that a very similar conclusion can be drawn if the sample size increases to 5000. However, the GNG model has significant worse performances at 5% and 10% quantile of the Gumbel distribution.

### 5.5.2 Bulk Model Based Tail Fraction Approach

The GkerG model has significantly outperformed the GNG and CONGNG at 1%, 5%, 10%, 95%, 99% quantile of the Gumbel distribution from Table 5.15. The RMSE differences have reached the level of significance. Similarly, GNG and GkerG have better performances at 1%, 5%, 95%, 99%, quantile of the negative Weibull distribution. According to Table 5.16, there is further evidence that bulk model based approach is not as flexible as the parameterised approach. The results obtained from the GkerG model at 1%, 5%, 10%, 90%, 95%, 99% quantile of the Gumbel and negative Weibull distribution are clearly significant. Moreover, the GkerG and CONGNG model achieve significantly better than the other models at Student-t($v = 4, \mu = 1$) as well.

## 5.6 Likelihood Inference For The Kernel GPD And Boundary Corrected Kernel GPD Model

The kernel GPD/PP and boundary corrected kernel GPD/PP models are fitted using the Bayesian inference in Section 5.3 and Section 5.4 by MacDonald et al. (2011) and MacDonald et al. (2013). Bayesian inference using MCMC is always computationally demanding especially as part of simulation studies.

These two models can also be fitted using the likelihood inference. Hence, this section focuses on the performance of these models when using the likelihood inference. It is not clear that whether the point process (PP) or GPD model should be chosen in the literature. Bayesian inference should perform much better using the point process tail representation as the point process parameters are independent of the threshold, thus prior specification and sampling from posterior much easier. However, the advantage of the point process representation in likelihood inference is less clear. Thus the GPD model has been used when applying likelihood inference for these mixture models, which also makes the results comparable to those from the other mixture models.

It should be pointed out that, as the cross-validation likelihood is not consistent with heavy tail, care must be taken when the bulk distribution is also heavy tailed to ensure the bandwidth chosen is not too large. Special attentions must be paid to the convergence of the (cross-validation) likelihood function if the lower tail is also heavy. The likelihood inference based kernel GPD model takes about 30 seconds for a sample size of 1000 and about 10 minutes for a sample size of 5000. The likelihood inference based boundary corrected kernel GPD model takes about 90 seconds for a sample size of 1000 and 15 minutes for a sample size of 5000.

Table 5.17 illustrates the results of two models when the sample size of population distribution is 1000. By comparing Table 5.17, Table 5.1, Table 5.5 and Table 5.6, the results indicate that there is no major difference between Bayesian inference and likelihood inference in terms of a similar bootstrap percentile interval. This is a pleasing result as it further confirms the diffuse priors proposed by MacDonald et al. (2011) are not influencing the tail fit performance.

## 5.7 Main Result

1 In a parametric mixture model: if the bulk model is correct, then it will perform well and you should use the bulk model approach for the tail fraction. A slight advantage is also achieved if the density is constrained to be continuous at the threshold, but the difference appears to be small.

2 However, if the bulk model is mis-specified, then the performance of the parametric and semi-parametric mixture models is more mixed. However, in general in this situation, it is better to use the parameterised tail fraction approach as this provides more flexibility. So that even if the bulk fit is poor, the tail model can still perform well. There is little to be gained by using the continuity constraint at the threshold and in some situations it makes the fit worse so should be avoided.

3 The hybrid Pareto model frequently performs poorly, seemingly as it does not include the tail fraction. It was designed towards to work well for asymmetric heavy tailed distributions, however, even in this case the performance is mixed.

4 If the bulk model is correctly specified, then the parametric mixture models are quick, generally simple, easy to explain to users and computations are efficient, so they are preferred. However, in the more usual situation of a mis-specified bulk model (i.e. from unknown population distribution), the non-parametric mixture models (i.e. nonparametric kernel density estimator for the bulk) perform consistently well for tall tail quantiles.

5 The only situation identified where the non-parametric mixture models did not perform well is where the bulk distribution was also heavy tailed (i.e in lower tail), for which it is known the cross validation likelihood bandwidth estimator is biased. This is easily resolved by using a two tailed non-parametric extreme value mixture model, where both the upper and lower tails are described by a GPD. Another alternative proposed by MacDonald et al. (2013) is to pre-specify the bandwidth, as the bulk density estimate does not have a substantial impact on the tail fit in this model.

Table 5.1: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using parameterised tail fraction approach with sample size 1000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| Normal($\mu = 0, \sigma = 3$) | normGPD | 0.176 (0.154, 0.198) | 0.186 (0.159, 0.213) | **0.299 (0.259, 0.338)** | 0.753 (0.658, 0.843) |
| | CONnormGPD | 0.162 (0.138, 0.186) | 0.210 (0.180, 0.241) | 0.325 (0.285, 0.363) | 0.739 (0.640, 0.837) |
| | hPD | 1.285 (1.096, 1.483) | 1.441 (1.293, 1.598) | 1.243 (1.128, 1.359) | 0.758 (0.655, 0.859) |
| | shPD | 1.618 (1.339, 1.887) | 1.702 (1.475, 1.929) | 1.383 (1.231, 1.538) | 0.781 (0.646, 0.914) |
| | kernelGPD | **0.155 (0.136, 0.173)** | **0.168 (0.146, 0.188)** | 0.304 (0.260, 0.349) | **0.710 (0.624, 0.794)** |
| T($v = 3, \mu = 0$) | normGPD | 0.143 (0.123, 0.162) | 0.207 (0.178, 0.234) | 0.507 (0.427, 0.587) | 2.586 (2.209, 2.978) |
| | CONnormGPD | 0.114 (0.097, 0.131) | 0.203 (0.175, 0.232) | **0.411 (0.350, 0.472)** | **2.432 (2.223, 2.632)** |
| | hPD | 2.432 (2.297, 2.568) | 3.040 (2.851, 3.229) | 3.462 (3.113, 3.795) | 2.770 (2.418, 3.113) |
| | shPD | 2.413 (2.262, 2.569) | 3.032 (2.831, 3.241) | 3.505 (3.158, 3.866) | 2.657 (2.195, 3.160) |
| | kernelGPD | **0.089 (0.079, 0.100)** | **0.140 (0.122, 0.158)** | 0.529 (0.454, 0.604) | 2.539 (2.218, 2.870) |
| Gumbel($\sigma = 1$) | normGPD | 0.112 (0.098, 0.127) | 0.133 (0.117, 0.149) | 0.264 (0.232, 0.296) | **0.778 (0.669, 0.881)** |
| | CONnormGPD | 0.134 (0.118, 0.150) | 0.208 (0.185, 0.231) | 0.379 (0.293, 0.464) | 1.050 (0.815, 1.288) |
| | hPD | 0.222 (0.204, 0.239) | 0.276 (0.252, 0.300) | 0.274 (0.236, 0.311) | 0.867 (0.777, 0.955) |
| | shPD | 0.376 (0.206, 0.526) | 0.434 (0.266, 0.600) | 0.429 (0.278, 0.602) | 0.845 (0.751, 0.935) |
| | kernelGPD | **0.077 (0.067, 0.087)** | **0.118 (0.101, 0.134)** | **0.261 (0.219, 0.302)** | 0.863 (0.730, 0.996) |
| N.Weibull($l = 10, k = 5$) | normGPD | 0.150 (0.130, 0.169) | 0.163 (0.142, 0.183) | 0.239 (0.213, 0.264) | **0.423 (0.376, 0.469)** |
| | CONnormGPD | 0.126 (0.109, 0.143) | 0.161 (0.140, 0.182) | **0.227 (0.197, 0.256)** | 0.425 (0.369, 0.480) |
| | hPD | 0.549 (0.495, 0.606) | 0.630 (0.576, 0.687) | 0.503 (0.446, 0.561) | 0.439 (0.384, 0.493) |
| | shPD | 0.952 (0.712, 1.172) | 0.906 (0.719, 1.085) | 0.602 (0.500, 0.701) | 0.441 (0.393, 0.488) |
| | kernelGPD | **0.122 (0.105, 0.139)** | **0.147 (0.125, 0.169)** | 0.228 (0.193, 0.263) | 0.424 (0.379, 0.468) |
| T($v = 4, \mu = 1$) | normGPD | 0.127 (0.110, 0.143) | 0.194 (0.168, 0.219) | 0.485 (0.420, 0.550) | 2.461 (2.023, 2.988) |
| | CONnormGPD | 0.108 (0.093, 0.122) | 0.179 (0.157, 0.201) | **0.419 (0.357, 0.481)** | **1.945 (1.682, 2.243)** |
| | hPD | 0.979 (0.933, 1.027) | 1.202 (1.134, 1.272) | 1.062 (0.921, 1.200) | 2.451 (2.269, 2.627) |
| | shPD | 1.072 (0.882, 1.280) | 1.251 (1.074, 1.454) | 1.000 (0.837, 1.174) | 2.525 (2.337, 2.706) |
| | kernelGPD | **0.094 (0.080, 0.107)** | **0.143 (0.126, 0.160)** | 0.533 (0.457, 0.608) | 2.755 (2.095, 2.837) |

Table 5.2: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using parameterised tail fraction approach with sample size 5000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| Normal($\mu = 0, \sigma = 3$) | normGPD | 0.063 (0.054, 0.072) | **0.077 (0.065, 0.088)** | 0.150 (0.127, 0.172) | 0.360 (0.315, 0.405) |
| | CONnormGPD | 0.072 (0.063, 0.080) | 0.096 (0.083, 0.110) | 0.161 (0.140, 0.181) | 0.316 (0.271, 0.360) |
| | hPD | 1.742 (1.560, 1.926) | 2.000 (1.834, 2.170) | 1.933 (1.797, 2.076) | 1.056 (0.923, 1.188) |
| | shPD | 1.967 (1.688, 2.235) | 2.187 (1.933, 2.435) | 2.096 (1.894, 2.297) | 1.256 (1.084, 1.424) |
| | kernelGPD | **0.069 (0.059, 0.079)** | 0.083 (0.072, 0.095) | **0.141 (0.123, 0.159)** | **0.302 (0.263, 0.341)** |
| T(v=3,$\mu$=0) | normGPD | 0.107 (0.087, 0.127) | 0.146 (0.132, 0.159) | 0.218 (0.189, 0.247) | 1.801 (1.649, 1.947) |
| | CONnormGPD | 0.083 (0.076, 0.090) | 0.172 (0.160, 0.185) | **0.201 (0.174, 0.227)** | 2.179 (2.058, 2.296) |
| | hPD | 2.905 (2.804, 3.010) | 3.710 (3.588, 3.834) | 4.531 (4.331, 4.730) | 2.489 (2.128, 2.844) |
| | shPD | 2.817 (2.708, 2.934) | 3.592 (3.463, 3.725) | 4.428 (4.225, 4.628) | 2.631 (2.309, 2.936) |
| | kernelGPD | **0.038 (0.034, 0.043)** | **0.063 (0.054, 0.072)** | 0.211 (0.181, 0.240) | **1.049 (0.917, 1.174)** |
| Gumbel($\sigma = 1$) | normGPD | 0.052 (0.046, 0.059) | 0.079 (0.069, 0.089) | 0.139 (0.118, 0.160) | 0.336 (0.293, 0.379) |
| | CONnormGPD | 0.103 (0.093, 0.114) | 0.197 (0.182, 0.211) | 0.305 (0.265, 0.345) | 0.408 (0.356, 0.460) |
| | hPD | 0.248 (0.225, 0.281) | 0.317 (0.287, 0.361) | 0.269 (0.207, 0.351) | 0.584 (0.532, 0.637) |
| | shPD | 0.465 (0.257, 0.649) | 0.530 (0.324, 0.716) | 0.481 (0.280, 0.653) | 0.573 (0.518, 0.628) |
| | kernelGPD | **0.040 (0.035, 0.045)** | **0.057 (0.050, 0.064)** | **0.117 (0.101, 0.132)** | **0.325 (0.281, 0.367)** |
| N.Weibull($l = 10, k = 5$) | normGPD | 0.065 (0.055, 0.075) | 0.072 (0.059, 0.086) | 0.098 (0.079, 0.120) | 0.176 (0.152, 0.200) |
| | CONnormGPD | 0.065 (0.056, 0.074) | 0.113 (0.101, 0.125) | 0.179 (0.162, 0.196) | 0.170 (0.150, 0.190) |
| | hPD | 0.900 (0.838, 0.960) | 1.033 (0.966, 1.096) | 0.933 (0.865, 0.998) | 0.462 (0.401, 0.522) |
| | shPD | 1.465 (1.196, 1.714) | 1.438 (1.201, 1.657) | 1.105 (0.946, 1.256) | 0.514 (0.420, 0.605) |
| | kernelGPD | **0.056 (0.048, 0.064)** | **0.063 (0.054, 0.072)** | **0.085 (0.072, 0.098)** | **0.165 (0.144, 0.185)** |
| T($v = 4, \mu = 1$) | normGPD | 0.058 (0.048, 0.067) | 0.095 (0.080, 0.112) | **0.189 (0.167, 0.211)** | 1.127 (0.984, 1.267) |
| | CONnormGPD | 0.057 (0.046, 0.069) | 0.104 (0.087, 0.122) | 0.240 (0.202, 0.279) | 1.076 (0.930, 1.222) |
| | hPD | 1.079 (1.044, 1.124) | 1.375 (1.328, 1.429) | 1.382 (1.291, 1.476) | 1.473 (1.345, 1.595) |
| | shPD | 1.088 (0.998, 1.230) | 1.374 (1.269, 1.534) | 1.373 (1.212, 1.587) | 1.605 (1.468, 1.739) |
| | kernelGPD | **0.047 (0.040, 0.053)** | **0.070 (0.060, 0.081)** | 0.212 (0.189, 0.235) | **0.877 (0.773, 0.976)** |

Table 5.3: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using bulk model based tail fraction approach with sample size 1000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| Normal($\mu = 0, \sigma = 3$) | normGPD | 0.138 (0.118, 0.157) | 0.190 (0.165, 0.215) | 0.330 (0.279, 0.379) | 0.767 (0.673, 0.857) |
| | CONnormGPD | **0.128 (0.108, 0.148)** | **0.163 (0.138, 0.188)** | **0.282 (0.245, 0.319)** | **0.728 (0.639, 0.815)** |
| | hPD | 1.285 (1.096, 1.483) | 1.441 (1.293, 1.598) | 1.243 (1.128, 1.359) | 0.758 (0.655, 0.859) |
| | shPD | 1.618 (1.339, 1.887) | 1.702 (1.475, 1.929) | 1.383 (1.231, 1.538) | 0.781 (0.646, 0.914) |
| | kernelGPD | 0.154 (0.132, 0.176) | 0.179 (0.156, 0.201) | 0.323 (0.278, 0.367) | 0.852 (0.743, 0.959) |
| T($v = 3, \mu = 0$) | normGPD | **0.292 (0.263, 0.323)** | 0.417 (0.363, 0.473) | 0.766 (0.646, 0.887) | 2.828 (2.461, 3.201) |
| | CONnormGPD | 0.329 (0.285, 0.379) | 0.395 (0.341, 0.458) | 0.695 (0.615, 0.771) | 3.150 (2.540, 3.738) |
| | hPD | 2.432 (2.297, 2.568) | 3.040 (2.851, 3.229) | 3.462 (3.113, 3.795) | 2.770 (2.418, 3.113) |
| | shPD | 2.413 (2.262, 2.569) | 3.032 (2.831, 3.241) | 3.505 (3.158, 3.866) | **2.657 (2.195, 3.160)** |
| | kernelGPD | 0.296 (0.164, 0.441) | **0.338 (0.181, 0.511)** | **0.592 (0.475, 0.729)** | 3.009 (2.518, 3.485) |
| Gumbel($\sigma = 1$) | normGPD | 0.207 (0.183, 0.230) | 0.271 (0.233, 0.308) | 0.337 (0.298, 0.376) | 0.793 (0.696, 0.885) |
| | CONnormGPD | 0.121 (0.104, 0.138) | 0.176 (0.153, 0.199) | **0.273 (0.233, 0.310)** | **0.773 (0.681, 0.863)** |
| | hPD | 0.222 (0.204, 0.239) | 0.276 (0.252, 0.300) | 0.274 (0.236, 0.311) | 0.867 (0.777, 0.955) |
| | shPD | 0.376 (0.206, 0.526) | 0.434 (0.266, 0.600) | 0.429 (0.278, 0.602) | 0.845 (0.751, 0.935) |
| | kernelGPD | **0.096 (0.079, 0.117)** | **0.141 (0.116, 0.168)** | 0.294 (0.249, 0.337) | 1.165 (0.944, 1.381) |
| N.Weibull($l = 10, k = 5$) | normGPD | 0.182 (0.161, 0.202) | 0.185 (0.163, 0.208) | 0.222 (0.184, 0.261) | 0.412 (0.354, 0.469) |
| | CONnormGPD | 0.139 (0.124, 0.155) | 0.177 (0.158, 0.195) | **0.210 (0.182, 0.238)** | **0.379 (0.328, 0.429)** |
| | hPD | 0.549 (0.495, 0.606) | 0.630 (0.576, 0.687) | 0.503 (0.446, 0.561) | 0.439 (0.384, 0.493) |
| | shPD | 0.952 (0.712, 1.172) | 0.906 (0.719, 1.085) | 0.602 (0.500, 0.701) | 0.441 (0.393, 0.488) |
| | kernelGPD | **0.114 (0.098, 0.130)** | **0.142 (0.119, 0.167)** | 0.218 (0.185, 0.249) | 0.466 (0.401, 0.530) |
| T($v = 4, \mu = 1$) | normGPD | 0.199 (0.159, 0.246) | 0.305 (0.248, 0.366) | 0.498 (0.429, 0.564) | 2.263 (2.033, 2.487) |
| | CONnormGPD | 0.131 (0.107, 0.156) | 0.198 (0.169, 0.227) | **0.474 (0.414, 0.531)** | **1.962 (1.711, 2.207)** |
| | hPD | 0.979 (0.933, 1.027) | 1.202 (1.134, 1.272) | 1.062 (0.921, 1.200) | 2.451 (2.269, 2.627) |
| | shPD | 1.072 (0.882, 1.280) | 1.251 (1.074, 1.454) | 1.000 (0.837, 1.174) | 2.525 (2.337, 2.706) |
| | kernelGPD | **0.106 (0.089, 0.122)** | **0.145 (0.089, 0.123)** | 0.538 (0.428, 0.648) | 2.915 (2.292, 3.539) |

Table 5.4: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using bulk model based tail fraction approach with sample size 5000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| Normal($\mu = 0, \sigma = 3$) | normGPD | **0.059 (0.051, 0.068)** | **0.074 (0.063, 0.085)** | **0.115 (0.102, 0.127)** | **0.279 (0.234, 0.326)** |
| | CONnormGPD | 0.056 (0.048, 0.064) | 0.065 (0.056, 0.074) | 0.125 (0.108, 0.141) | 0.334 (0.283, 0.383) |
| | hPD | 1.742 (1.560, 1.926) | 2.000 (1.834, 2.170) | 1.933 (1.797, 2.076) | 1.056 (0.923, 1.188) |
| | shPD | 1.967 (1.688, 2.235) | 2.187 (1.933, 2.435) | 2.096 (1.894, 2.297) | 1.256 (1.084, 1.424) |
| | kernelGPD | 0.068 (0.059, 0.076) | 0.093 (0.083, 0.103) | 0.157 (0.135, 0.179) | 0.322 (0.282, 0.361) |
| T($v = 3, \mu = 0$) | normGPD | 0.343 (0.305, 0.382) | 0.474 (0.411, 0.539) | 0.529 (0.443, 0.621) | 2.082 (1.914, 2.241) |
| | CONnormGPD | 0.293 (0.275, 0.312) | 0.349 (0.326, 0.372) | 0.629 (0.567, 0.690) | 1.636 (1.352, 1.909) |
| | hPD | 2.905 (2.891, 3.117) | 3.736 (3.792, 4.127) | 4.625 (4.996, 5.784) | 2.778 (4.316, 6.438) |
| | shPD | 2.817 (2.708, 2.934) | 3.592 (3.463, 3.725) | 4.428 (4.225, 4.628) | 2.631 (2.309, 2.936) |
| | kernelGPD | **0.204 (0.157, 0.269)** | **0.228 (0.175, 0.302)** | **0.427 (0.371, 0.523)** | **1.401 (1.273, 1.660)** |
| Gumbel($\sigma = 1$) | normGPD | 0.216 (0.198, 0.234) | 0.304 (0.274, 0.332) | 0.302 (0.269, 0.335) | 0.491 (0.444, 0.538) |
| | CONnormGPD | 0.090 (0.081, 0.098) | 0.135 (0.124, 0.146) | 0.129 (0.112, 0.145) | 0.514 (0.467, 0.559) |
| | hPD | 0.248 (0.225, 0.281) | 0.317 (0.287, 0.361) | 0.269 (0.207, 0.351) | 0.584 (0.532, 0.637) |
| | shPD | 0.465 (0.257, 0.649) | 0.530 (0.324, 0.716) | 0.481 (0.280, 0.653) | 0.573 (0.518, 0.628) |
| | kernelGPD | **0.041 (0.036, 0.045)** | **0.057 (0.049, 0.064)** | **0.115 (0.100, 0.129)** | **0.331 (0.288, 0.373)** |
| N.Weibull($l = 10, k = 5$) | normGPD | 0.156 (0.145, 0.167) | 0.160 (0.145, 0.175) | 0.123 (0.108, 0.137) | **0.171 (0.152, 0.190)** |
| | CONnormGPD | 0.069 (0.059, 0.079) | 0.097 (0.085, 0.110) | 0.131 (0.112, 0.150) | 0.196 (0.173, 0.218) |
| | hPD | 0.900 (0.838, 0.960) | 1.033 (0.966, 1.096) | 0.933 (0.865, 0.998) | 0.462 (0.401, 0.522) |
| | shPD | 1.465 (1.196, 1.714) | 1.438 (1.201, 1.657) | 1.105 (0.946, 1.256) | 0.514 (0.420, 0.605) |
| | kernelGPD | **0.046 (0.041, 0.051)** | **0.059 (0.052, 0.066)** | **0.084 (0.073, 0.095)** | 0.174 (0.154, 0.193) |
| T($v = 4, \mu = 1$) | normGPD | 0.128 (0.115, 0.142) | 0.235 (0.200, 0.267) | 0.326 (0.275, 0.375) | 1.211 (1.073, 1.351) |
| | CONnormGPD | **0.066 (0.049, 0.086)** | 0.111 (0.098, 0.125) | **0.213 (0.186, 0.240)** | 1.228 (1.081, 1.375) |
| | hPD | 1.079 (1.044, 1.124) | 1.375 (1.328, 1.429) | 1.382 (1.291, 1.476) | 1.473 (1.345, 1.595) |
| | shPD | 1.088 (0.998, 1.230) | 1.374 (1.269, 1.534) | 1.373 (1.212, 1.587) | 1.605 (1.468, 1.739) |
| | kernelGPD | 0.070 (0.054, 0.085) | **0.089 (0.073, 0.105)** | 0.244 (0.200, 0.287) | **1.026 (0.840, 1.200)** |

Table 5.5: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using parameterised tail fraction approach with sample size 1000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| Gamma($\alpha = 5, \beta = 1$) | gammaGPD | 0.180 (0.156, 0.203) | **0.190 (0.163, 0.215)** | **0.342(0.301, 0.382)** | 0.974 (0.854, 1.095) |
| | CONgammaGPD | **0.155 (0.134, 0.175)** | 0.222 (0.191, 0.254) | 0.397 (0.347, 0.446) | 1.076 (0.956, 1.190) |
| | WeibullGPD | 0.182 (0.156, 0.208) | 0.222 (0.186, 0.261) | 0.352 (0.300, 0.402) | 1.080 (0.896, 1.266) |
| | CONWeibullGPD | 0.168 (0.145, 0.192) | 0.246 (0.212, 0.279) | 0.434 (0.380, 0.486) | **0.926 (0.815, 1.036)** |
| | gamMixGPD(m=2) | 0.166 (0.145, 0.186) | 0.203 (0.172, 0.236) | 0.412 (0.331, 0.503) | 1.061 (0.915, 1.205) |
| | DWM | 0.314 (0.244, 0.365) | 0.443 (0.378, 0.500) | 0.549 (0.478, 0.679) | 1.089 (1.004, 1.291) |
| | BCkernelGPD | 0.162 (0.143, 0.181) | 0.217 (0.188, 0.247) | 0.356 (0.303, 0.408) | 1.055 (0.907, 1.202) |
| Gamma($\alpha = 2, \beta = 2$) | gammaGPD | 0.235 (0.210, 0.260) | 0.304 (0.264, 0.339) | 0.499 (0.424, 0.574) | 1.743 (1.464, 2.037) |
| | CONgammaGPD | 0.229 (0.197, 0.260) | 0.324 (0.276, 0.371) | **0.484 (0.406, 0.571)** | 1.657 (1.444, 1.865) |
| | WeibullGPD | 0.234 (0.204, 0.265) | 0.329 (0.285, 0.374) | 0.589 (0.514, 0.663) | 1.799 (1.594, 1.997) |
| | CONWeibullGPD | 0.267 (0.227, 0.307) | 0.358 (0.294, 0.427) | 0.667 (0.569, 0.771) | 1.634 (1.446, 1.816) |
| | gamMixGPD(m=2) | 0.231 (0.188, 0.274) | **0.291 (0.238, 0.345)** | 0.543 (0.470, 0.613) | 1.642 (1.397, 1.883) |
| | DWM | 0.249 (0.213, 0.285) | 0.328 (0.283, 0.371) | 0.570 (0.498, 0.640) | **1.410 (1.241, 1.572)** |
| | BCkernelGPD | **0.204 (0.178, 0.229)** | 0.303 (0.266, 0.338) | 0.581 (0.504, 0.665) | 1.683 (1.413, 1.978) |
| Gamma($\alpha = 1, \beta = 2$) | gammaGPD | 0.199 (0.165, 0.234) | 0.258 (0.221, 0.294) | 0.548 (0.473, 0.622) | 1.631 (1.426, 1.826) |
| | CONgammaGPD | 0.171 (0.143, 0.197) | 0.252 (0.215, 0.287) | **0.472 (0.403, 0.538)** | 1.542 (1.362, 1.719) |
| | WeibullGPD | 0.211 (0.185, 0.236) | 0.251 (0.211, 0.290) | 0.481 (0.414, 0.547) | 1.544 (1.354, 1.726) |
| | CONWeibullGPD | 0.145 (0.127, 0.162) | 0.236 (0.204, 0.264) | 0.500 (0.429, 0.569) | 1.656 (1.407, 1.906) |
| | gamMixGPD(m=2) | 0.188 (0.164, 0.211) | 0.268 ( 0.231, 0.303) | 0.548 (0.481, 0.613) | 1.545 (1.368, 1.718) |
| | DWM | 0.166 (0.142, 0.190) | **0.235 (0.206, 0.265)** | 0.561 (0.500, 0.619) | 1.719 (1.515, 1.917) |
| | BCkernelGPD | **0.163 (0.141, 0.185)** | 0.246 (0.212, 0.279) | 0.558 (0.467, 0.654) | **1.435 (1.259, 1.607)** |
| N.Chi-Squared($v = 2, \lambda = 2$) | gammaGPD | 0.323 (0.279, 0.367) | 0.396 (0.352, 0.438) | 0.739 (0.626, 0.848) | 2.378 (2.006, 2.775) |
| | CONgammaGPD | 0.244 (0.205, 0.283) | 0.379 (0.326, 0.432) | 0.715 (0.635, 0.792) | 2.116 (1.825, 2.405) |
| | WeibullGPD | 0.322 (0.279, 0.364) | 0.444 (0.387, 0.502) | 0.804 (0.699, 0.905) | 2.248 (1.950, 2.537) |
| | CONWeibullGPD | 0.258 (0.222, 0.293) | 0.392 (0.344, 0.438) | 0.735 (0.641, 0.827) | 2.425 (2.063, 2.795) |
| | gamMixGPD(m=2) | 0.375 (0.321, 0.430) | 0.419 (0.356, 0.479) | 0.731 (0.612, 0.848) | 2.295 (1.946, 2.648) |
| | DWM | **0.240 (0.211, 0.269)** | **0.325 (0.286, 0.363)** | 0.739 (0.642, 0.835) | 2.074 (1.811, 2.332) |
| | BCkernelGPD | 0.282 (0.245, 0.316) | 0.389 (0.338, 0.439) | **0.684 (0.583, 0.786)** | **2.024 (1.716, 2.334)** |

Table 5.6: Continued:The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using parameterised tail fraction approach with sample size 1000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| N.Chi-Squared($v = 2, \lambda = 6$) | gammaGPD | 0.410 (0.346, 0.472) | **0.456 (0.399, 0.514)** | **0.815 (0.696, 0.934)** | 2.420 (2.138, 2.701) |
| | CONgammaGPD | **0.353 (0.301, 0.403)** | 0.559 (0.478, 0.638) | 0.874 (0.758, 0.985) | 2.361 (1.992, 2.734) |
| | WeibullGPD | 0.450 (0.391, 0.508) | 0.521 (0.447, 0.592) | 0.871 (0.758, 0.984) | 2.577 (2.227, 2.951) |
| | CONWeibullGPD | 0.359 (0.305, 0.412) | 0.523 (0.446, 0.597) | 0.867 (0.755, 0.975) | 2.391 (2.062, 2.709) |
| | gamMixGPD(m=2) | 0.453 (0.404, 0.502) | 0.489 (0.431, 0.546) | 1.074 (0.875, 1.217) | 2.439 (2.063, 2.804) |
| | DWM | 0.386 (0.330, 0.439) | 0.539 (0.463, 0.613) | 1.024 (0.887, 1.158) | 2.618 (2.366, 2.860) |
| | BCkernelGPD | 0.408 (0.349, 0.466) | 0.581 (0.490, 0.674) | 1.002 (0.860, 1.141) | **2.323 (2.040, 2.600)** |
| Inverse gamma($\alpha = 4, \beta = 4$) | gammaGPD | 0.093 (0.078, 0.108) | 0.143 (0.110, 0.181) | 0.456 (0.328, 0.616) | 2.032 (1.550, 2.645) |
| | CONgammaGPD | 0.124 (0.102, 0.149) | 0.245 (0.192, 0.303) | 0.560 (0.442, 0.683) | **1.630 (1.407, 1.856)** |
| | WeibullGPD | 0.102 (0.086, 0.118) | 0.142 (0.121, 0.163) | 0.367 (0.318, 0.414) | 1.685 (1.487, 1.896) |
| | CONWeibullGPD | 0.104 (0.093, 0.115) | 0.201 (0.177, 0.225) | 0.544 (0.445, 0.640) | 1.903 (1.572, 2.253) |
| | gamMixGPD(m=2) | 0.160 (0.138, 0.180) | 0.237 (0.201, 0.273) | 0.531 (0.444, 0.617) | 2.004 (1.706, 2.301) |
| | DWM | 0.156 (0.126, 0.192) | 0.281 (0.236, 0.336) | 0.429 (0.334, 0.531) | 2.006 (1.756, 2.286) |
| | BCkernelGPD | **0.079(0.067, 0.092)** | **0.125 (0.105, 0.145)** | **0.347 (0.282, 0.411)** | 1.779 (1.493, 2.060) |
| Inverse gamma($\alpha = 4, \beta = 8$) | GammaGPD | 0.171 (0.147, 0.194) | 0.232 (0.202, 0.260) | **0.658 (0.575, 0.736)** | 3.229 (2.859, 3.587) |
| | CONgammaGPD | 0.183 (0.160, 0.205) | 0.279 (0.243, 0.313) | 0.723 (0.627, 0.818) | **3.025 (2.691, 3.356)** |
| | WeibullGPD | 0.180 (0.158, 0.201) | 0.253 (0.221, 0.284) | 0.773 (0.661, 0.887) | 3.698 (3.220, 4.183) |
| | CONWeibullGPD | 0.194 (0.172, 0.215) | 0.298 (0.257, 0.337) | 0.673 (0.557, 0.788) | 3.193 (2.852, 3.533) |
| | gamMixGPD(m=2) | 0.171 (0.150, 0.191) | 0.293 (0.257, 0.328) | 0.752 (0.649, 0.853) | 3.520 (2.988, 4.058) |
| | DWM | 0.337 (0.288, 0.386) | 0.396 (0.338, 0.457) | 0.850 (0.744, 0.951) | 3.858 (3.499, 4.193) |
| | BCkernelGPD | **0.158 (0.134, 0.183)** | **0.249 (0.213, 0.284)** | 0.764 (0.661, 0.865) | 3.449 (3.015, 3.862) |

Table 5.7: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using parameterised tail fraction approach with sample size 5000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| Gamma($\alpha = 5, \beta = 1$) | gammaGPD | 0.072 (0.063, 0.082) | 0.091 (0.078, 0.104) | **0.155 (0.132, 0.177)** | 0.461 (0.397, 0.523) |
| | CONgammaGPD | 0.071 (0.060, 0.081) | 0.104 (0.088, 0.120) | 0.169 (0.146, 0.192) | 0.450 (0.388, 0.512) |
| | WeibullGPD | 0.076 (0.066, 0.086) | 0.096 (0.082, 0.111) | 0.169 (0.144, 0.194) | 0.373 (0.328, 0.416) |
| | CONWeibullGPD | 0.109 (0.097, 0.120) | 0.205 (0.188, 0.222) | 0.353 (0.322, 0.383) | 0.463 (0.403, 0.521) |
| | gamMixGPD(m=2) | **0.062 (0.053, 0.071)** | **0.081 (0.071, 0.090)** | 0.161 (0.139, 0.183) | **0.415 (0.359, 0.469)** |
| | DWM | 0.219 (0.198, 0.241) | 0.320 (0.290, 0.349) | 0.423 (0.376, 0.469) | 0.581 (0.514, 0.650) |
| Gamma($\alpha = 2, \beta = 2$) | gammaGPD | 0.117 (0.103, 0.132) | 0.152 (0.130, 0.173) | 0.282 (0.246, 0.317) | 0.773 (0.675, 0.869) |
| | CONgammaGPD | **0.098 (0.084, 0.112)** | 0.151 (0.128, 0.173) | 0.281 (0.244, 0.317) | 0.719 (0.625, 0.811) |
| | WeibullGPD | 0.114 (0.094, 0.134) | **0.124 (0.103, 0.143)** | 0.275 (0.239, 0.311) | 0.829 (0.743, 0.911) |
| | CONWeibullGPD | 0.120 (0.106, 0.134) | 0.207 (0.181, 0.233) | 0.382 (0.329, 0.434) | 0.764 (0.661, 0.865) |
| | gamMixGPD(m=2) | 0.108 (0.091, 0.124) | 0.140 (0.117, 0.163) | **0.267 (0.218, 0.317)** | **0.715 (0.618, 0.810)** |
| | DWM | 0.176 (0.158, 0.194) | 0.267 (0.2410, 0.2917) | 0.321 (0.272, 0.368) | 0.820 (0.710, 0.928) |
| Gamma($\alpha = 1, \beta = 2$) | gammaGPD | **0.071 (0.061, 0.080)** | **0.101 (0.087, 0.116)** | 0.250 (0.216, 0.282) | 0.683 (0.587, 0.776) |
| | CONgammaGPD | 0.083 (0.071, 0.094) | 0.125 (0.109, 0.141) | 0.233 (0.200, 0.266) | 0.673 (0.577, 0.768) |
| | WeibullGPD | 0.092 (0.081, 0.102) | 0.118 (0.101, 0.135) | 0.251 (0.222, 0.279) | 0.741 (0.629, 0.858) |
| | CONWeibullGPD | 0.082 (0.071, 0.093) | 0.126 (0.111, 0.140) | 0.231 (0.205, 0.255) | 0.682 (0.586, 0.775) |
| | gamMixGPD(m=2) | 0.089 (0.075, 0.103) | 0.111 (0.092, 0.129) | 0.221 (0.193, 0.250) | 0.725 (0.629, 0.821) |
| | DWM | 0.084 (0.071, 0.095) | 0.110 (0.093, 0.126) | **0.214 (0.181, 0.247)** | **0.650 (0.541, 0.757)** |
| | BCkernelGPD | 0.082 (0.073, 0.091) | 0.120 (0.106, 0.135) | 0.234 (0.200, 0.268) | 0.681 (0.584, 0.778) |
| N.Chi-Squared($v = 2, \lambda = 2$) | gammaGPD | 0.124 (0.107, 0.141) | 0.181 (0.153, 0.209) | 0.335 (0.288, 0.385) | **0.801 (0.692, 0.909)** |
| | CONgammaGPD | **0.115 (0.099, 0.130)** | 0.228 (0.201, 0.255) | 0.398 (0.350, 0.445) | 0.908 (0.774, 1.041) |
| | WeibullGPD | 0.130 (0.111, 0.147) | 0.172 (0.146, 0.198) | 0.347 (0.290, 0.403) | 1.034 (0.868, 1.197) |
| | CONWeibullGPD | 0.136 (0.117, 0.155) | 0.209 (0.178, 0.238) | 0.344 (0.294, 0.393) | 0.864 (0.749, 0.979) |
| | gamMixGPD(m=2) | 0.240 (0.211, 0.269) | 0.229 (0.195, 0.263) | 0.395 (0.333, 0.456) | 1.418 (1.226, 1.608) |
| | DWM | 0.123 (0.106, 0.139) | **0.168 (0.145, 0.190)** | **0.306 (0.268, 0.344)** | 0.905 (0.796, 1.011) |
| | BCkernelGPD | 0.127 (0.111, 0.142) | 0.175 (0.148, 0.203) | 0.312 (0.266, 0.354) | 0.813 (0.721, 0.902) |

Table 5.8: Continued:The RMSE of quantile estimate and 95% bootstrap percentile interval using parameterised tail fraction approach with sample size 5000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| N.Chi-Squared($v = 2, \lambda = 6$) | GammaGPD | 0.196 (0.172, 0.220) | 0.253 (0.221, 0.283) | 0.444 (0.388, 0.501) | 1.177 (1.053, 1.297) |
| | CONgammaGPD | 0.174 (0.146, 0.202) | 0.388 (0.346, 0.429) | 0.556 (0.485, 0.626) | 1.259 (1.107, 1.406) |
| | WeibullGPD | 0.181 (0.154, 0.206) | 0.244 (0.203, 0.289) | **0.382 (0.328, 0.437)** | **1.010 (0.883, 1.133)** |
| | CONWeibullGPD | 0.175 (0.150, 0.198) | 0.261 (0.221, 0.299) | 0.493 (0.427, 0.558) | 1.357 (1.169, 1.543) |
| | gamMixGPD(m=2) | 0.340 (0.279, 0.398) | 0.346 (0.275, 0.419) | 0.541 (0.458, 0.623) | 1.194 (1.023, 1.367) |
| | DWM | **0.158 (0.136, 0.179)** | **0.203 (0.178, 0.228)** | 0.417 (0.368, 0.464) | 1.490 (1.266, 1.726) |
| Inverse gamma($\alpha = 4, \beta = 4$) | gammaGPD | **0.056 (0.048, 0.063)** | 0.078 (0.066, 0.090) | 0.222 (0.164, 0.295) | 1.171 (1.057, 1.285) |
| | CONgammaGPD | 0.074 (0.066, 0.083) | 0.112 (0.098, 0.126) | 0.202 (0.175, 0.228) | 1.166 (1.051, 1.279) |
| | WeibullGPD | 0.057 (0.050, 0.063) | **0.077 (0.069, 0.087)** | **0.181 (0.158, 0.204)** | 1.155 (1.035, 1.271) |
| | CONWeibullGPD | 0.088 (0.083, 0.093) | 0.124 (0.109, 0.139) | 0.230 (0.179, 0.284) | 1.433 (1.335, 1.529) |
| | gamMixGPD(m=2) | 0.058 (0.046, 0.070) | 0.084 (0.068, 0.101) | 0.233 (0.187, 0.284) | **1.107 (0.932, 1.283)** |
| | DWM | 0.142 (0.126, 0.158) | 0.273 (0.248, 0.299) | 0.294 (0.240, 0.350) | 1.298 (1.146, 1.446) |
| Inverse gamma($\alpha = 4, \beta = 8$) | GammaGPD | 0.077 (0.067, 0.088) | 0.120 (0.103, 0.135) | 0.365 (0.316, 0.414) | 1.767 (1.548, 1.974) |
| | CONgammaGPD | 0.161 (0.148, 0.174) | 0.241 (0.215, 0.266) | 0.431 (0.356, 0.512) | 2.242 (2.035, 2.443) |
| | WeibullGPD | 0.119 (0.103, 0.136) | 0.155 (0.135, 0.175) | 0.358 (0.308, 0.407) | 2.370 (2.115, 2.620) |
| | CONWeibullGPD | 0.181 (0.165, 0.197) | 0.254 (0.221, 0.290) | 0.421 (0.347, 0.497) | 2.835 (2.648, 3.021) |
| | gamMixGPD(m=2) | 0.069 (0.060, 0.078) | 0.125 (0.106, 0.144) | 0.551 (0.449, 0.649) | 2.119 (1.801, 2.433) |
| | DWM | 0.263 (0.235, 0.290) | 0.286 (0.253, 0.320) | 0.515 (0.444, 0.587) | 2.814 (2.500, 3.115) |
| | BCkernelGPD | **0.065 (0.058, 0.072)** | **0.113 (0.098, 0.129)** | **0.335 (0.295, 0.374)** | **1.522 (1.341, 1.704)** |

Table 5.9: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using bulk model based tail fraction approach with sample size 1000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| Gamma($\alpha = 5, \beta = 1$) | GammaGPD | 0.150 (0.131, 0.168) | 0.206 (0.177, 0.235) | 0.404 (0.345, 0.463) | 1.005 (0.881, 1.124) |
| | CONgammaGPD | 0.142 (0.123, 0.161) | 0.178 (0.153, 0.202) | 0.325 (0.281, 0.367) | 1.140 (0.986, 1.295) |
| | WeibullGPD | 0.239 (0.210, 0.267) | 0.302 (0.259, 0.344) | 0.443 (0.397, 0.488) | 1.090 (0.962, 1.216) |
| | CONWeibullGPD | 0.165 (0.142, 0.188) | 0.229 (0.201, 0.256) | 0.370 (0.313, 0.427) | 0.968 (0.836, 1.099) |
| Gamma($\alpha = 2, \beta = 2$) | GammaGPD | 0.200 (0.171, 0.228) | 0.320 (0.272, 0.368) | 0.654 (0.584, 0.723) | 1.719 (1.512, 1.929) |
| | CONgammaGPD | 0.218 (0.192, 0.243) | 0.269 (0.235, 0.302) | 0.593 (0.511, 0.674) | 2.058 (1.823, 2.283) |
| | WeibullGPD | 0.287 (0.251, 0.322) | 0.336 (0.291, 0.379) | 0.593 (0.522, 0.662) | 1.706 (1.532, 1.876) |
| | CONWeibullGPD | 0.227 (0.196, 0.256) | 0.326 (0.282, 0.370) | 0.547 (0.468, 0.623) | 1.772 (1.431, 2.113) |
| Gamma($\alpha = 1, \beta = 2$) | GammaGPD | 0.187 (0.163, 0.211) | 0.248 (0.214, 0.281) | 0.631 (0.541, 0.719) | 1.562 (1.364, 1.752) |
| | CONgammaGPD | 0.171 (0.150, 0.192) | 0.217 (0.190, 0.245) | 0.411 (0.351, 0.470) | 1.666 (1.449, 1.881) |
| | WeibullGPD | 0.167 (0.144, 0.189) | 0.228 (0.202, 0.253) | 0.515 (0.448, 0.581) | 1.516 (1.353, 1.673) |
| | CONWeibullGPD | 0.169 (0.144, 0.195) | 0.243 (0.206, 0.280) | 0.558 (0.486, 0.628) | 1.883 (1.676, 2.088) |
| N.Chi-Squared($v = 2, \lambda = 2$) | GammaGPD | 0.374 (0.330, 0.418) | 0.457 (0.378, 0.535) | 0.741 (0.626, 0.864) | 2.433 (2.012, 2.854) |
| | CONgammaGPD | 0.360 (0.307, 0.414) | 0.517 (0.451, 0.580) | 0.779 (0.663, 0.895) | 2.069 (1.809, 2.323) |
| | WeibullGPD | 0.301 (0.263, 0.340) | 0.443 (0.376, 0.508) | 0.817 (0.712, 0.919) | 2.255 (1.986, 2.520) |
| | CONWeibullGPD | 0.291 (0.252, 0.330) | 0.410 (0.353, 0.470) | 0.765 (0.668, 0.860) | 2.238 (1.965, 2.500) |

Table 5.10: Continued:The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using bulk model based tail fraction approach with sample size 1000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| N.Chi-Squared($v = 2, \lambda = 6$) | GammaGPD | 0.717 (0.653, 0.780) | 0.739 (0.653, 0.827) | 0.986 (0.856, 1.112) | 2.565 (2.268, 2.853) |
| | CONgammaGPD | 0.700 (0.631, 0.769) | 0.995 (0.909, 1.076) | 1.353 (1.174, 1.532) | 2.947 (2.268, 2.853) |
| | WeibullGPD | 0.404 (0.351, 0.455) | 0.546 (0.474, 0.617) | 1.067 (0.918, 1.209) | 2.717 (2.373, 3.049) |
| | CONWeibullGPD | 0.377 (0.323, 0.430) | 0.477 (0.409, 0.546) | 0.834 (0.715, 0.949) | 2.585 (2.230, 2.917) |
| Inverse gamma($\alpha = 4, \beta = 4$) | gammaGPD | 0.161 (0.142, 0.180) | 0.289 (0.243, 0.333) | 0.570 (0.470, 0.663) | 1.894 (1.655, 2.124) |
| | CONgammaGPD | 0.082 (0.071, 0.094) | 0.122 (0.104, 0.141) | 0.372 (0.310, 0.437) | 1.787 (1.486, 2.149) |
| | WeibullGPD | 0.167 (0.151, 0.184) | 0.301 (0.265, 0.336) | 0.526 (0.459, 0.592) | 1.789 (1.624, 1.950) |
| | CONWeibullGPD | 0.100 (0.086, 0.116) | 0.139 (0.118, 0.160) | 0.366 (0.324, 0.406) | 1.806 (1.654, 1.953) |
| Inverse gamma($\alpha = 4, \beta = 8$) | GammaGPD | 0.216 (0.185, 0.247) | 0.303 (0.253, 0.353) | 0.728 (0.622, 0.837) | 3.198 (2.851, 3.533) |
| | CONgammaGPD | 0.172 (0.149, 0.195) | 0.238 (0.201, 0.273) | 0.670 (0.582, 0.760) | 3.306 (2.979, 3.624) |
| | WeibullGPD | 0.228 (0.192, 0.263) | 0.377 (0.307, 0.4438) | 0.834 (0.711, 0.959) | 3.420 (2.975, 3.864) |
| | CONWeibullGPD | 0.220 (0.194, 0.245) | 0.303 (0.264, 0.340) | 0.696 (0.618, 0.775) | 3.433 (3.114, 3.739) |

Table 5.11: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using bulk model based tail fraction approach with sample size 5000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| Gamma($\alpha = 5, \beta = 1$) | GammaGPD | 0.061 (0.052, 0.070) | 0.092 (0.079, 0.106) | 0.164 (0.146, 0.183) | 0.508 (0.444, 0.571) |
| | CONgammaGPD | 0.062 (0.055, 0.069) | 0.077 (0.067, 0.088) | 0.130 (0.111, 0.150) | 0.450 (0.393, 0.504) |
| | WeibullGPD | 0.218 (0.204, 0.232) | 0.262 (0.240, 0.282) | 0.255 (0.240, 0.283) | 0.505 (0.447, 0.561) |
| | CONWeibullGPD | 0.103 (0.089, 0.118) | 0.177 (0.158, 0.196) | 0.225 (0.201, 0.248) | 0.561 (0.491, 0.629) |
| Gamma($\alpha = 2, \beta = 2$) | GammaGPD | 0.081 (0.072, 0.090) | 0.132 (0.115, 0.149) | 0.277 (0.242, 0.311) | 0.855 (0.760, 0.946) |
| | CONgammaGPD | 0.084 (0.074, 0.094) | 0.106 (0.094, 0.119) | 0.221 (0.191, 0.250) | 0.873 (0.760, 0.982) |
| | WeibullGPD | 0.214 (0.194, 0.234) | 0.225 (0.198, 0.252) | 0.302 (0.264, 0.337) | 0.835 (0.723, 0.944) |
| | CONWeibullGPD | 0.113 (0.099, 0.127) | 0.161 (0.140, 0.182) | 0.226 (0.189, 0.261) | 0.812 (0.716, 0.906) |
| Gamma($\alpha = 1, \beta = 2$) | GammaGPD | 0.075 (0.065, 0.086) | 0.106 (0.091, 0.122) | 0.258 (0.217, 0.299) | 1.045 (0.720, 1.439) |
| | CONgammaGPD | 0.086 (0.063, 0.116) | 0.116 (0.084, 0.157) | 0.225 (0.157, 0.310) | 0.790 (0.642, 0.946) |
| | WeibullGPD | 0.082 (0.071, 0.094) | 0.120 (0.102, 0.137) | 0.232 (0.201, 0.264) | 0.724 (0.625, 0.822) |
| | CONWeibullGPD | 0.078 (0.067, 0.089) | 0.108 (0.093, 0.124) | 0.236 (0.209, 0.263) | 0.848 (0.737, 0.962) |
| N.Chi-Squared($v = 2, \lambda = 2$) | GammaGPD | 0.309 (0.289, 0.327) | 0.284 (0.260, 0.308) | 0.378 (0.333, 0.422) | 1.104 (0.969, 1.238) |
| | CONgammaGPD | 0.235 (0.214, 0.255) | 0.334 (0.304, 0.363) | 0.398 (0.344, 0.449) | 1.197 (1.075, 1.318) |
| | WeibullGPD | 0.183 (0.163, 0.203) | 0.206 (0.179, 0.231) | 0.378 (0.329, 0.428) | 0.960 (0.830, 1.087) |
| | CONWeibullGPD | 0.160 (0.139, 0.180) | 0.237 (0.209, 0.264) | 0.393 (0.337, 0.446) | 1.256 (1.084, 1.430) |

Table 5.12: Continued:The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using bulk model based tail fraction approach with sample size 5000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| N.Chi-Squared($v = 2, \lambda = 6$) | GammaGPD | 0.638 (0.606, 0.670) | 0.612 (0.565, 0.659) | 0.702 (0.621, 0.783) | 1.073 (0.935, 1.208) |
| | CONgammaGPD | 0.604 (0.573, 0.634) | 0.916 (0.873, 0.958) | 1.092 (0.986, 1.196) | 1.366 (1.198, 1.531) |
| | WeibullGPD | 0.177 (0.153, 0.200) | 0.236 (0.204, 0.268) | 0.455 (0.390, 0.522) | 0.991 (0.859, 1.119) |
| | CONWeibullGPD | 0.157 (0.140, 0.176) | 0.221 (0.196, 0.245) | 0.418 (0.369, 0.468) | 1.121 (0.980, 1.256) |
| Inverse gamma($\alpha = 4, \beta = 4$) | gammaGPD | 0.140 (0.122, 0.159) | 0.236 (0.196, 0.273) | 0.397 (0.323, 0.470) | 1.831 (1.140, 2.629) |
| | CONgammaGPD | 0.063 (0.057, 0.069) | 0.083 (0.074, 0.091) | 0.183 (0.160, 0.205) | 1.336 (1.238, 1.430) |
| | WeibullGPD | 0.136 (0.124, 0.149) | 0.219 (0.189, 0.248) | 0.357 (0.311, 0.402) | 1.381 (1.271, 1.486) |
| | CONWeibullGPD | 0.084 (0.078, 0.090) | 0.096 (0.088, 0.104) | 0.186 (0.165, 0.208) | 1.601 (1.520, 1.682) |
| Inverse gamma($\alpha = 4, \beta = 8$) | GammaGPD | 0.243 (0.214, 0.271) | 0.369 (0.306, 0.428) | 0.654 (0.529, 0.776) | 2.136 (1.944, 2.321) |
| | CONgammaGPD | 0.110 (0.099, 0.121) | 0.143 (0.126, 0.159) | 0.371 (0.324, 0.417) | 2.701 (2.519, 2.877) |
| | WeibullGPD | 0.253 (0.227, 0.278) | 0.450 (0.382, 0.515) | 0.726 (0.617, 0.834) | 2.527 (2.281, 2.764) |
| | CONWeibullGPD | 0.164 (0.151, 0.177) | 0.191 (0.171, 0.211) | 0.407 (0.358, 0.454) | 3.217 (3.046, 3.388) |

Table 5.13: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using parameterised tail fraction approach with sample size 1000

| Distribution | Model | 0.1% | 1% | 5% | 10% | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|---|---|---|---|
| | GNG | 0.759 | 0.325 | 0.202 | 0.155 | 0.152 | 0.185 | 0.306 | 0.740 |
| Normal($\mu = 0, \sigma = 3$) | | (0.656, 0.862) | (0.279, 0.372) | (0.173, 0.231) | (0.134, 0.175) | (0.131, 0.171) | (0.164, 0.205) | (0.266, 0.345) | (0.645, 0.833) |
| | CONGNG | **0.651** | **0.278** | **0.182** | **0.130** | **0.149** | 0.197 | 0.298 | 0.754 |
| | | **(0.575, 0.728)** | **(0.239, 0.317)** | **(0.153, 0.209)** | **(0.113, 0.148)** | **(0.127, 0.171)** | (0.169, 0.225) | (0.259, 0.337) | (0.614, 0.912) |
| | GkerG | 0.752 | 0.346 | 0.198 | 0.157 | 0.172 | **0.176** | **0.272** | **0.739** |
| | | (0.656, 0.845) | (0.307, 0.385) | (0.174, 0.220) | (0.139, 0.175) | (0.146, 0.196) | **(0.150, 0.202)** | **(0.236, 0.308)** | **(0.642, 0.843)** |
| | GNG | 2.347 | 0.526 | 0.237 | 0.098 | **0.097** | 0.190 | 0.524 | **2.328** |
| T($v = 3, \mu = 0$) | | (1.978, 2.729) | (0.427, 0.625) | (0.169, 0.314) | (0.083, 0.113) | **(0.084, 0.110)** | (0.160, 0.222) | (0.424, 0.627) | **(2.068, 2.581)** |
| | CONGNG | **2.284** | **0.445** | 0.205 | 0.324 | 0.309 | 0.184 | **0.412** | 2.395 |
| | | **(1.940, 2.651)** | **(0.366, 0.521)** | (0.154, 0.255) | (0.291, 0.354) | (0.276, 0.340) | (0.142, 0.227) | **(0.357, 0.467)** | (1.930, 2.886) |
| | GkerG | 2.424 | 0.505 | **0.171** | **0.097** | 0.105 | **0.164** | 0.470 | 2.430 |
| | | (2.002, 2.864) | (0.409, 0.609) | **(0.141, 0.202)** | **(0.083, 0.115)** | (0.091, 0.120) | **(0.139, 0.188)** | (0.408, 0.533) | (2.083, 2.775) |
| | GNG | 0.180 | 0.074 | 0.056 | 0.044 | 0.099 | **0.115** | 0.246 | 0.836 |
| Gumbel($\sigma = 1$) | | (0.127, 0.233) | (0.065, 0.083) | (0.048, 0.064) | (0.038, 0.050) | (0.084, 0.114) | **(0.098, 0.131)** | (0.214, 0.278) | (0.710, 0.963) |
| | CONGNG | **0.131** | **0.065** | 0.051 | 0.046 | **0.098** | 0.143 | 0.308 | 0.845 |
| | | **(0.116, 0.147)** | **(0.056, 0.075)** | (0.044, 0.058) | (0.040, 0.052) | **(0.084, 0.111)** | (0.123, 0.164) | (0.259, 0.356) | (0.707, 0.980) |
| | GkerG | 0.165 | 0.080 | **0.048** | **0.036** | 0.100 | 0.129 | **0.240** | **0.826** |
| | | (0.117, 0.217) | (0.067, 0.093) | **(0.042, 0.054)** | **(0.030, 0.042)** | (0.088, 0.112) | (0.113, 0.144) | **(0.207, 0.274)** | **(0.667, 1.008)** |
| | GNG | 0.374 | **0.155** | **0.105** | 0.098 | 0.122 | 0.144 | 0.204 | 0.420 |
| N.Weibull($l = 10, k = 5$) | | (0.323, 0.424) | **(0.133, 0.178)** | **(0.089, 0.121)** | (0.084, 0.112) | (0.107, 0.137) | (0.126, 0.162) | (0.175, 0.233) | (0.363, 0.476) |
| | CONGNG | **0.363** | 0.158 | 0.118 | **0.094** | **0.108** | 0.148 | 0.210 | 0.446 |
| | | **(0.316, 0.410)** | (0.136, 0.180) | (0.100, 0.138) | **(0.079, 0.112)** | **(0.090, 0.125)** | (0.124, 0.174) | (0.180, 0.240) | (0.358, 0.545) |
| | GkerG | 0.366 | 0.157 | 0.111 | 0.095 | 0.123 | **0.121** | **0.195** | **0.407** |
| | | (0.308, 0.423) | (0.137, 0.178) | (0.096, 0.126) | (0.079, 0.111) | (0.105, 0.141) | **(0.101, 0.140)** | **(0.162, 0.233)** | **(0.352, 0.461)** |
| | GNG | 0.795 | 0.226 | 0.104 | 0.071 | **0.114** | 0.190 | 0.507 | 2.149 |
| T($v = 4, \mu = 1$) | | (0.674, 0.940) | (0.186, 0.266) | (0.088, 0.120) | (0.057, 0.087) | **(0.100, 0.130)** | (0.165, 0.215) | (0.446, 0.565) | (1.873, 2.443) |
| | CONGNG | **0.749** | **0.188** | **0.082** | **0.058** | 0.115 | 0.196 | **0.502** | **2.115** |
| | | **(0.672, 0.822)** | **(0.166, 0.209)** | **(0.071, 0.092)** | **(0.050, 0.065)** | (0.099, 0.132) | (0.170, 0.221) | **(0.437, 0.566)** | **(1.846, 2.380)** |
| | GkerG | 0.823 | 0.231 | 0.104 | 0.071 | 0.121 | **0.176** | 0.556 | 2.394 |
| | | (0.722, 0.922) | (0.197, 0.265) | (0.088, 0.121) | (0.060, 0.082) | (0.105, 0.137) | **(0.149, 0.204)** | (0.455, 0.665) | (2.046, 2.737) |

Table 5.14: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using parameterised tail fraction approach with sample size 5000

| Distribution | Model | 0.1% | 1% | 5% | 10% | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|---|---|---|---|
| | GNG | 0.324 | 0.151 | **0.081** | 0.069 | 0.072 | **0.087** | 0.170 | 0.346 |
| Normal($\mu = 0, \sigma = 3$) | | (0.279, 0.369) | (0.126, 0.177) | **(0.069, 0.092)** | (0.060, 0.078) | (0.058, 0.087) | **(0.075, 0.100)** | (0.143, 0.199) | (0.301, 0.390) |
| | CONGNG | **0.319** | **0.140** | 0.085 | **0.063** | 0.063 | 0.089 | **0.140** | **0.278** |
| | | **(0.273, 0.363)** | **(0.119, 0.161)** | (0.072, 0.097) | (0.054, 0.071) | (0.052, 0.075) | (0.074, 0.103) | **(0.119, 0.160)** | **(0.238, 0.319)** |
| | GkerG | 0.352 | 0.170 | 0.097 | 0.071 | **0.065** | 0.095 | 0.166 | 0.334 |
| | | (0.299, 0.405) | (0.138, 0.203) | (0.079, 0.115) | (0.058, 0.083) | **(0.058, 0.073)** | (0.081, 0.109) | (0.145, 0.188) | (0.279, 0.388) |
| | GNG | 1.289 | **0.188** | 0.084 | 0.046 | 0.039 | 0.082 | 0.223 | 1.270 |
| T($v = 3, \mu = 0$) | | (1.174, 1.400) | **(0.157, 0.221)** | (0.070, 0.099) | (0.040, 0.053) | (0.034, 0.044) | (0.070, 0.094) | (0.187, 0.259) | (1.054, 1.527) |
| | CONGNG | **1.113** | 0.284 | 0.082 | **0.042** | 0.042 | 0.087 | 0.294 | **1.181** |
| | | **(0.987, 1.238)** | (0.234, 0.337) | (0.070, 0.095) | **(0.036, 0.048)** | (0.036, 0.048) | (0.074, 0.100) | (0.247, 0.342) | **(1.016, 1.344)** |
| | GkerG | 1.380 | 0.217 | **0.069** | 0.043 | **0.038** | **0.061** | **0.204** | 1.378 |
| | | (1.164, 1.594) | (0.182, 0.251) | **(0.057, 0.082)** | (0.037, 0.049) | **(0.033, 0.043)** | **(0.050, 0.071)** | **(0.170, 0.238)** | (1.181, 1.581) |
| | GNG | 0.124 | 0.060 | 0.061 | 0.065 | 0.049 | **0.058** | 0.120 | 0.372 |
| Gumbel($\sigma = 1$) | | (0.056, 0.196) | (0.039, 0.084) | (0.032, 0.089) | (0.035, 0.092) | (0.042, 0.056) | **(0.050, 0.066)** | (0.103, 0.138) | (0.320, 0.425) |
| | CONGNG | **0.052** | **0.033** | 0.021 | **0.019** | **0.048** | 0.067 | 0.127 | **0.336** |
| | | **(0.044, 0.059)** | **(0.029, 0.037)** | (0.019, 0.024) | **(0.017, 0.022)** | **(0.041, 0.055)** | (0.058, 0.076) | (0.110, 0.144) | **(0.277, 0.402)** |
| | GkerG | 0.057 | 0.037 | **0.020** | 0.020 | 0.050 | 0.063 | **0.114** | 0.382 |
| | | (0.048, 0.066) | (0.032, 0.042) | **(0.018, 0.023)** | (0.017, 0.023) | (0.044, 0.055) | (0.055, 0.070) | **(0.099, 0.129)** | (0.326, 0.440) |
| | GNG | 0.157 | 0.090 | 0.054 | 0.047 | 0.066 | 0.068 | 0.088 | **0.182** |
| N.Weibull($l = 10, k = 5$) | | (0.137, 0.177) | (0.073, 0.108) | (0.047, 0.061) | (0.040, 0.054) | (0.057, 0.076) | (0.058, 0.078) | (0.074, 0.103) | **(0.154, 0.211)** |
| | CONGNG | **0.149** | **0.069** | 0.049 | 0.045 | **0.048** | 0.072 | 0.107 | 0.184 |
| | | **(0.129, 0.169)** | **(0.058, 0.081)** | (0.043, 0.055) | (0.039, 0.050) | **(0.042, 0.055)** | (0.064, 0.081) | (0.094, 0.119) | (0.160, 0.206) |
| | GkerG | 0.158 | 0.076 | **0.048** | **0.044** | 0.054 | **0.062** | **0.087** | 0.216 |
| | | (0.138, 0.178) | (0.066, 0.085) | **(0.043, 0.054)** | **(0.038, 0.050)** | (0.046, 0.061) | **(0.054, 0.070)** | **(0.074, 0.101)** | (0.183, 0.249) |
| | GNG | 0.374 | 0.103 | 0.036 | 0.034 | 0.060 | 0.089 | **0.184** | 1.026 |
| T($v = 4, \mu = 1$) | | (0.330, 0.417) | (0.087, 0.120) | (0.032, 0.041) | (0.030, 0.039) | (0.052, 0.068) | (0.074, 0.106) | **(0.161, 0.205)** | (0.841, 1.254) |
| | CONGNG | **0.340** | 0.120 | 0.040 | 0.028 | 0.056 | 0.107 | 0.254 | **0.931** |
| | | **(0.293, 0.386)** | (0.104, 0.135) | (0.034, 0.046) | (0.024, 0.032) | (0.049, 0.064) | (0.092, 0.121) | (0.223, 0.285) | **(0.818, 1.037)** |
| | GkerG | 0.393 | **0.095** | **0.034** | **0.027** | **0.045** | **0.074** | 0.205 | 1.310 |
| | | (0.343, 0.441) | **(0.080, 0.110)** | **(0.030, 0.039)** | **(0.023, 0.031)** | **(0.039, 0.052)** | **(0.064, 0.085)** | (0.178, 0.232) | (1.062, 1.588) |

Table 5.15: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using bulk model based tail fraction approach with sample size 1000

| Distribution | Model | 0.1% | 1% | 5% | 10% | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|---|---|---|---|
| | GNG | **0.674** | 0.322 | 0.190 | **0.137** | **0.131** | 0.188 | 0.298 | **0.698** |
| Normal($\mu = 0, \sigma = 3$) | | **(0.586, 0.760)** | (0.283, 0.361) | (0.164, 0.215) | **(0.117, 0.156)** | **(0.112, 0.150)** | (0.164, 0.213) | (0.260, 0.334) | **(0.608, 0.788)** |
| | CONGNG | 0.742 | **0.254** | **0.167** | 0.145 | 0.144 | **0.164** | **0.266** | 0.820 |
| | | (0.652, 0.831) | **(0.218, 0.290)** | **(0.146, 0.188)** | (0.128, 0.161) | (0.127, 0.160) | **(0.144, 0.183)** | **(0.228, 0.303)** | (0.712, 0.935) |
| | GkerG | 0.752 | 0.346 | 0.198 | 0.157 | 0.172 | 0.176 | 0.272 | 0.739 |
| | | (0.656, 0.845) | (0.307, 0.385) | (0.174, 0.220) | (0.139, 0.175) | (0.146, 0.196) | (0.150, 0.202) | (0.236, 0.308) | (0.642, 0.843) |
| | GNG | **2.161** | **0.392** | 0.164 | 0.107 | 0.102 | **0.162** | **0.448** | 2.242 |
| T($v = 3, \mu = 0$) | | **(1.944, 2.371)** | **(0.338, 0.446)** | (0.142, 0.185) | (0.094, 0.119) | (0.090, 0.114) | **(0.138, 0.185)** | **(0.382, 0.516)** | (1.937, 2.591) |
| | CONGNG | 2.232 | 0.493 | **0.163** | **0.089** | **0.090** | 0.155 | 0.453 | 2.441 |
| | | (1.945, 2.514) | (0.426, 0.559) | **(0.141, 0.185)** | **(0.078, 0.101)** | **(0.079, 0.100)** | (0.139, 0.170) | (0.394, 0.509) | (2.104, 2.808) |
| | GkerG | 2.424 | 0.505 | 0.171 | 0.097 | 0.105 | 0.164 | 0.470 | **2.430** |
| | | (2.002, 2.864) | (0.409, 0.609) | (0.141, 0.202) | (0.083, 0.115) | (0.091, 0.120) | (0.139, 0.188) | (0.408, 0.533) | **(2.083, 2.775)** |
| | GNG | 0.182 | 0.137 | 0.102 | 0.080 | 0.167 | 0.215 | 0.327 | **0.801** |
| Gumbel($\sigma = 1$) | | (0.107, 0.264) | (0.099, 0.174) | (0.085, 0.119) | (0.068, 0.091) | (0.148, 0.186) | (0.188, 0.241) | (0.278, 0.377) | **(0.680, 0.926)** |
| | CONGNG | **0.140** | 0.177 | 0.087 | 0.050 | 0.137 | 0.188 | 0.395 | 1.369 |
| | | **(0.118, 0.161)** | (0.156, 0.198) | (0.073, 0.100) | (0.043, 0.057) | (0.120, 0.154) | (0.164, 0.212) | (0.324, 0.462) | (1.006, 1.720) |
| | GkerG | 0.165 | **0.080** | **0.048** | **0.036** | **0.100** | **0.129** | **0.240** | 0.826 |
| | | (0.117, 0.217) | **(0.067, 0.093)** | **(0.042, 0.054)** | **(0.030, 0.042)** | **(0.088, 0.112)** | **(0.113, 0.144)** | **(0.207, 0.274)** | (0.667, 1.008) |
| | GNG | **0.349** | 0.178 | 0.130 | 0.114 | 0.143 | 0.161 | 0.204 | 0.413 |
| N.Weibull($l = 10, k = 5$) | | **(0.306, 0.391)** | (0.152, 0.203) | ( 0.111, 0.148) | (0.099, 0.130) | (0.121, 0.166) | (0.138, 0.184) | (0.177, 0.231) | (0.358, 0.468) |
| | CONGNG | 0.406 | 0.244 | 0.177 | 0.118 | 0.153 | 0.211 | 0.279 | 0.438 |
| | | (0.331, 0.489) | (0.214, 0.273) | (0.158, 0.195) | (0.103, 0.134) | (0.129, 0.178) | (0.183, 0.240) | (0.249, 0.308) | (0.392, 0.483) |
| | GkerG | 0.366 | **0.157** | **0.111** | **0.095** | **0.123** | **0.121** | **0.195** | **0.407** |
| | | (0.308, 0.423) | **(0.137, 0.178)** | **(0.096, 0.126)** | **(0.079, 0.111)** | **(0.105, 0.141)** | **(0.101, 0.140)** | **(0.162, 0.233)** | **(0.352, 0.461)** |
| | GNG | **0.781** | 0.279 | 0.120 | 0.082 | 0.182 | 0.269 | 0.487 | 2.027 |
| T($v = 4, \mu = 1$) | | **(0.657, 0.909)** | (0.216, 0.347) | (0.100, 0.141) | (0.073, 0.091) | (0.162, 0.201) | (0.234, 0.304) | (0.433, 0.540) | (1.835, 2.208) |
| | CONGNG | 0.860 | **0.184** | **0.085** | **0.068** | **0.110** | **0.167** | **0.440** | **1.960** |
| | | (0.677, 1.064) | **(0.154, 0.216)** | **(0.072, 0.098)** | **(0.059, 0.076)** | **(0.094, 0.126)** | **(0.142, 0.192)** | **(0.389, 0.489)** | **(1.741, 2.173)** |
| | GkerG | 0.823 | 0.231 | 0.104 | 0.071 | 0.121 | 0.176 | 0.556 | 2.394 |
| | | (0.722, 0.922) | (0.197, 0.265) | (0.088, 0.121) | (0.060, 0.082) | (0.105, 0.137) | (0.149, 0.204) | (0.455, 0.665) | (2.046, 2.737) |

Table 5.16: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using bulk model based tail fraction approach with sample size 5000

| Distribution | Model | 0.1% | 1% | 5% | 10% | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|---|---|---|---|
| | GNG | **0.267** | 0.136 | 0.078 | 0.068 | 0.066 | 0.094 | 0.152 | **0.302** |
| Normal($\mu = 0, \sigma = 3$) | | **(0.233, 0.301)** | (0.119, 0.153) | (0.068, 0.088) | (0.058, 0.077) | (0.057, 0.075) | (0.080, 0.108) | (0.133, 0.171) | **(0.256, 0.346)** |
| | CONGNG | 0.314 | **0.134** | **0.073** | **0.062** | **0.056** | **0.068** | **0.125** | 0.325 |
| | | (0.274, 0.354) | **(0.115, 0.154)** | **(0.064, 0.082)** | **(0.055, 0.069)** | **(0.048, 0.064)** | **(0.058, 0.078)** | **(0.106, 0.144)** | (0.283, 0.366) |
| | GkerG | 0.352 | 0.170 | 0.097 | 0.071 | 0.065 | 0.095 | 0.166 | 0.334 |
| | | (0.299, 0.405) | (0.138, 0.203) | (0.079, 0.115) | (0.058, 0.083) | (0.058, 0.073) | (0.081, 0.109) | (0.145, 0.188) | (0.279, 0.388) |
| | GNG | **1.258** | 0.245 | 0.121 | 0.056 | 0.057 | 0.119 | 0.238 | 1.330 |
| T($v = 3, \mu = 0$) | | **(1.123, 1.388)** | (0.201, 0.291) | (0.107, 0.134) | (0.049, 0.063) | (0.050, 0.064) | (0.104, 0.133) | (0.209, 0.265) | (1.184, 1.471) |
| | CONGNG | 1.274 | **0.202** | 0.073 | 0.044 | 0.045 | 0.078 | **0.196** | **1.152** |
| | | (1.137, 1.408) | **(0.169, 0.233)** | (0.064, 0.082) | (0.037, 0.051) | (0.039, 0.051) | (0.066, 0.090) | **(0.163, 0.229)** | **(1.044, 1.257)** |
| | GkerG | 1.380 | 0.217 | **0.069** | **0.043** | **0.038** | **0.061** | 0.204 | 1.378 |
| | | (1.164, 1.594) | (0.182, 0.251) | **(0.057, 0.082)** | **(0.037, 0.049)** | **(0.033, 0.043)** | **(0.050, 0.071)** | (0.170, 0.238) | (1.181, 1.581) |
| | GNG | 0.064 | 0.080 | 0.068 | 0.074 | 0.165 | 0.197 | 0.212 | 0.406 |
| Gumbel($\sigma = 1$) | | (0.056, 0.071) | (0.059, 0.107) | (0.056, 0.082) | (0.063, 0.084) | (0.150, 0.179) | (0.175, 0.219) | (0.186, 0.238) | (0.354, 0.460) |
| | CONGNG | 0.139 | 0.158 | 0.081 | 0.039 | 0.105 | 0.151 | 0.164 | 0.498 |
| | | (0.110, 0.172) | (0.134, 0.181) | (0.064, 0.097) | (0.033, 0.046) | (0.085, 0.125) | (0.124, 0.178) | (0.128, 0.201) | (0.453, 0.543) |
| | GkerG | **0.057** | **0.037** | **0.020** | **0.020** | **0.050** | **0.063** | **0.114** | **0.382** |
| | | **(0.048, 0.066)** | **(0.032, 0.042)** | **(0.018, 0.023)** | **(0.017, 0.023)** | **(0.044, 0.055)** | **(0.055, 0.070)** | **(0.099, 0.129)** | **(0.326, 0.440)** |
| | GNG | 0.180 | 0.123 | 0.090 | 0.097 | 0.114 | 0.125 | 0.122 | **0.193** |
| N.Weibull($l = 10, k = 5$) | | (0.156, 0.203) | (0.101, 0.146) | (0.077, 0.103) | (0.089, 0.104) | (0.105, 0.122) | (0.114, 0.136) | (0.105, 0.138) | **(0.160, 0.229)** |
| | CONGNG | 0.231 | 0.283 | 0.181 | 0.089 | 0.086 | 0.156 | 0.234 | 0.205 |
| | | (0.198, 0.264) | (0.266, 0.301) | (0.170, 0.192) | (0.082, 0.096) | (0.077, 0.094) | (0.145, 0.167) | (0.219, 0.249) | (0.180, 0.230) |
| | GkerG | **0.158** | **0.076** | **0.048** | **0.044** | **0.054** | **0.062** | **0.087** | 0.216 |
| | | **(0.138, 0.178)** | **(0.066, 0.085)** | **(0.043, 0.054)** | **(0.038, 0.050)** | **(0.046, 0.061)** | **(0.054, 0.070)** | **(0.074, 0.101)** | (0.183, 0.249) |
| | GNG | 0.426 | 0.163 | 0.083 | 0.080 | 0.155 | 0.222 | 0.304 | 1.180 |
| T($v = 4, \mu = 1$) | | (0.369, 0.483) | (0.142, 0.184) | (0.074, 0.093) | (0.071, 0.088) | (0.139, 0.170) | (0.196, 0.247) | (0.266, 0.342) | (1.065, 1.291) |
| | CONGNG | 0.406 | 0.114 | 0.036 | 0.032 | 0.055 | 0.110 | **0.186** | **1.305** |
| | | (0.358, 0.453) | (0.100, 0.128) | (0.031, 0.041) | (0.028, 0.037) | (0.049, 0.061) | (0.098, 0.120) | **(0.164, 0.208)** | **(1.191, 1.412)** |
| | GkerG | **0.393** | **0.095** | **0.034** | **0.027** | **0.045** | **0.074** | 0.205 | 1.310 |
| | | **(0.343, 0.441)** | **(0.080, 0.110)** | **(0.030, 0.039)** | **(0.023, 0.031)** | **(0.039, 0.052)** | **(0.064, 0.085)** | (0.178, 0.232) | (1.062, 1.588) |

Table 5.17: The RMSE of quantile estimate and 95% bootstrap percentile confidence interval using likelihood inference approach with sample size 1000

| Distribution | Model | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| Normal($\mu = 0, \sigma = 3$) | kernelGPD | 0.169 (0.149, 0.190) | 0.187 (0.163, 0.210) | 0.315 (0.259, 0.376) | 0.690 (0.591, 0.789) |
| T($v = 3, \mu = 0$) | kernelGPD | 0.103 (0.088, 0.119) | 0.172 (0.144, 0.182) | 0.524 (0.421, 0.632) | 2.238 (1.829, 2.623) |
| Gumbel($\sigma$=1) | kernelGPD | 0.106 (0.091, 0.120) | 0.129 (0.110, 0.148) | 0.272 (0.235, 0.308) | 0.877 (0.760, 0.994) |
| N.Weibull ($l = 10, k = 5$) | kernelGPD | 0.127 (0.109, 0.146) | 0.141 (0.122, 0.159) | 0.209 (0.180, 0.237) | 0.445 (0.387, 0.502) |
| T($v = 4, \mu = 1$) | kernelGPD | 0.107 (0.093, 0.121) | 0.158 (0.138, 0.177) | 0.446 (0.387, 0.503) | 2.456 (2.037, 2.912) |
| Distribution | Model | 90% | 95% | 99% | 99.9% |
| Gamma($\alpha = 5, \beta = 1$) | BCkernelGPD | 0.167 (0.145, 0.188) | 0.217 (0.187, 0.246) | 0.431 (0.370, 0.492) | 1.115 (0.925, 1.305) |
| gamma($\alpha = 2, \beta = 2$)) | BCkernelGPD | 0.235 (0.208, 0.263) | 0.314 (0.271, 0.358) | 0.562 (0.485, 0.637) | 1.633 (1.415, 1.844) |
| Gamma($\alpha = 1, \beta = 2$) | BCkernelGPD | 0.210 (0.188, 0.239) | 0.290 (0.250, 0.329) | 0.559 (0.481, 0.638) | 1.497 (1.309, 1.684) |
| N.Chi-Squared($v = 2, \lambda = 2$) | BCkernelGPD | 0.271 (0.237, 0.304) | 0.344 (0.307, 0.380) | 0.707 (0.586, 0.835) | 2.055 (1.796, 2.302) |
| N.Chi-Squared($v = 2, \lambda = 6$) | BCkernelGPD | 0.430 (0.374, 0.484) | 0.543 (0.475, 0.610) | 0.966 (0.855, 1.075) | 2.140 (1.840, 2.434) |
| Inverse gamma($\alpha = 4, \beta = 4$) | BCkernelGPD | 0.080 (0.069, 0.091) | 0.133 (0.112, 0.152) | 0.382 (0.337, 0.425) | 1.958 (1.650, 2.327) |
| Inverse gamma($\alpha = 4, \beta = 8$) | BCkernelGPD | 0.148 (0.130, 0.165) | 0.231 (0.197, 0.264) | 0.757 (0.639, 0.876) | 3.318 (2.888, 3.728) |

# Chapter 6

# Applications

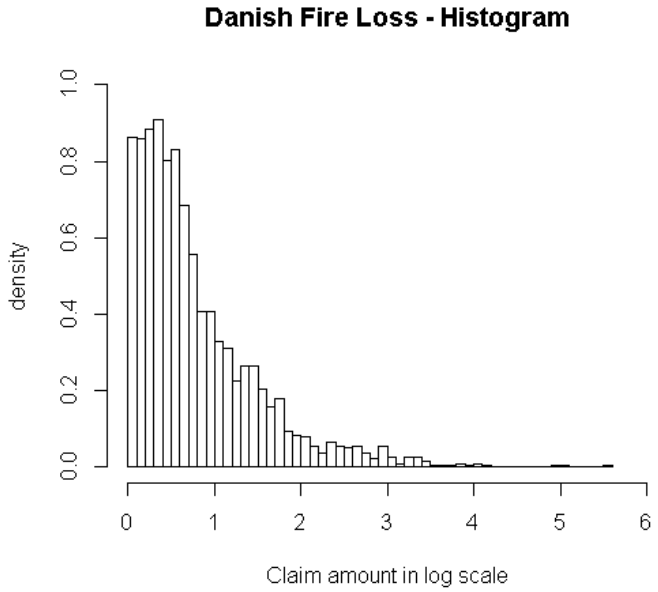
## 6.1 Introduction

This chapter will look at applications of extreme value mixture models to some of the classic datasets considered in extreme value applications: Danish fire insurance claims, FTSE 100 index daily return and Fort Collins precipitation data . All the relevant extreme value mixture models are applied to these three datasets by likelihood inference.

## 6.2 Application to Danish Fire Insurance Data

The Danish dataset contains 2167 fire losses in Denmark from January 1980 to December 1990. It is available in `evir` library in `R`. It will be rescaled by -1 and have a small residual added to it, so the data starts just above the zero. If the data is heavily rounded, the bandwidth parameter of KDE could be zero. Likelihood for KDE can be effected in this case.

Figure 6.1: The histogram of the Danish fire insurance data in log scale.



The Danish fire data has been intensively studied in the extreme value literature. McNeil (1997) and Embrechts (1999) apply the peaks over threshold method to estimate the quantile of the fire loss. Resnick (1997) uses several diagnostics for the independence assumption of the Dainish data and conclude that there is insufficient evidence to reject the independence hypothesis. He further acknowledges that McNeil's work is of great importance in insurance claim application and an excellent example of adopting the extreme value models. In the extreme value mixture model field, Frigessi et al. (2002), Carreau and Bengio (2008), Charpentier and Oulidi (2009) and Lee et al. (2012) have applied their mixture model to this heavy tailed data. Here all the relevant extreme value mixture models are fitted.

Figure 6.2 illustrates the mean residual life plot and Figure 6.3 gives the parameter stability plots for the Danish fire insurance data. McNeil (1997) has tried some subjective thresholds using POT method. Based on these 2 figures, I have considered 3 thresholds (4.5, 9, 10) as the initial value of the threshold for the likelihood inference. These three thresholds are corresponding to 90%, 95% and 99% quantile of the data. According to the Figure 6.2, the threshold is chosen the place where the mean excess above the threshold is linear. The parameter stability plots fits GPD over a certain values of thresholds against the shape and scale parameter. The threshold $u$ should be chosen where the shape and modified scale parameter remain constant after taking the sampling variability into account. All 3 choices are valid under general guideline for choosing the threshold. Hence, I decide to choose the threshold 4.5 (90% quantile) as the initial value for the threshold, which generates more excesses compared to using threshold 9 and 10.

Figure 6.2: The mean residual life plot for the Danish fire insurance data.

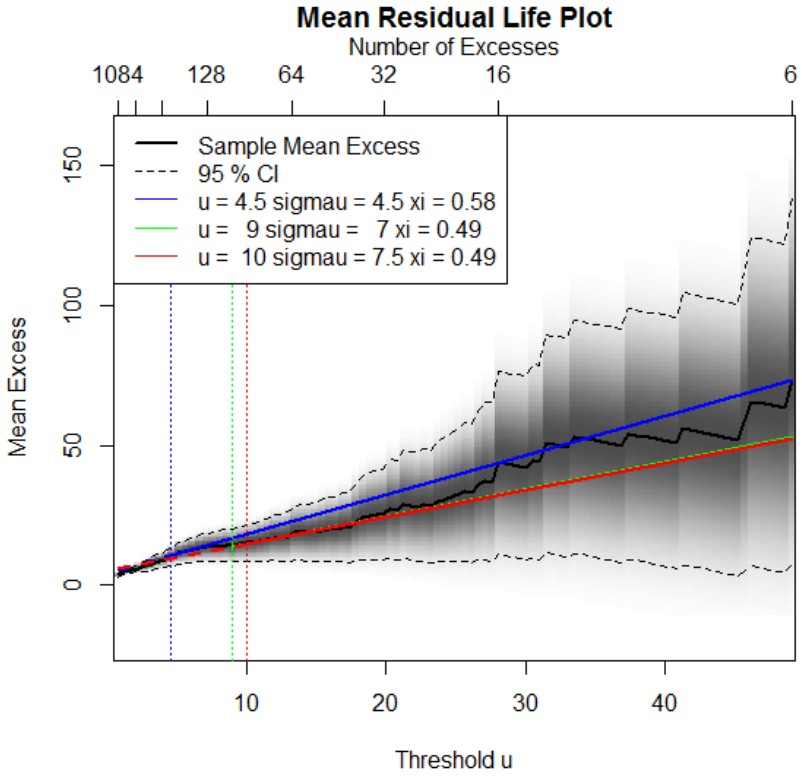
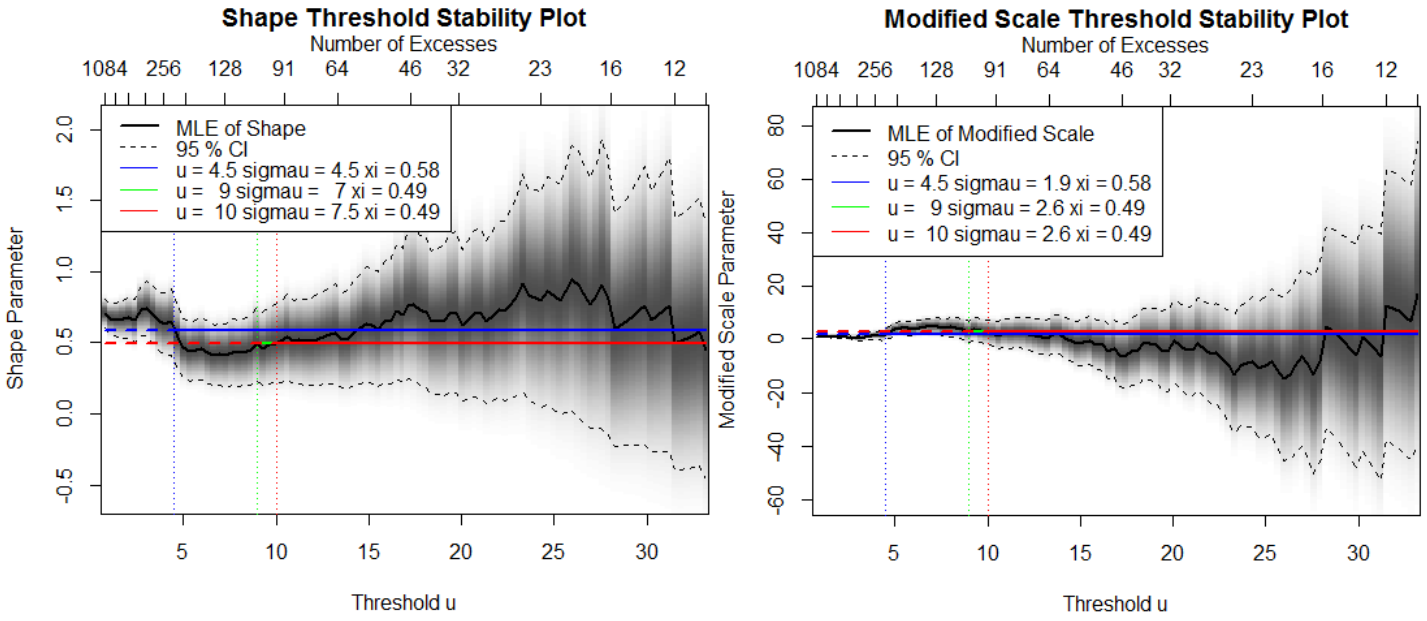


Figure 6.3: The parameter stability plot for the Danish fire insurance data.




Figures 6.4 and 6.5 are the corresponding return level plot for the Danish fire data. The return level plot is very useful for the model validation and particularly for evaluating the performance in the tail extrapolation. According to Figure 6.4, all the parametric bulk form mixture models give a reasonable and similar model fit to the upper tail up to the 99% quantile level. It seems that the bulk tail fraction approach based models (gammaGPD, WeibullGPD and CONweibullGPD) have a better fit between 99% and 99.9% quantile level. Above the 99.9% quantile, all 8 models can give a good fit to the very extreme upper tail. The differences between the two approaches will not be significant if the 95% confidence interval for the return level is added to the plot.

The simple boundary correction method of Jones (1993) is the default method for the boundary correction kernel GPD model. However, this method does not work well for this dataset. This method tends to overestimate the bandwidth and this issue will be investigated in the future. Alternatively, the renormalisation method of Diggle (1985) is used as the boundary correction method for the boundary corrected kernel GPD model. The dynamically weighted mixture and the boundary correction kernel GPD models can provide a reasonable and similar fit to the upper tail. Above the 99.9% quantile level, both two models can still give a good tail extrapolation.

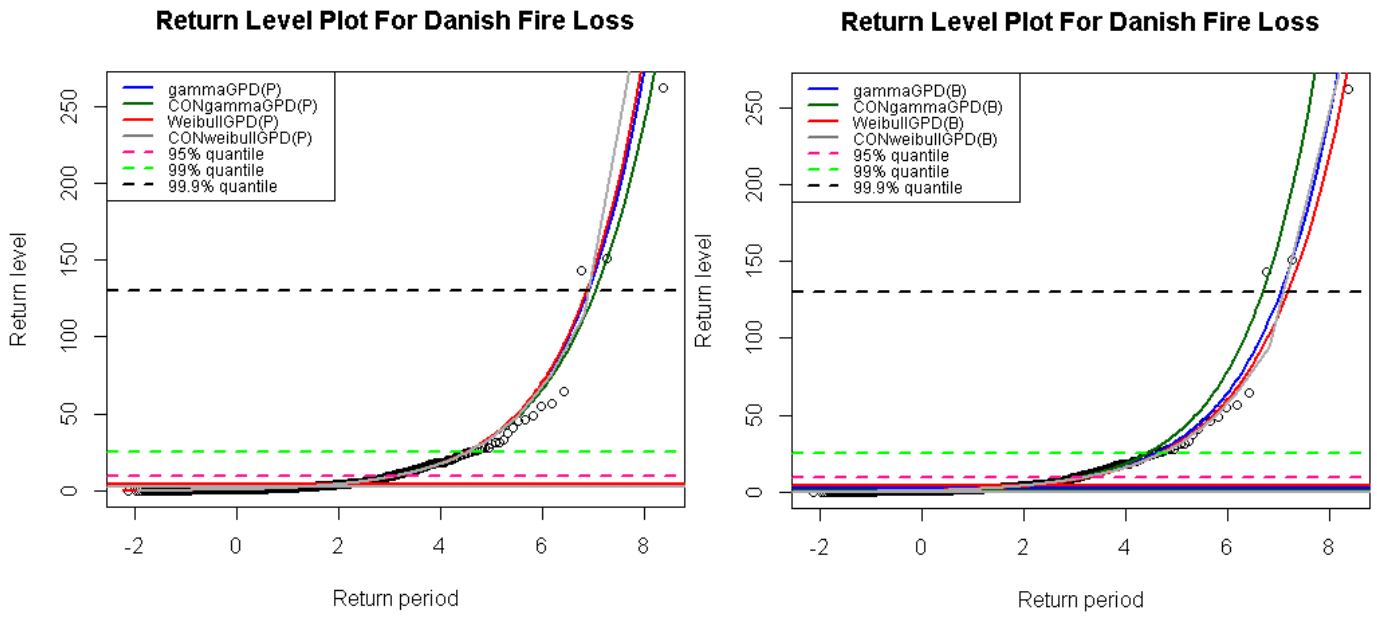



Figure 6.4: The fitted return level plot of Danish fire insurance claims data of parametric bulk mixture model using both parameterised and bulk approaches, where the horizontal line represents the corresponding threshold for each model and the horizontal dashed line represents the corresponding return level. The plot is constructed at a Gumbel scale.

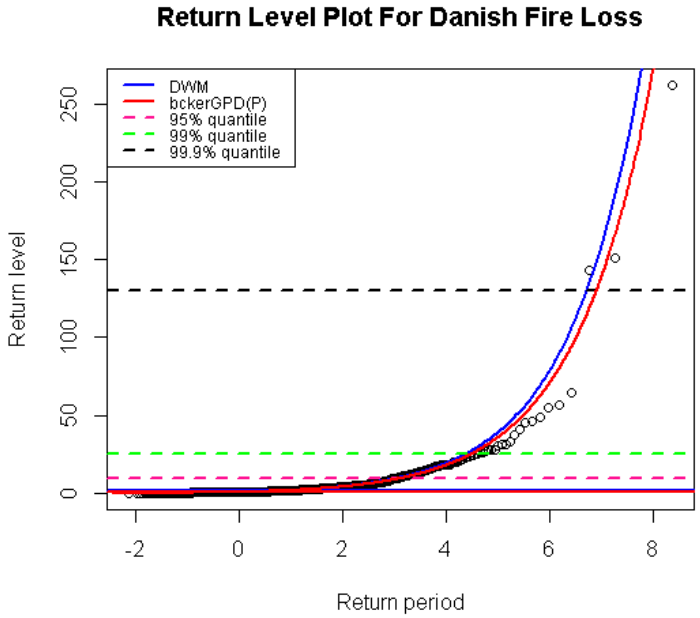



Figure 6.5: The fitted return level plot of Danish fire insurance claims data of dynamically weighted mixture and boundary corrected kernel GPD model, where the horizontal line represents the corresponding threshold for each model and the horizontal dashed line represents the corresponding return level. The plot is constructed at a Gumbel scale.

## 6.3 Application to FTSE 100 Index Log Daily Return Data

This section will examine all the relevant mixture models applied to data on from the FTSE 100 index. The FTSE 100 index is one of the most important stock indices in the world, which comprises 100 companies on the London stock exchange market. The FTSE data can be found at Yahoo Finance and the data is from 02/04/1984 to 28/02/2012 for the daily closing index. The log daily return of the FTSE 100 index is defined as : $r = \log\left(\frac{P_{i+1}}{P_i}\right)$, where $r$ is the log daily return and $P_i$ is the today's closing index. It is worth mentioning that financial data tends to exhibit heavy tails with dependence. The dependence issue will be ignored for the following analysis. All the positive bounded support extreme value mixture models considered in the simulation study will be applied to this data. The upper tail is of the interest.

Figure 6.6 provides the mean residual life plot for the FTSE 100 index log daily return data and Figure 6.7 gives the parameter threshold stability plots. The thresholds of 0.012, 0.017, and 0.031 have been considered as possible choices. The threshold 0.012 will be selected as the initial value for likelihood inference.

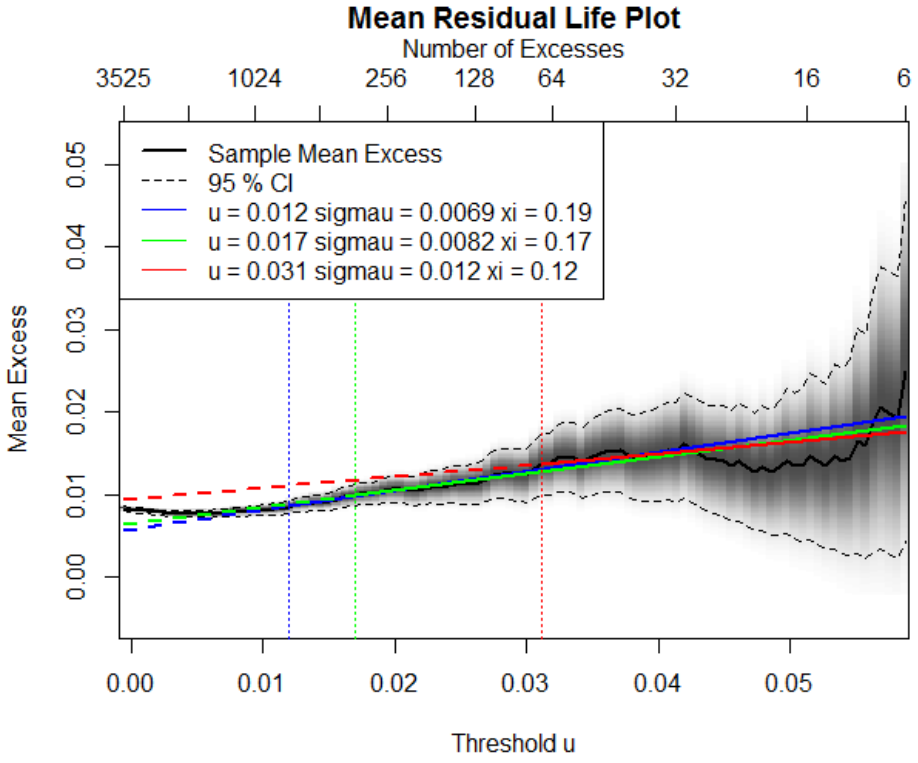



Figure 6.6: The mean residual life plot for FTSE 100 index log daily return data of the upper tail.

Figures 6.8 provides the histogram of the data overlaid with the fitted density plot. As can be seen, the parameterised approach based normal GPD and continuous normal GPD model tend to overestimate the mode of the FTSE data, resulting from the estimated standard deviation

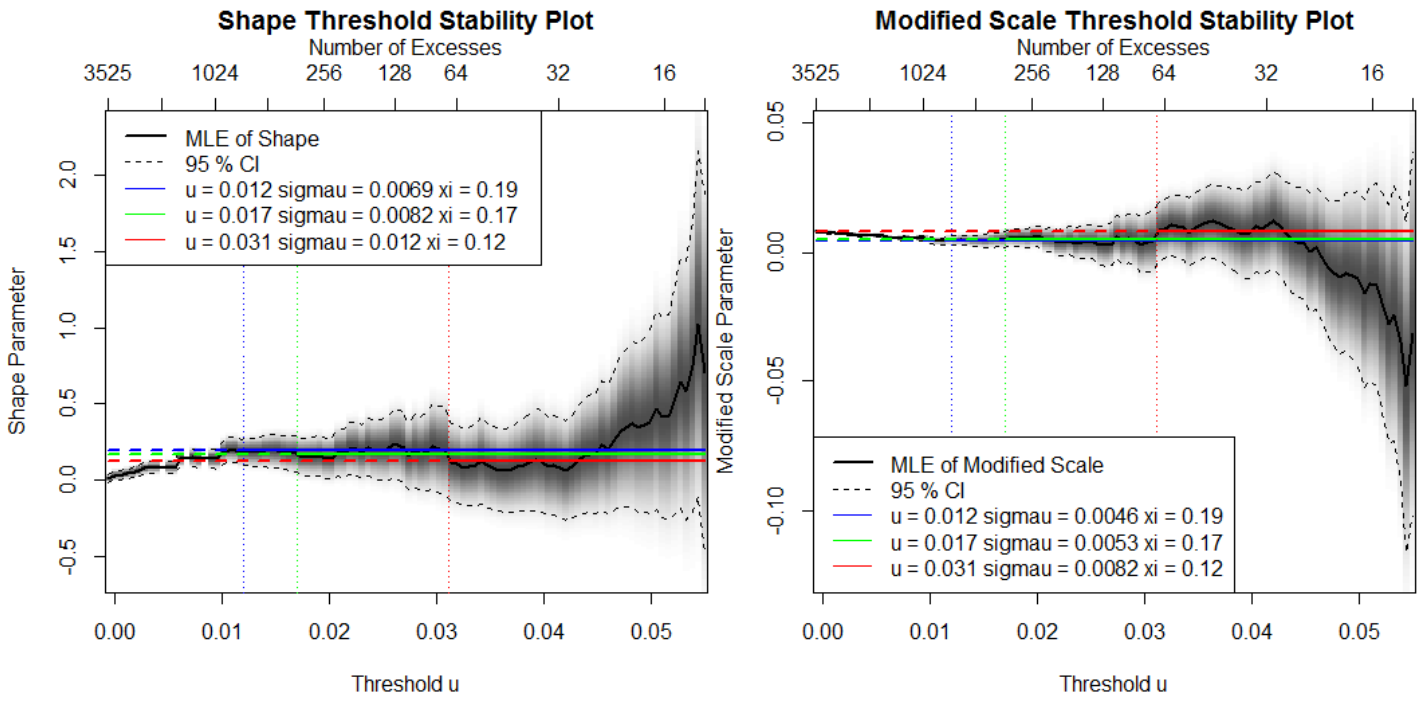



Figure 6.7: The parameter stability plot for FTSE 100 index log daily return data of the upper tail.

being too small. On the other hand, the bulk approach based normal GPD and continuous normal GPD model tend to underestimate the mode, resulting from having too large standard deviation. Both the overestimation and underestimation are caused by the heavy lower tail. It is commonly known that normal distribution cannot cope with a heavy tail.

It is also interesting to note that even when the bulk model does not provide a good fit, the GPD is still able to cope with the heavy upper tail very well even if the bulk model is mis-specified. However, the hybrid Pareto and single continuity constraint hybrid Pareto model are completely mis-specified for the FTSE data, which lead to very poor fit. Both models fail to provide sufficient flexibility to the FTSE data to overcome the lack of the tail fraction $\phi_u$ term. This could explain the relatively poor performance of these two models. It is clear that the kernel GPD model gives a very good fit for both tails and the corresponding return level plot is not shown for brevity.

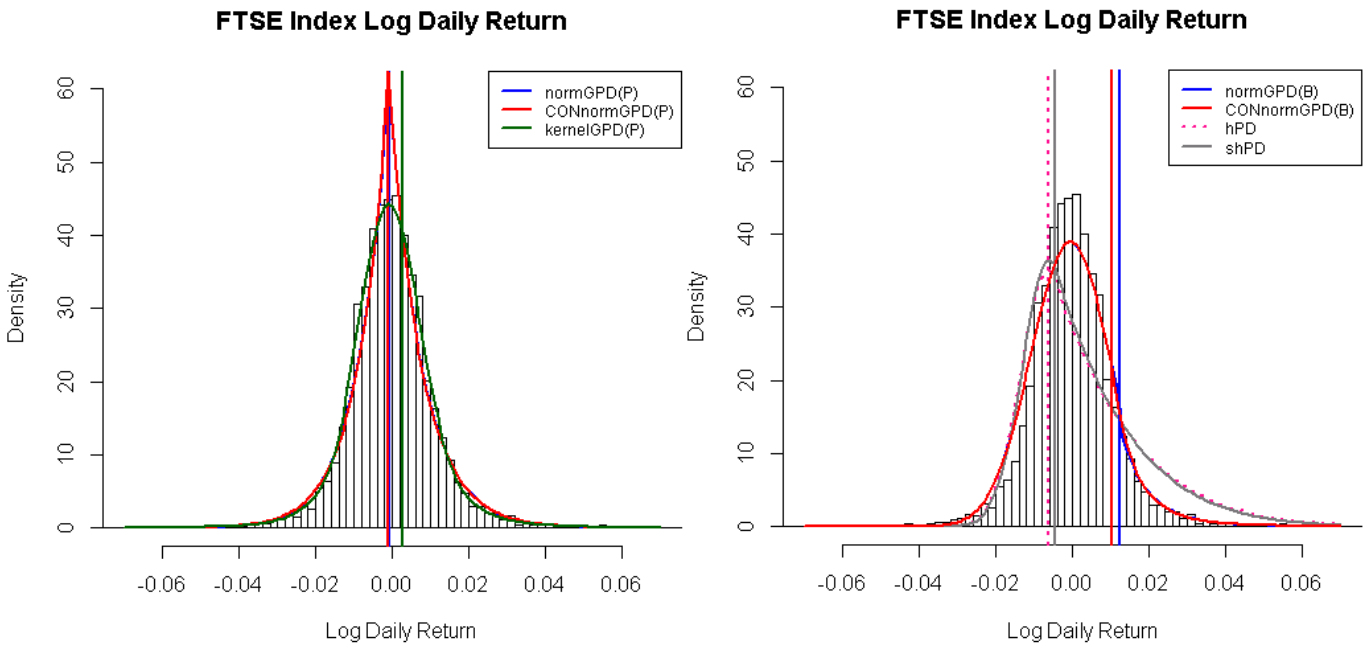



Figure 6.8: Left: Histogram of FTSE 100 index log daily return data fitted by parametric bulk mixture and kernel GPD models using parameterised approach. Right: Histogram of FTSE 100 Index Log Daily Return fitted by parametric bulk mixture model using bulk approach and two hybrid Pareto models, where the vertical line represents the corresponding threshold for each model.

### Application to FTSE 100 Index Log Daily Return Data(Two tailed)

As explained earlier, both the normal GPD and continuous normal GPD model are affected by the heavy lower tail. Moreover, the cross-validation likelihood based kernel GPD model is sometimes problematic due to the inconsistency of the cross-validation likelihood based bandwidth estimator for heavy tailed population distribution. One way of solving such issues is to extend the single tail model to a two tailed model by splicing the bulk model with both lower and upper extreme value models (GPD-normal-GPD, continuous GPD-normal-GPD and GPD-kernel-GPD).

Figure 6.9 gives the histogram of the data overlaid with the fitted density. All 5 models are able to provide a reasonable fit. It can be observed from Figure 6.10 that a good tail exploration can be obtained in continuous GNG and two tailed kernel GPD models, which highlights the usefulness of the two tailed extreme value mixture models. The results also highlight the feature of these two tailed models is that these models can capture all sorts of tail behavior, especially for both heavy lower and upper tail. The two tailed kernel GPD model is more flexible to other models as the choice of the bulk distribution could affect the tail estimation.

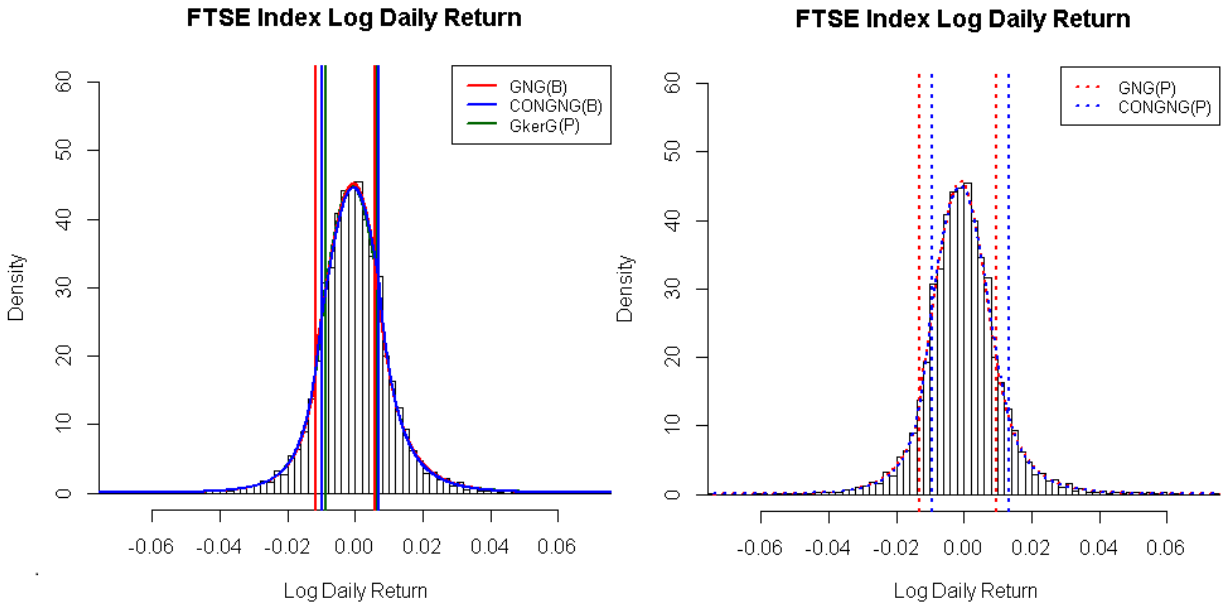



Figure 6.9: Left: Histogram of FTSE 100 index log daily return data fitted by two tailed normal GPD, continuous two tailed normal GPD using bulk approach and two tail kernel GPD model using parameterised approach. Right: Histogram of data fitted by two tailed normal GPD, continuous two tailed normal GPD using parameterised approach. The vertical lines represents either the corresponding lower or upper threshold for each model.

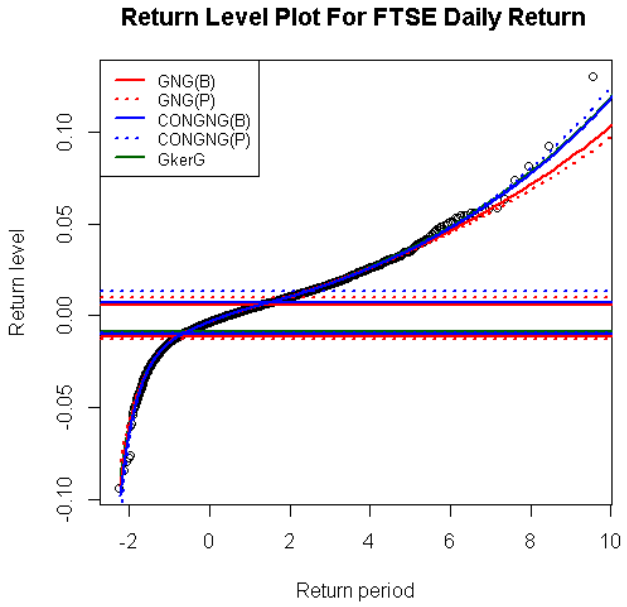



Figure 6.10: The return level plot of FTSE 100 index log daily return data by the 5 two tailed mixture models, where the horizontal lines represent either the lower or upper corresponding threshold for each model. The plot is constructed at a Gumbel scale.

## 6.4 Application to Fort Collins Precipitation Data

The Fort Collins precipitation rate is a very popular dataset in the hydrology field, which contains the daily precipitation rate in Fort Collins, Colorado, USA. This data is available in `extRemes` package in `R`. It will be shifted by removing the zero values for the following analysis as zero is not inside the support of gamma and Weibull distributions.

The feature of this data is that there are multiple choices for the threshold. Scarrott and MacDonald (2012) has demonstrated the multiple choices of the threshold selection for this data in detail and this data is discussed in Section 2.2.2 as well. In this section, all the positive bounded support models will apply this data with the thresholds of 0.395, 0.85 and 1.2. Katz et al. (2002) suggested the threshold $u$ to be 0.395 and Scarrott and MacDonald (2012) considered the thresholds of 0.85 and 1.2. These 3 different choices will be used as the initial value of the threshold in each model. The best fit model is determined by the maximum likelihood function value (or minimum negative likelihood function value). Figures 6.11 and 6.12 are the corresponding mean residual life and parameter stability plots for the Fort Collins data.

This data is also heavily rounded, so the bandwidth can be zero as the cross validation likelihood is utilized in the boundary correction kernel GPD model. Hence, a small residual has been added to the data. The default simple boundary correction method by Jones (1993) does not work well for this data. I have found that this default method tends to overestimate the bandwidth, which leads to a poor fit to the distribution below the threshold. This may be caused by the cross-validation likelihood which used for bandwidth parameter estimation. This issue requires further investigations. Therefore, I adopt the refection method of Boneva et al. (1971) as the boundary correction method for the boundary correction kernel GPD model.

Figures 6.13 and 6.14 provide the histogram of the data overlaid with the fitting density, which suggests that all models can achieve a reasonable fit. Figures 6.15 and 6.16 are the corresponding return level plot for data. Most mixture models can only give a reasonable fit up to the 99% quantile level. Beyond the 99% quantile level, the parameterised approach based gamma GPD and boundary correction kernel GPD model can still provide a good tail exploration.

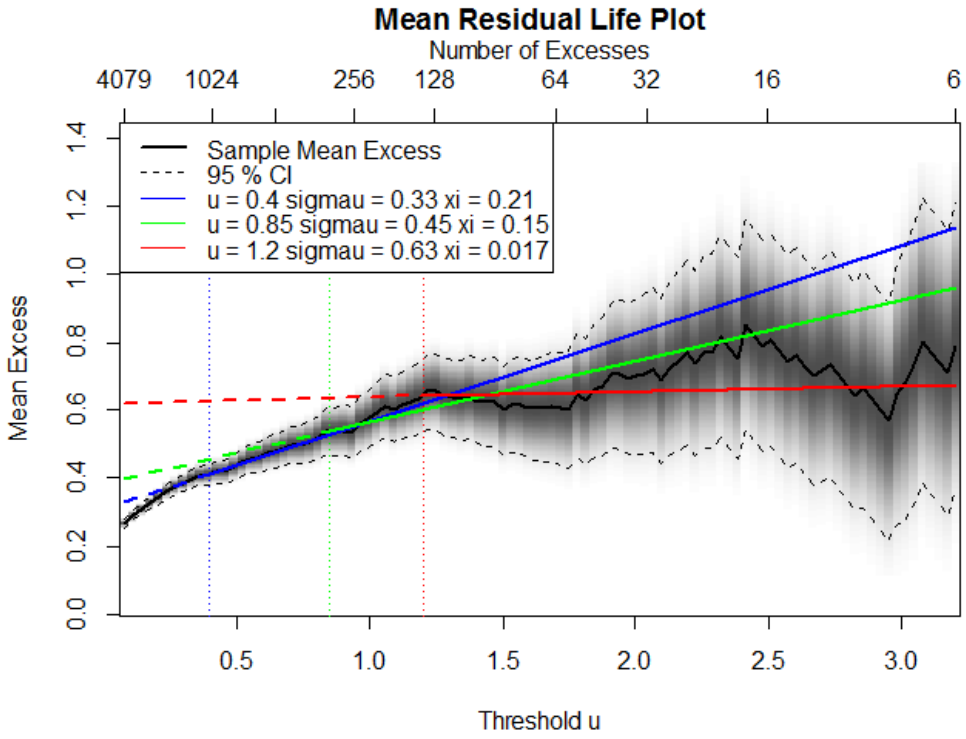



Figure 6.11: The mean residual life plot for Fort Collins precipitation data.

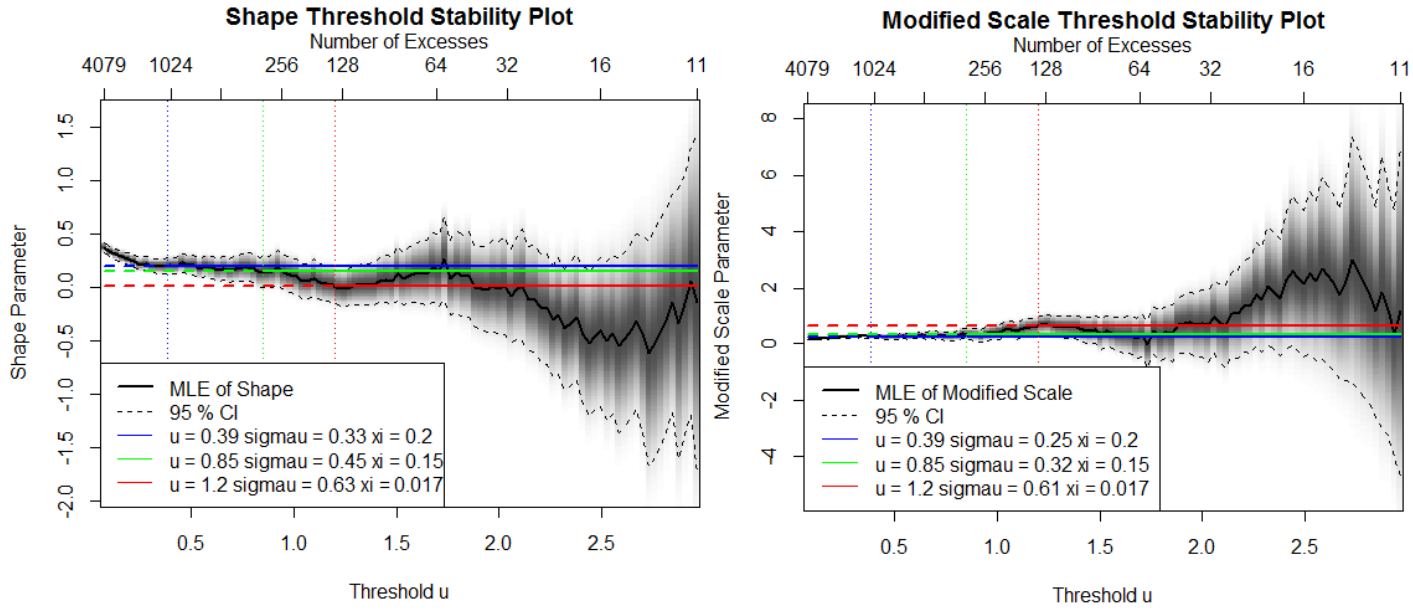



Figure 6.12: The parameter stability plot for Fort Collins precipitation data.

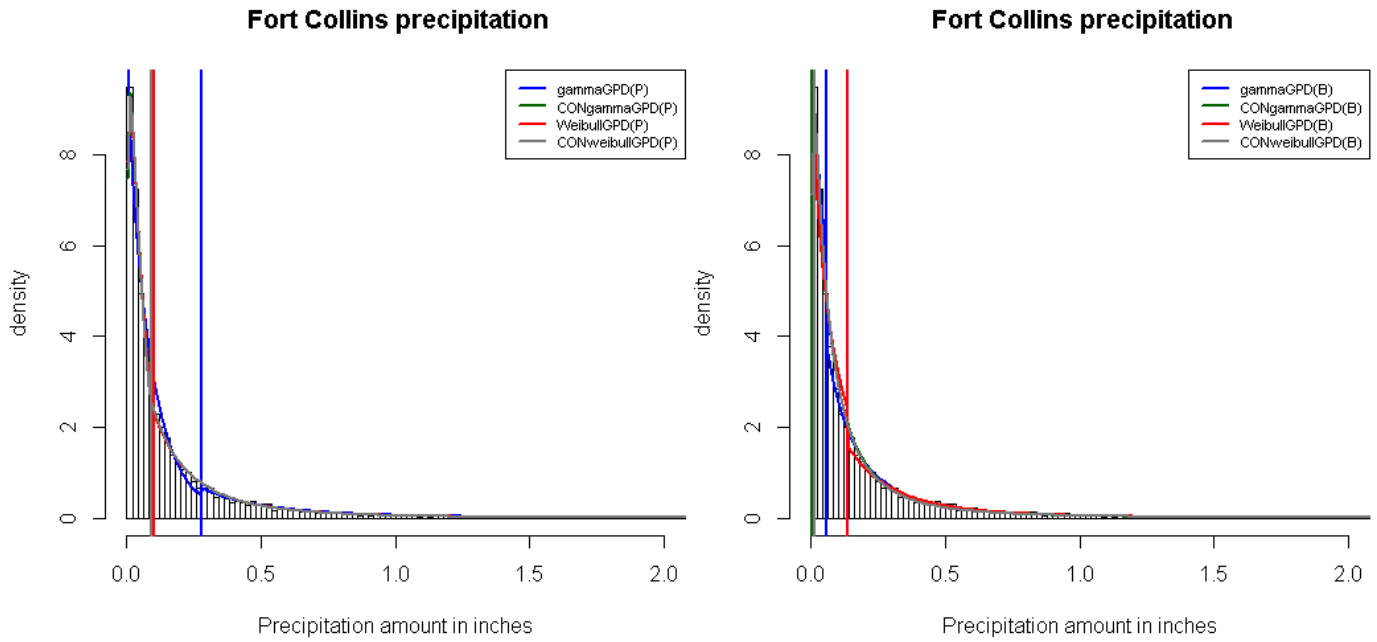



Figure 6.13: Histogram of Fort Collins precipitation data fitted by parametric bulk mixture model using parameterised and bulk approaches, where the vertical line represents the corresponding threshold for each model.

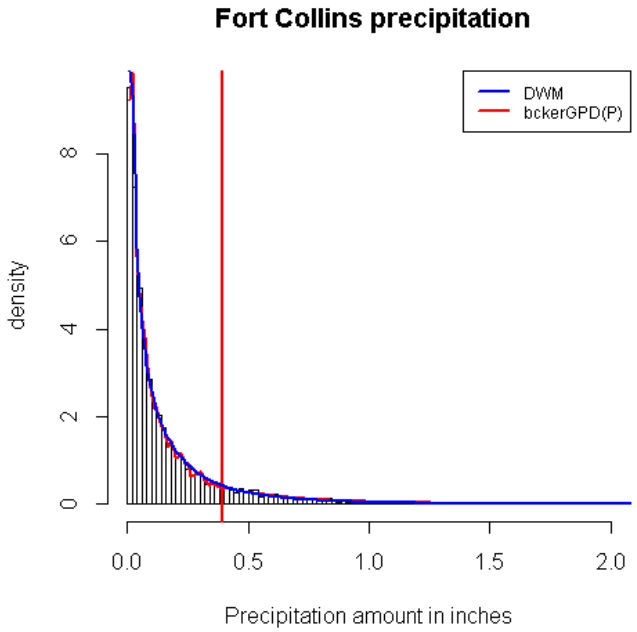



Figure 6.14: Histogram of Fort Collins precipitation data fitted by dynamically weighted mixture and boundary correction kernel GPD model, where the vertical line represents the corresponding threshold for each model.

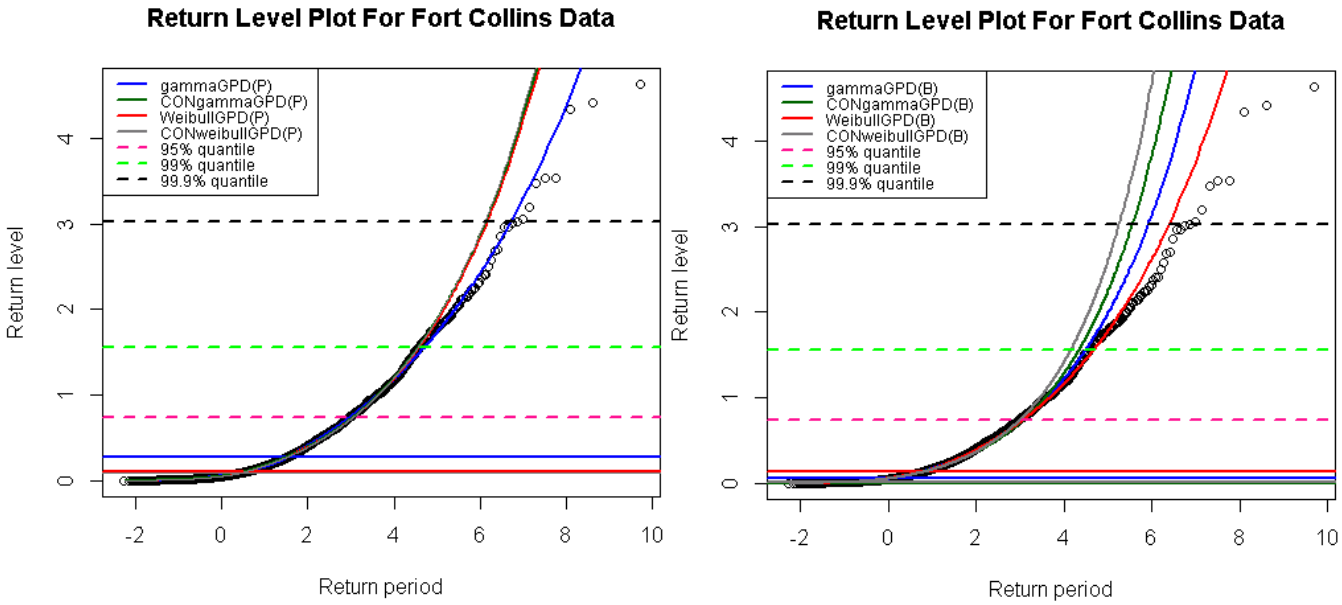



Figure 6.15: The return level plot of Fort Collins precipitation data fitted by parametric bulk mixture model using bulk and parameterised approaches, where the vertical line represents the corresponding threshold for each model. The 95%, 99%, 99.9% quantile of the Fort Collins data are also indicated in the plot. The plot is constructed at a Gumbel scale.

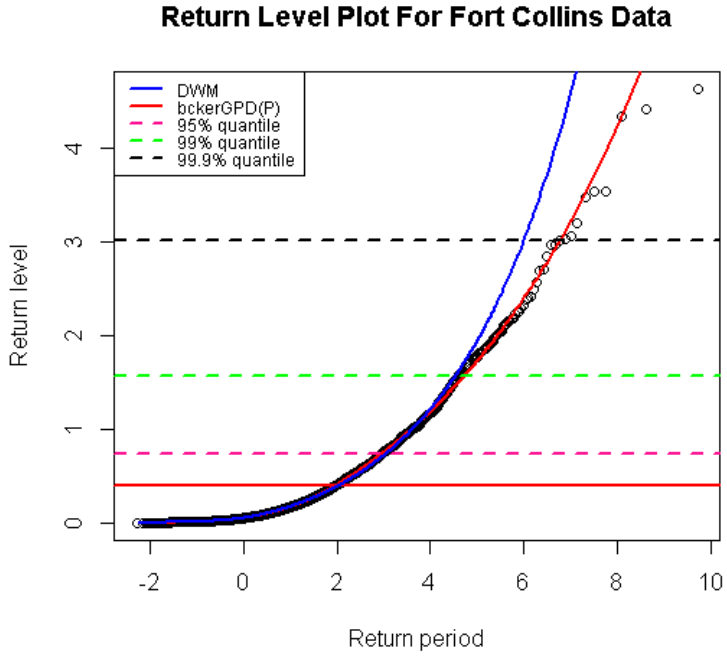



Figure 6.16: The return level plot of Fort Collins precipitation data fitted by dynamically weighted mixture and boundary correction kernel GPD model, where the vertical line represents the corresponding threshold for each model. The 95%, 99%, 99.9% quantile of the Fort Collins data are also indicated in the plot. The plot is constructed at a Gumbel scale.

# Chapter 7

# Conclusion and Future Work

## 7.1 Summary

This thesis evaluates the majority of the of extreme value mixture models in the literature, including implementing them in the statistical computing package `R` and carrying out a simulation study to compare their performance. I extended some of the existing models using different parametric components in the mixture and placing them in a more generalized framework to compare the performance under the various specifications used in the literature. The `R` package I created called `evmix` will be available on CRAN to increase the usability of such mixture models and enable these models to be further developed.

Chapter 2 reviews both the extreme value theory and the kernel density estimator. Chapter 3 presents a detailed literature review of existing extreme value models with some discussions of their properties. Some new model extensions which have continuous density at the threshold are also introduced. Chapter 4 details the user's guide to the `evmix` package, demonstrating the usage of the package with examples.

The simulation study which investigates the model performance is carried out in Chapter 5. The extreme value mixture models in the `evmix` package can be classified into three groups: a single tail with full support, a single tail with bounded positive support and two tailed with full support model by the support of the distribution. The simulation study provides some definitive conclusions about the properties of these mixture models. High quantile estimates (99.9%) are fairly robust to the choice of bulk model while the lower quantile estimates (90%, 95%, 99%) are very sensitive to the choice of the bulk model.

The results also indicate that the kernel density estimator based mixture models are superior or close to the best performer. However, Bayesian inference based kernel GPD and boundary corrected kernel GPD suffer huge computational burden. The likelihood inference based kernel GPD and boundary corrected kernel GPD are good alternatives. MacDonald et al. (2011) and MacDonald et al. (2013) develop three kernel density estimator based extreme value mixture models by Bayesian inference. Two of the three models are implemented by Bayesian inference, and all three models are implemented by the likelihood inference as well in the `evmix` library.

However, if the bulk model is correctly specified, then they are preferred due to their simplicity and good performance. Furthermore, in this case the bulk model based tail fraction approach is

beneficial as the tail inference essentially benefits from the extra data in the bulk, but provides a substantial disadvantage if the bulk model is mis-specified. The continuity constraint at the threshold is beneficial if the bulk model is correctly specified, but again is a disadvantage if the bulk model is mis-specified. If the bulk model is mis-specified, it is better to adopt the parameterised tail fraction approach, as it provides more flexibility and reduces the sensitivity of the tail fit on the bulk model.

In the more general situations, it is very difficult to judge whether the bulk model is mis-specified or not before fitting the data (e.g. an unknown population distribution or a complex data with bimodal or multimodal modes). The kernel density estimator based mixture models perform well in general under different population distributions, as the kernel density estimator does not assume any particular function form. A key problem with the kernel density estimator is that it does not perform well under a heavy tailed distribution due to the cross-validation likelihood based bandwidth is biased high. In this case, the kernel density estimator based mixture models can be included to have GPDs for both tails to overcome the issue. Therefore, it is better to adopt the non-parametric form mixture model rather than the parametric and semi-parametric form mixture models.

I have implemented three recent extremal mixture models in the package including two tailed normal GPD (GNG) by Zhao et al. (2010), two tailed kernel GPD by MacDonald et al. (2011) and the boundary corrected kernel GPD model by MacDonald et al. (2013), which are the most relevant models for financial applications. This work may provide a useful new insight for researchers in the field of extreme value statistics and financial models. These mixture models have been applied to the application in Danish fire insurance data and FTSE 100 index data in Chapter 6.

## 7.2 My Contributions

1 I have implemented most of the existing extreme value mixture models in the literature in the `evmix` package for `R`, which provides a very useful tool for researchers and practitioners in the field of extreme value. For each mixture model, the density, distribution, quantile, random number generation, likelihood and maximum likelihood based fitting function are provided in the `evmix` package and a Bayesian inference approach via MCMC has also been provided.

2 There is no consistent definition of the tail fraction in the extreme value mixture models in the literature. Some have used the proportion of the bulk model, which would have been above the threshold if it continued. Others have parameterised the tail fraction and estimated it using the sample proportion. I have placed all the mixture models (except hybrid Pareto and dynamically weighted mixture, which do not have a tail fraction scaling) into both frameworks, to evaluate their performance under each choice. Further the implementation in the `evmix` package also allows the user to utilise either approach.

3 I also extended some mixture models which have parametric model for the bulk distribution to be continuous at the threshold . The continuity constraint is achieved on the density by replacing the GPD scale parameter in terms of other parameters. I considered this extension for various different choices of bulk model (e.g. gamma, Weibull, and normal distribution).

A continuity version of the two tailed GPD-normal-GPD (GNG) model was also obtained by replacing the GPD scale parameters for both lower and upper tail. A single continuity constraint version of the hybrid Pareto model was developed. However, this model does not include the tail fraction in its derivation and this is a key difference compared to the normal GPD model with a single continuity constraint. Overall, the model performance is poor which is consistent with the original hybrid Pareto model. These two models highlight the importance of use of the tail fraction.

4 A detailed simulation study and applications have been used to compare the performance of these different mixture models, including whether the tail fraction specification and continuity constraint at the threshold improve the performance.

## 7.3 Future Work

The kernel density estimator has bias at the boundary. MacDonald et al. (2013) utilise Jones and Foster (1996)'s non-negative boundary corrected kernel density estimator to overcome the boundary bias. Another possibility is to adopt Marron and Ruppert (1994)'s transformation based boundary correction approach of the kernel density estimator

Future work could also consider alternative non-parametric density estimator to describe the bulk distribution. Kooperberg and Stone (1991) propose a non-parametric density estimator for data estimation, which is known as the logspline density estimation. The advantage of the logspline density estimation is that it works well for a sample size of the data as small as 50. They further claim that the corresponding confidence intervals for the quantiles can be easily obtained.

In the simulation study, only 100 simulations are carried out. In future research, it would be better to perform a larger simulation study.

# Bibliography


Behrens, C. N., H. F. Lopes, and D. Gamerman (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling 4*(3), 227–244.

Boneva, L., D. Kendall, and I. Stefanov (1971). Spline transformations: Three new diagnostic aids for the statistical data-analyst. *Journal of the Royal Statistical Society. Series B 33*(1), 1–71.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika 71*(2), 353–360.

Brewer, M. J. (2000). A Bayesian model for local smoothing in kernel density estimation. *Statistics and Computing 10*(4), 299–309.

Carreau, J. and Y. Bengio (2008). A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes 12*(1), 53–76.

Carreau, J. and Y. Bengio (2009). A hybrid Pareto mixture for conditional asymmetric fat-tailed distributions. *IEEE transactions on neural networks 20*(7), 1087–1101.

Charpentier, A. and A. Oulidi (2009). Beta kernel quantile estimators of heavy-tailed loss distributions. *Statistics and Computing 20*(1), 35–55.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer-Verlag: London.

Cooley, D., D. Nychka, and P. Naveau (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association 102*(479), 824–840.

Diggle, P. (1985). A kernel method for smoothing point process data. *Applied Statistics 34*(2), 138–147.

Duin, R. (1976). On the choice of smoothing parameters for Parzen estimators of probability density functions. *IEEE Transactions on Computers 25*(11), 1175–1179.

Embrechts, P. (1999). Extreme value theory as a risk management tool. *North American Actuarial 3*(2), 37–41.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Applications of Mathematics. Springer-Verlag: New York.

Frigessi, A., O. Haug, and H. Rue (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes 5*(3), 219–235.

Green, P. J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika 82*(4), 711.

Habbema, J., H. J, and v. d. B. K (1974). *A stepwise discriminant analysis program using density estimation*. Number 1. Physica-Verlag: Vienna.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) value of meteorological events. *Quarterly Journal of the Royal Meteorological Society 81*(348), 158–172.

Jones, M. and P. Foster (1996). A simple nonnegative boundary correction method for kernel density estimation. *Statistica Sinica 6*, 1005–1013.

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing 3*, 135–146.

Kang, S. and R. Serfozo (1999). Extreme values of phase-type and mixed random variables with parallel-processing examples. *Journal of Applied Probability 36*(1), 194–210.

Katz, R. W., M. B. Parlange, and P. Naveau (2002). Statistics of extremes in hydrology. *Advances in Water Resources 25*(8-12), 1287–1304.

Kooperberg, C. and C. J. Stone (1991). A study of logspline density estimation. *Computational Statistics & Data Analysis 12*(3), 327–347.

Leadbetter, M. R., G. Lindgren, and H. Rootzen (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. Springer-Verlag: New York.

Lee, D., W. K. Li, and T. S. T. Wong (2012). Modeling insurance claims via a mixture exponential model combined with peaks-over-threshold approach. *Insurance: Mathematics and Economics 51*(3), 538–550.

MacDonald, A. E., C. J. Scarrott, and D. S. Lee (2013). Boundary correction, consistency and robustness of kernel densities using extreme value theory. *Submitted*.

MacDonald, A. E., C. J. Scarrott, D. S. Lee, B. Darlow, M. Reale, and G. Russell (2011). A flexible extreme value mixture model. *Computational Statistics and Data Analysis 55*(6), 2137–2157.

Marron, J. and D. Ruppert (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological) 56*(4), 653–671.

McNeil, A. (1997). Estimating the tails of loss severity distributions using extreme value theory. *Astin Bulletin 27*, 117–137.

McNeil, A. J. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance 7*(3-4), 271–300.

Mendes, B. V. D. M. and H. F. Lopes (2004). Data driven estimates for mixtures. *Computational Statistics and Data Analysis 47*(3), 583–598.

Nascimento, F. F., D. Gamerman, and H. F. Lopes (2011). A semiparametric Bayesian approach to extreme value estimation. *Statistics and Computing 22*(2), 661–675.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics 33*(3), 1065–1076.

Pickands, J. (1971). The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability 8*(4), 745–756.

Reiss, R. and M. Thomas (2001). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhäuser: Berlin.

Resnick, S. (1997). Discussion of the Danish data on large fire insurance losses. *Astin Bulletin 1*, 1–12.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer Series in Operations Research and Financial Engineering. Springer-Verlag: New York.

Robert, C. P. and G. Casella (2004). *Monte Carlo statistical methods*. Springer Series in Statistics. Springer: New York.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics 27*(3), 832–837.

Rossi, F., M. Fiorentino, and P. Versace (1984). Two-component extreme value distribution for flood frequency analysis. *Water Resources Research 20*(7), 847–856.

Scarrott, C. J. and A. E. MacDonald (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT Statistical Journal 10*(1), 33–60.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall: London.

Smith, R. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika 72*(1), 67–90.

Smith, R. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science 4*(4), 367–393.

Tancredi, A., C. Anderson, and A. O'Hagan (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes 9*, 87–106.

Tawn, J. (1990). Modelling multivariate extreme value distributions. *Biometrika 77*(2), 245–253.

Thompson, P., Y. Cai, D. Reeve, and J. Stander (2009). Automated threshold selection methods for extreme wave analysis. *Coastal Engineering 56*(10), 1013–1021.

von Mises, R. (1954). La distribution de la plus grande de n valeurs. *American Mathematical Society: Providence RI II*, 271–294.

Wadsworth, J. L. and J. A. Tawn (2012). Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74*(3), 543–567.

Wand, M. and M. Jones (1995). *Kernel Smoothing*. Monographs on Statistics and Applied Probability. Chapman & Hall: London.

Watson, G. and M. R. Leadbetter (1963). On the estimation of the probability density, I. *Annals of Mathematical Statistics 34*(2), 480–491.

Zhao, X., C. J. Scarrott, L. Oxley, and M. Reale (2010). Extreme value modelling for forecasting market crisis impacts. *Applied Financial Economics 20*(1-2), 63–72.

Zhao, X., C. J. Scarrott, L. Oxley, and M. Reale (2011). GARCH dependence in extreme value models with Bayesian inference. *Mathematics and Computers in Simulation 81*(7), 1430–1440.