Delft Institute of Applied Mathematics
Faculty of Electrical Engineering, Mathematics and Computer Science

Non-parametric extreme quantile estimation for the common shaped tail model

Forecasting extreme precipitation by post-processing precipitation from a numerical weather prediction model

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfilment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

Jasper Jonathan Velthoen

Delft, the Netherlands
October 2016

MSc Thesis APPLIED MATHEMATICS

"Non-parametric extreme quantile estimation for the common shaped tail model"

Jasper Jonathan Velthoen

Delft University of Technology

Daily Supervisors	**Responsible professor**
Dr. J. Cai	Prof. Dr. Ir. G. Jongbloed
Prof. Dr. Ir. G. Jongbloed	

Other committee members

Dr. P. Cirillo	Dr. M. Schmeits (KNMI)

October, 2016 Delft

Abstract

The estimation of extreme quantile curves of a family of conditional distributions is a non-trivial problem, due to the data-sparseness in the tail of the distribution. This thesis considers the problem of post-processing extreme precipitation forecasts in Friesland from the numerical weather prediction model HARMONIE. Assuming forecasts are accurate, it is natural to assume a linear relationship between the precipitation observations and the forecasts. However, in practice this relationship is not linear, due to large uncertainties in the modelling process. To deal with this problem of non-linearity a non-parametric common shaped tail estimator (CST) is proposed to adequately estimate the non-linear shape of the relationship. Performance of the CST estimator is shown and compared to other extreme quantile estimators from the literature in an extensive simulation study. The estimators are also compared using the quantile verification skill score on a three year precipitation dataset in the province of Friesland. It is shown that the estimator has a lot of bias in the estimation, which prevents the CST estimator to outperform all other methods. However, the verification scores for the real data example exceed all scores of the other estimators.

Contents

Chapter 1

Introduction

Weather has a large influence on our daily lives. Everyday people make simple decisions based on the weather. For example whether they should take the car or the bike to work, or how to dress to be comfortable on a certain day. These are all relatively easy choices, that have little impact, though there are also more critical problems which are heavily influenced by weather conditions. For example wind turbines on the North Sea need to be stopped in order to prevent damage when wind gusts become too severe. Control of the water levels is another important issue in the Netherlands. Proper precautions need to be taken in the case of a forecast extreme precipitation event, in order to prevent flooding.

All the above examples require accurate weather forecasts, but the wind turbine and extreme precipitation examples require extreme weather forecasts. Low wind speeds and little precipitation are not of interest in the prevention of the turbines breaking down and flooding problems.

Weather forecasts are generally generated from a numerical weather prediction (NWP) model. This model approximates the solution to the physical differential equations, which govern the weather. These differential equations are non-linear which means that a small misspecification in the initial or the boundary conditions can result in a completely different forecast. Furthermore, as not all processes can be explicitly resolved, the forecast is sensitive to model errors in sub grid scale parametrisations. This makes deterministic forecasts less useful, especially for extreme weather forecasts. Instead statistical methods are used to be able to make a probabilistic forecast of the weather event using the historical relationship between the forecast and the observations.

Consider Y to be the weather event and denote by X the forecast from the NWP model. The interest lies in the conditional distribution of the weather event given the forecast, i.e. the distribution of $Y|X$. When proper inference can be made on the distribution of $Y|X$, decision rules can be created.

To make this more concrete, consider an alert system for extreme precipitation, i.e. a system that issues an alarm when a heavy precipitation event is expected e.g. [14]. Therefore two values need to be fixed. First let y_{ext} denote a threshold value such that, when the amount of precipitation exceeds y_{ext}, it is considered an extreme event. Second, let τ denote a pre-specified probability such that an alarm is given when $\mathrm{P}(Y > y_{ext}|X = x) > \tau$, i.e. the probability of observing an extreme precipitation event is larger than the pre-specified probability.

This alarm system could also be written in terms of the quantile function $Q_{Y|X}$ of the random variable $Y|X$, i.e. issue an alarm when $Q_{Y|X}(1 - \tau|X = x) > y_{ext}$. An example of an alarm system is the code orange weather alert of the KNMI. Here a code orange is issued for $\tau = 0.6$.

Clearly the performance of these alarm systems relies on adequately modelling the distribution of $Y|X$. When the modelling of the distribution improves, also the alarm system will improve. This thesis discusses the estimation of the quantile function in the tail, i.e. what

happens with the quantile function $Q_{Y|X}(\tau|X = x)$ when $\tau \to 1$. Four approaches to the estimation of $Q_{Y|X}(\tau|X = x)$ are discussed, a linear quantile estimator [19], a non-parametric quantile estimator [2], a likelihood estimation of the generalized Pareto distribution to the exceedances above a high threshold [7] and a non-parametric kernel estimator for the distribution function [6]. These four estimators allow for a varying extreme value index over the covariates X. This allows for a lot of freedom in the modelling process, but it also introduces large uncertainties in the estimation. In Chapter 3 it is shown that these estimators introduce indeed a lot of uncertainty when the data has the property of a stationary extreme value index. In addition, in the linear quantile estimator the problem of quantile crossing appears, which will be elaborated on in Chapter 2.

The aim of this thesis is to develop a non-parametric estimator, not affected by the quantile crossing problem, for data that has the property of a stationary extreme value index. The common shaped tail model is proposed, which is an additive model for the quantile function by splitting the dependence on the covariate X and the dependence on τ hence,

$$Q(\tau|X = x) = r(x) + \epsilon(\tau).$$

Based on this model a common shaped tail (CST) estimator is proposed.

The outline is as follows, first in Chapter 2 basic theory on extreme value statistics, quantile regression and non-parametric estimation is discussed. Then the four approaches in the literature for the extreme quantile estimation are discussed in Chapter 3. Chapter 4 introduces the new CST estimator. An extensive simulation study will be done in Chapter 5 to show the performance of the CST estimator and to compare it with the estimators from the literature based on a finite data sample. In Chapter 6 the methods will be applied and verified on a real data example. Finally, in Chapter 7, concluding remarks are made and future work is discussed. In Appendix B documentation is listed for the R-code that is written for the implementation of the CST estimator and the two non-parametric estimators from Chapter 3.

Chapter 2

Prerequisite information

The general topic covered in this thesis is the estimation of extreme quantiles of the conditional distribution $Y|X$. Before discussing different approaches to this estimation problem, some statistical theory is introduced in the current chapter which will be important later on. This chapter is subdivided in three sections. The first section covers basic theory on quantile regression that was first introduced by Koenker in [13]. The second section, covers the basic theory on univariate extreme value theory and finally, the last section explains an approach to perform non-parametric quantile regression.

In order to maintain consistency through the entire thesis, the basic notation will be fixed here. Let $(Y_1, X_1), \ldots, (Y_n, X_n)$ be paired response and covariates. Denote $F(y|X = x)$ as the distribution function of $Y|X = x$ and let the conditional quantile function be given by $Q(\tau|X = x) = F^{-1}(\tau|X = x)$, where $\tau \in [0, 1]$ is a percentile level.

2.1 Quantile regression

Regression techniques are applied in statistics to obtain a dependence relation between a response variable and covariates. Inference about such a relation is always restricted to a specific part of the conditional distribution $Y|X$. For example least squares regression obtains the relation between Y and X by modelling the conditional mean, i.e. $\mathbf{E}(Y|X = x)$. Assume that $Y|X = x$ is a symmetric distribution for all values of x, and additionally assume the expectation exists; in that case it must hold that $\mathbf{E}(Y|X = x) = Q(0.5|X = x)$ for all x. This means the conditional mean is exactly the same as the conditional median. This is not the case when $Y|X = x$ is not symmetric or even heavily skewed. It would therefore be interesting to have a regression method to obtain inference about the conditional median, which always exists, instead of the conditional mean.

Fixing a sequence of percentile levels $\tau_1, \ldots, \tau_p$ and obtaining inference about the conditional quantiles at each of these percentile levels allows us to estimate the conditional quantile function $Q(\tau|X = x)$ as a smooth function of x and a discrete function in τ.

For a fixed percentile level τ, a quantile curve as a function of x, $x \mapsto Q(\tau|X = x)$ is called a regression quantile. A regression quantile corresponding to percentile level τ can intuitively be considered as a curve which splits the observations $(Y_1, X_1), \ldots, (Y_n, X_n)$ in a portion τ under the curve and $1 - \tau$ above it for every interval of X.

Using this intuitive idea of the regression quantile, a proper distance measure which leads to a consistent estimator for the quantile curve is given by,

$$\rho_\tau(u) = u(\tau - \mathbf{1}_{u<0}) = \max(u, 0)\tau + \max(-u, 0)(1 - \tau). \tag{2.1}$$

Here u is defined by the ℓ_1 distance between the curve, $r(x)$, and the response, $|Y - r(X)|$. Equation 2.1 is referred to in the literature as the quantile check function. The regression quan-

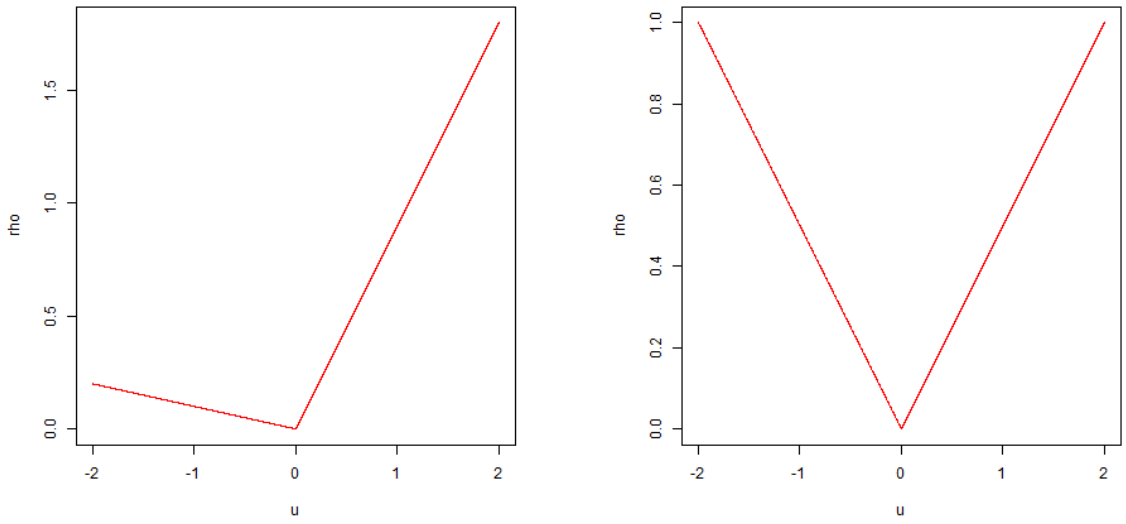

Figure 2.1: The quantile check function with $\tau = 0.5$ (left) and $\tau = 0.9$ (right).

tile is obtained by minimizing the expectation of the check function. Intuitively, the obtained regression quantile minimizes the distance between the observations above and below the regression quantile relative to the expected number of observations above and below. In Figure 2.1 the quantile check function is shown for $\tau = 0.5$ and $\tau = 0.9$. It can be seen from the figure that the function is symmetric for $\tau = 0.5$, which is intuitive should have half the points above and half the points below, i.e. $u > 0$ and $u < 0$ respectively. However, for $\tau = 0.9$ this the observations above and below the line are 1 to 9, which means that the observations above, $u > 0$ should have higher weight than observations below $u < 0$. Now given the observations denoted as above, the quantile curve $r_\tau(x)$ is given by the following minimization problem,

$$\hat{f}_\tau(x) = \text{argmin}_{r_\tau} \sum_{i=1}^{n} \rho_\tau(Y_i - r_\tau(x_i)). \tag{2.2}$$

Now in order to perform the minimization a choice is to be made for the class of the functions r_τ, as otherwise the minimization gives $r_\tau(X_i) = Y_i$. Often this class is chosen to be the linear functions, i.e. $r_\tau = \alpha + \beta X_i$, but this can be any class of functions in general.

Below a common encountered problem will be discussed, namely the quantile crossing problem, which appears when regression quantiles are estimated for a sequence of percentiles $\tau_1, \ldots, \tau_p$. Finally, the section will finish with a short discussion on the computational aspects of regression quantiles.

2.1.1 The quantile crossing problem

A problem which arises ofr the independent estimation of regression quantiles for a sequence of percentiles $\tau_1, \ldots, \tau_p$ is the quantile crossing problem. Note that the crossing of two quantile curves corresponding to two distinct percentile levels τ_1 and τ_2 contradicts the monotonicity of the quantile function. For example, consider the median $\tau = 0.5$, and the third quartile $\tau = 0.75$. In the case that the median curve and the third quartile curve cross, i.e. the median curve is higher in comparison to the thrid quartile curve, the following contradiction will arise,

$$0.75 = \text{P}(Y \leq Q(0.75)) \leq \text{P}(Y \leq Q(0.5)) = 0.5.$$

When estimators of quantile curves cross, then they lose their interpretation as quantile curves. The reason why the estimates of the quantile curves cross for different percentile levels comes from the fact that they are estimated independently. Assume for simplicity that the quantile curves $r_\tau(x)$ are linear, i.e. $r_\tau(x) = \alpha(\tau) + \beta(\tau)x$. Consider two distinct percentile levels $\tau_1 > \tau_2$ and the corresponding estimated quantile curves,

$$\hat{Q}_{\tau_1}(x) = \hat{\alpha}_1 + \hat{\beta}_1 x$$
$$\hat{Q}_{\tau_2}(x) = \hat{\alpha}_2 + \hat{\beta}_2 x.$$

According to the monotonicity constrain of the quantile function, the estimated quantile functions must follow the following inequality constrain, $\hat{Q}_{\tau_1}(x) > \hat{Q}_{\tau_2}(x)$. Though, as the parameters of the quantile curves are estimated independently of each other, the slopes of the curves are generally not equal, i.e. $\hat{\beta}_1 \neq \hat{\beta}_2$. In this case, there exists a value $x_1 = \frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\hat{\beta}_2 - \hat{\beta}_1}$ where the curves cross. This means that either for $x > x_1$ or $x < x_1$, $\hat{Q}_{\tau_1}(x) < \hat{Q}_{\tau_2}(x)$ and so in the respective areas the quantile curves lose their natural interpretation.

In the paper of Chernozhukov [4] a monotone rearrangement is proposed to deal with the quantile crossing problem. Denote the estimated quantile levels at a fixed point x^* by $\hat{Q}_1(x^*), \ldots, \hat{Q}_p(x^*)$, corresponding to an increasing percentile sequence $\tau_1, \ldots, \tau_p$. Then in the case that the sequence $\hat{Q}_i(x^*)$ is not monotonically increasing, the monotone rearrangement uses the monotonically reordered estimates as the new quantile estimators, i.e. the order statistics of the estimates, $\hat{Q}_{1,p}(x^*), \ldots, \hat{Q}_{p,p}(x^*$. These reordered estimates have the monotone property of the quantile function and are therefore interpretable as quantiles.

In Figure 2.2 an example of the crossing problem is shown, on the left crossing quantile curves are shown and on the right side the monotone rearrangement of the curves.

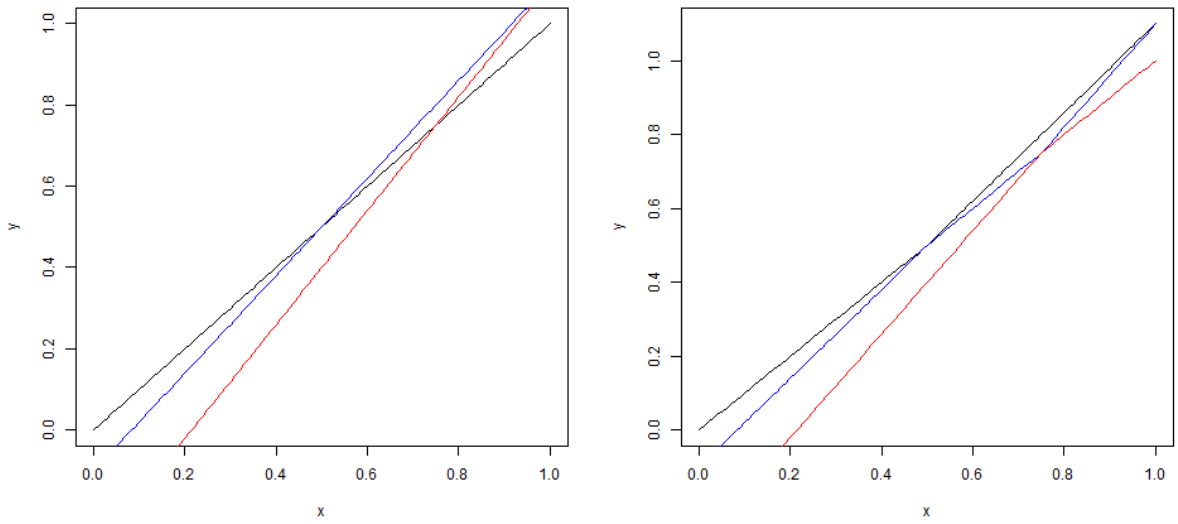

Figure 2.2: Crossing quantile curves (left) and the monotone rearrangement method applied to the quantile curve (right).

In Chapter 4 this problem will be tackled by proposing a common shaped tail model.

2.1.2 Computational aspects of quantile regression

From a practical point of view, the computation of regression quantiles is of significant importance. It will be shown below that the minimization problem as presented in Equation 2.2 can be written as a linear optimization problem. The aim of this section is not to provide the reader with direct optimization methods for solving the quantile regression minimization. Instead, showing that the minimization problem can be written as a linear optimization problem, enables the use of the extensive literature and software on this subject, which is beyond the scope of this thesis.

The quantile regression minimization problem for a linear model is given by,

$$\hat{\beta} = \text{argmin}_\beta \sum_{i=1}^{n} \rho_\tau(Y_i - X_i'\beta),$$

where $X_i' = (1, X_i)$ with X_i the covariate variable. Now define $E_i(\beta) = Y_i - X_i'\beta$ and split it in a positive part and a negative part, hence $U_i(\beta) = \max(E_i, 0)$ and $V_i(\beta) = \max(-E_i, 0)$. Now the check function within the sum becomes,

$$\begin{aligned}\rho_\tau(Y_i - X_i'\beta) = \rho_\tau(E_i(\beta)) &= \max(E_i(\beta), 0)\tau + \max(-E_i(\beta), 0)(1-\tau) \\ &= U_i(\beta)\tau + V_i(\beta)(1-\tau).\end{aligned}$$

Let $\overline{U} = (U_1(\beta), \dots U_n(\beta))$ and $\overline{V} = (V_1(\beta), \dots V_n(\beta))$ and $\beta = \beta^+ - \beta^-$ where β^+ and β^- are the positive and negative part of β respectively. The minimization problem is now given by,

$$\begin{cases} \min c(\overline{U}, \overline{V}, \beta) & = \sum_{i=1}^{n} U_i(\beta)\tau + V_i(\beta)(1-\tau) \\ \text{Subject to} & \\ & Y_i - X_i(\beta^+ - \beta^-) = U_i(\beta) - V_i(\beta) \\ & U_i(\beta) \geq 0 \ \forall 1 \leq i \leq n \\ & V_i(\beta) \geq 0 \ \forall 1 \leq i \leq n \\ & \beta^+ \geq 0 \\ & \beta^- \geq 0. \end{cases}$$

By introducing extra slack variables a_i in the equality constrain, this minimization problem can be written as follows,

$$\begin{cases} \min c(\overline{U}, \overline{V}, \beta) & = \sum_{i=1}^{n} U_i(\beta)\tau + V_i(\beta)(1-\tau) \\ \text{Subject to} & \\ & U_i(\beta) - V_i(\beta) - X_i(\beta^+ - \beta^-) \leq Y_i + a_i \\ & U_i(\beta) \geq 0 \ \forall 1 \leq i \leq n \\ & V_i(\beta) \geq 0 \ \forall 1 \leq i \leq n \\ & \beta^+ \geq 0 \\ & \beta^- \geq 0. \end{cases}$$

Where this can be written as a linear optimization problem in standard form,

$$\begin{cases} \min c^T x & \\ \text{Subject to} & \\ & Ax \leq b \\ & x \geq 0, \end{cases}$$

Where $x = (\bar{(}U), \bar{(}V), \beta^+, \beta^-, a_1, \ldots, a_n)$, A and b are formed such that the linear system $U_i - V_i - X_i'\beta \leq Y_i + a_i$ holds for every $1 \leq i \leq n$ and c such that $c^T x$ is given by $\sum_{i=1}^{n} U_i\tau + V_i(1-\tau)$.

2.2 Extreme value theory

Extreme value theory is the part of statistics that is concerned with a tail of the distribution. Therefore, it enables the estimation of extreme quantiles. The estimation of quantiles from a set of i.i.d. observations $Y_1, \ldots Y_n$ is a relatively simple problem. Consider, the ordered observations $Y_{1,n}, \ldots Y_{n,n}$, with $Y_{1,n}$ the sample minimum and $Y_{n,n}$ the sample maximum. An estimator for $Q(\frac{i}{n+1})$ is given by $Y_{i,n}$. The highest tail quantile, which could be estimated with this approach is then $Q(\frac{n}{n+1})$. Extreme value theory provides the methods, which allow to extrapolate beyond the range of the data and obtain quantile estimates for $\tau > \frac{n}{n+1}$.

In order to be able to extrapolate beyond the range of the data, some extra knowledge about the limiting behaviour of the sample maxima $Y_{n,n}$ is needed when n tends to infinity. Equation 2.3 states in mathematical terms the limiting behaviour and raises the question what is the limiting distribution $G(y)$ of the sample maximum $Y_{n,n}$ when $n \to \infty$. From this point all theory is stated in terms of the sample maxima and the right tail of the distribution, all the theory on the sample minima or left tail can be obtained in a similar way.

$$\lim_{n\to\infty} \mathrm{P}\left(\frac{Y_{n,n} - b_n}{a_n} \leq y\right) = \lim_{n\to\infty} F^n(a_n y + b_n) = G(y) \tag{2.3}$$

Note that in order for G to be a non-degenerate limiting distribution, normalisation with a_n and b_n is necessary, otherwise $Y_{n,n} \xrightarrow{p} x^*$, where $x^* = \sup\{x : F(x) < 1\}$.

The limiting distribution G turns out to be the extreme value distribution, which is proven by Fisher, Tippet [9] and Gnedenko [11];

Theorem 2.2.1. *The class of extreme value distributions is given by*

$$G_\gamma(y) = \exp\left(-(1+\gamma y)^{-\frac{1}{\gamma}}\right), \quad 1 + \gamma y > 0, \tag{2.4}$$

with $\gamma \in \mathbb{R}$ called the extreme value index and for $\gamma = 0$, G_γ is given by $\exp(-e^{-y})$.

The convergence of the normalized maxima to an extreme value distribution does not hold in general; but for most well known distributions it does hold. To make this more precise the concept of domain of attraction is introduced,

Definition 2.2.2. *The distribution function F is said to be in the domain of attraction of an extreme value distribution, denoted by $F \in \mathcal{D}(G_\gamma)$, if there exists normalising constants $a_n > 0$ and $b_n \in \mathbf{R}$ such that,*

$$\lim_{n\to\infty} F^n(a_n y + b_n) = G_\gamma(y) = \exp\left(-(1+\gamma y)^{-1/\gamma}\right),$$

for all y such that $1 + \gamma y > 0$

The distribution of the maxima is now given by a one-parameter family of distributions, which can be subdivided in three distinct classes of tail distributions with their respective properties,

- The case $\gamma > 0$ or the Fréchet domain of attraction: are heavy tails with an infinite endpoint, $y^* = \infty$. Because of the heaviness of the tails moments of $1/\gamma$ do not exist.

- The case $\gamma = 0$ or the Gumbel domain of attraction: are light tails with a finite or infinite endpoint, $y^* \leq \infty$. All moments always exist with these types of tails.
- The case $\gamma < 0$ or the Weibull domain of attraction: are short tails with a finite endpoint, $y^* < \infty$. As a result of the short tails also all the moments exist.

Some examples of distributions, which belong to these three domains are the student-t distirbution with 1 degree of freedom, which belongs to the Fréchet domain of attraction with $\gamma = 1$. The exponential distribution belongs to the Gumbel domain as it has an extreme value index of 0 and the standard uniform distribution belongs to the Weibull domain of attraction with $\gamma = -1$

The following proposition gives a necessary and sufficient condition for the domain of attraction of the Fréchet class,

Proposition 2.2.3. *Let Y be a random variable with distribution function F, then $Y \in \mathcal{D}(G_\gamma)$ with $\gamma > 0$ if and only if,*

$$\lim_{t\to\infty} \frac{1 - F(ty)}{1 - F(t)} = y^{-\frac{1}{\gamma}}.$$

Two distributions are said to be tail equivalent if they share one common extreme value index, i.e.

Definition 2.2.4. *Two random variables Y_1 and Y_2 with distribution functions F_1 and F_2 which are both in the domain of attraction of an extreme value distribution with γ_1 and γ_2 respectively are said to be tail equivalent if $\gamma_1 = \gamma_2$*

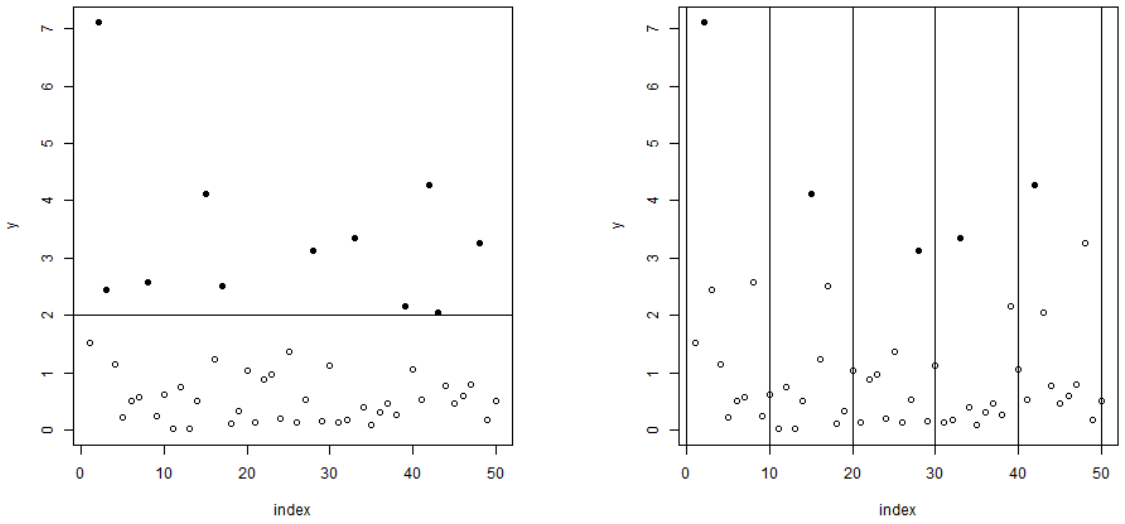

Figure 2.3: The selection of data for extreme value theory, the peaks over threshold method (left) and the block maxima method (right). The data is simualted from a generalized Pareto distribution with $\gamma = 0.1$.

The extreme value distributions as presented in Theorem 2.2.1 are the possible limiting distribution of the sample maxima. In extreme value theory there is a second approach in obtaining a limit distribution for the tail observations. This approach is instead concerned with the behaviour of exceedances above a high threshold. The limiting theorem is due to Pickands, Balkema and de Haan,

Theorem 2.2.5. *Let X be a random variable with distribution function F. Then $F \in \mathcal{D}(G_\gamma)$ if and only if there exists a positive function $f(t)$ such that the following limit relation holds,*

$$\lim_{t \to y^*} P\left(\frac{Y-t}{f(t)} > y \middle| Y > t \right) = \lim_{t \to y^*} \frac{1 - F(t + yf(t)))}{1 - F(t)} = (1+\gamma y)^{-1/\gamma}, \tag{2.5}$$

with $x^* = \sup\{y : F(y) < 1\}$. The distribution $H_\gamma(y) = 1 - (1+\gamma y)^{-1/\gamma}$ is called the generalized Pareto distribution (GPD).

In extreme value theory, as the interest lies in the tail and not in the whole distribution, only the k largest observations are included in the estimation process, with $k << n$. There are two approaches of selecting the "largest" observations for the estimation of the tail behaviour. Let $Y_1, \ldots, Y_n$ i.i.d. random variables. The first approach is called the block maxima approach, which splits the observations $Y_1, \ldots Y_m$ up in k blocks of observations. This is generally done with respect to the time of measurement of the observations, for example blocks of a year or a month. Assume there are k blocks then the maxima of these blocks are denoted by, $M_1, \ldots, M_k$. From Theorem 2.2.1 it is clear that these converge to an extreme value distribution when $\frac{n}{k}$ tends to infinity. The second approach considers all the upper order statistics exceeding a high threshold d, i.e. $Y_{(n-m+1)} - d, \ldots, Y_{(n)} - d$. From Theorem 2.2.5 these exceedances will asymptotically follow the generalized Pareto distribution. See Figure 2.3 for both methods.

The second approach is preferred in this thesis as the extension to the tail is most natural by modelling it as the exceedances above a high threshold, which is referred to as the peaks over threshold method.

Two estimators of the extreme value index will be presented below; the first is called the Hill estimator, [12]. This is the first estimator which appeared in the literature for the extreme value index. It is however restricted to positive values of γ. This means that it estimates the extreme value index for heavy tailed distributions, but not for light tailed or short tailed distributions.

Definition 2.2.6. *The estimator $\hat{\gamma}_H(k)$ is called the Hill estimator and is defined by*

$$\hat{\gamma}_H(k) = \frac{1}{k} \sum_{i=1}^{k} \log Y_{n-i+1,n} - \log Y_{n-k,n} \tag{2.6}$$

Where $Y_{i,n}$ is the ith order statistic of an iid random sample.

The second estimator is the maximum likelihood estimator, which maximizes the likelihood of the parameters of the generalized Pareto distribution over the exceedances above a high threshold, [8]. This estimator is proven to be consistent for the extreme value index in the range of $[-1, \infty)$, i.e. it is consistent for heavy tailed, light tailed and a range of short tailed distributions.

Definition 2.2.7. *The maximum likelihood estimators for the scale and shape are defined as the values $\hat{\gamma}(k), \hat{\sigma}(k)$ which maximize the following likelihood equations,*

$$\begin{cases} \frac{1}{k} \sum_{i=1}^{k} \log\left(1 + \frac{\gamma}{\sigma}(Y_{n-i-1} - Y_{n-k,n})\right) = \gamma \\ \frac{1}{k} \sum_{i=1}^{k} \frac{1}{1+\frac{\gamma}{\sigma}(Y_{n-i-1} - Y_{n-k,n})} = \frac{1}{\gamma+1} \end{cases}$$

Note however that this estimator is based on the assumption that exceedances above $Y_{n-k,n}$ follow a generalized Pareto distribution. The equations are then derived from the likelihood of these exceedances assuming a generalized Pareto distribution. The choice of k is of high importance here as the approximation of the GPD requires the threshold, i.e. $Y_{n-k,n}$, to be high.

2.3 Non-parametric estimators

Models within statistics can roughly be divided into two classes, the so called parametric models and the non-parametric models. The main difference between the two comes from the assumptions which are made before the estimation procedure. A parametric model assumes a specific parametric structure on the data and a non-parametric model does not assume such a structure. For example in the estimation of the distribution a parametric estimator assumes that the data $Y_1, \ldots, Y_n$ are from a distributional family smoothly parametrized by $F = F_\theta$ for some finite dimensional parameter θ. Consequently in order to estimate the distribution it is only required to estimate the parameter θ.

These estimators are generally quite good when the parametric assumption holds, though the problems with these parametric estimators arise when the assumption does not hold. As a result, there can be a lot of bias in the estimation resulting from a bad assumption. Consider for example to use an estimator for the normal distribution on data which is from a bimodal distribution. The left plot in Figure 2.4 shows the exact bimodal distribution together with the estimated normal density based on a dataset of $n = 500$. It is clear that the bad assumption of normality leads to a large bias.

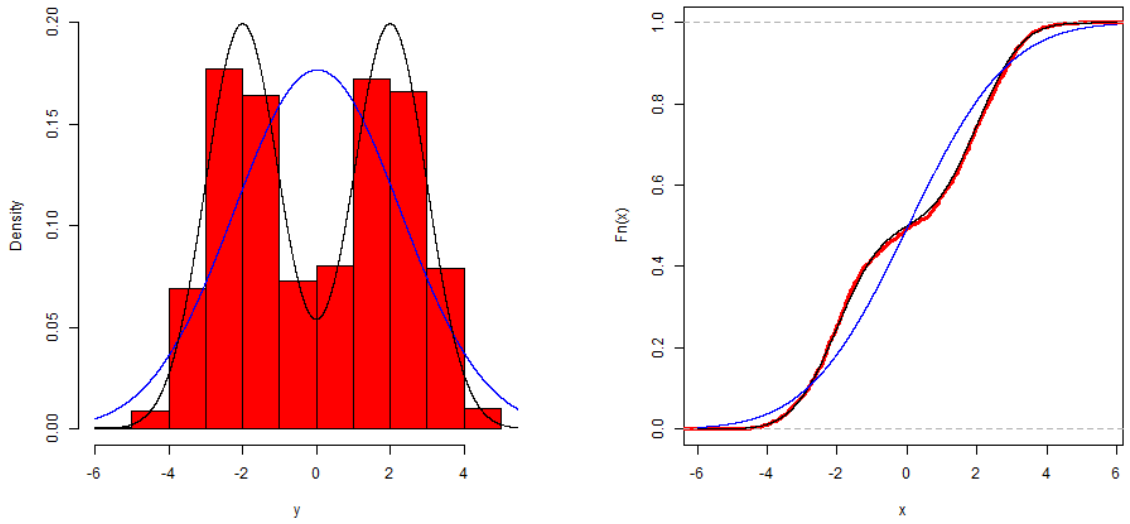

Figure 2.4: Comparison of non-parametric estimation (in red) and parametric estimation based on normality assumption (in blue). Shown for both the density (left) and the distribution function (right).

This motivates the introduction of non-parametric models, i.e. models which do not have a parametric assumption on the data. Therefore they obtain the information in a more data driven way. Staying with the same example as above, a well known non-parametric estimator for the estimation of the distribution function is the empirical cumulative distribution function which is defined by,

$$\overline{F}_n(y) = \sum_{i=1}^{n} \mathbf{1}_{(y \leq Y_i)}$$

This estimator looks very simple the probability that $Y \leq y$ is estimated by the relative frequency of occurrence, but is in fact very useful and well studied. In the next chapter this

estimator will serve as a basis for an extreme quantile estimator. The right plot in Figure 2.4 shows the bimodal cumulative distribution function together with the empirical cumulative distribution function. It is clear that this estimator is significantly better than the estimator based on the normality assumption.

In the quantile regression setting as introduced above, often it is assumed that the quantile curves are linear in the covariate, i.e. the relation between quantiles of Y and X is linear for all values of τ. The quantile curve can then be parametrized by $r_\tau(x) = \alpha(\tau) + \beta(\tau)x$. The regression quantile estimator is then obtained estimating $(\alpha(\tau), \beta(\tau))$. This would introduce large biases when the quantile curves are in fact not linear.

A non-parametric regression estimator models the dependence relation between Y and X in a local way. Consider a fixed covariate level x^* and let $h > 0$ be a bandwidth. Non-parametric quantile regression estimates $f_\tau(x^*)$ by approximating the function in the neighbourhood of x^*, $[x^* - h, x^* + h]$, by a finite Taylor expansion. Let $m > 0$ be the order of the Taylor expansion, then the quantile function can be approximated in a neighbourhood of x^* by,

$$Q(\tau|X = x) \approx \sum_{i=0}^{m} \left. \frac{\mathrm{d}^i Q(\tau|X = x)}{\mathrm{d}x^i} \right|_{x=x^*} \frac{(x - x^*)^i}{i!} = \sum_{i=0}^{m} \beta_i (x - x^*)^i,$$

where $\beta_i = \left. \frac{\mathrm{d}^i Q(\tau|X=x)}{\mathrm{d}x^i} \right|_{x=x^*} \frac{1}{i!}$. Note that this does not assume any parametric structure on the quantile curves. The approximation becomes more accurate when m increases. Taking $m \to \infty$ is however impossible as the number of parameters to estimate will go to infinity. In practice the approximating function is chosen either local constant $m = 0$, local linear $m = 1$ or local quadratic $m = 2$. In addition note that $\beta_0 = Q(\tau|X = x^*)$, i.e. β_0 is the point estimate of $\hat{f}_\tau(x^*)$.

The parameters in this local Taylor expansion can be estimated by solving the minimization given in Equation 2.2, rewriting the minimization problem for the current setting,

$$\underset{\beta_0,\ldots,\beta_m}{\arg\min} \sum_{i=1}^{n} \mathbf{1}(|X_i - x^*| < h) \rho_\tau \left(Y_i - \sum_{j=0}^{m} \beta_j (X_i - x^*)^j \right).$$

Here the choice of h is a non-trivial and very important problem. The same bias variance trade-off needs to be made as with the choice of k in extreme value theory. For small h the local approximation is expected to hold quite well, which has a small bias as a resutl. However a small h will limit the number of observations used for the estimation and will therefore blow up the variance.

In the above minimization, every observation in the neighbourhood of x^*, $[x^* - h, x^* + h]$, has equal weight in the minimization problem. Observations which lie further from x^* could potentially introduce more bias, as the errors in the Taylor expansion are larger. Therefore it would be better to put more weight on observations (Y_i, X_i) for which X_i is closer to x^*. Let $w_1(x^*), \ldots, w_n(x^*)$ denote a sequence of weights which sum up to 1. The minimization problem can now be written as,

$$\underset{\beta_0,\ldots,\beta_m}{\arg\min} \sum_{i=1}^{n} w_i(x^*) \rho_\tau \left(Y_i - \sum_{j=0}^{m} \beta_j (X_i - x^*)^j \right).$$

Often a kernel function, $K(.)$ is used for these weights, where the kernel is defined as a probability density function on $\mathbf{R}$. Generally the choice is made for a uni-modal symmetric

distribution around 0 to put more weight on the observations close to x^* and less weight on observations far from x^*. A weight for the non-parametric quantile estimator at x^* with bandwidth h is defined by,

$$w_i(x^*) = K\left(\frac{x^* - X_i}{h}\right).$$

Chapter 3

Extreme quantile estimators in the literature

The conditional distribution function of Y depending on X is denoted as $F(y|X = x) = \mathrm{P}(Y \leq y|X = x)$. The focus in this thesis will be on modelling the tail of $F(y|X = x)$, which can be denoted by $\mathrm{P}(Y \leq c|X = x)$ for c large. There are already a variety of approaches for the tail estimation of the conditional distribution. This chapter introduces four of these methods, each method represents one of four approaches to this specific estimation problem.

Obtaining inference about the tail of the conditional distribution is in general a hard problem. Tail observations are sparse by definition and additionally the tail distribution may depend on the covariate information X. Therefore, instead of the estimation of a single extreme value index, a function, $\gamma(x)$, is to be estimated.

Clearly, this imposes quite a challenge as direct extrapolation as is done in univariate extreme value theory is not possible, due to the lack of observations for the specific covariate levels x^*. Instead, estimators are obtained in a two step procedure. To illustrate the procedure, the tail distribution can be split into two parts.

1. Intermediate quantiles, $Q(\tau_n)$ s.t. $n(1 - \tau_n) \to \infty$ when $n \to \infty$

2. Extreme quantiles, $Q(\tau_n)$ s.t $n(1 - \tau_n) \to c^*$ when $n \to \infty$ with c^* a constant

The intermediate quantiles can be interpreted as tail quantiles, which are in the range of the data, i.e. q is an intermediate quantile if $P(Y > q|X = x) > \frac{1}{n}$. The extreme quantiles can be interpreted as the tail quantiles which exceed the range of the data, i.e. q is an extreme quantile if $P(Y > q|X = x) < \frac{1}{n}$.

The first step obtains an estimator for the intermediate part of the tail distribution. The second step uses the estimator for the intermediate quantiles for a given $X = x^*$ and extrapolates them to the extreme quantiles using methods from univariate extreme value theory.

Roughly, all estimators in the literature can be categorized in two groups. For the first group, inference about the intermediate quantiles is obtained from the conditional distribution function. The second group, on the other hand, obtains inference about the intermediate quantiles via quantile regression. For each of the two groups both a parametric estimator and a non-parametric estimator will be discussed.

The following section will introduce the general structure of the estimators and clarify the choices that have to be made in the estimation process.

3.1 General estimation structure

The structure of every estimator in the literature is similar, as also described above. This section describes a general structure of the two step approach. All methods will be compared in this general structure.

First, let $\tau_1, \ldots, \tau_p$ denote a sequence of intermediate percentiles and let the corresponding quantile curves be denoted by $Q(\tau_1|X = x), \ldots, Q(\tau_p|X = x)$. Note that the quantile function is the inverse of the distribution function and therefore estimators of the conditional quantile function can be obtained by taking the left continuous inverse of the estimated conditional distribution function, i.e. $\hat{Q}(\tau|x) = \hat{F}^{\leftarrow}(\tau|x) = \inf\{t : \hat{F}(t|x) \geq \tau\}$.

In the second step the intermediate quantiles are extrapolated using methods from extreme value theory. Note that quantile curves evaluated at a fixed covariate level, $X = x_0$, are an increasing sequence of p distinct values assuming a continuous distribution, i.e. $\hat{Q}(\tau_1|x_0), \ldots, \hat{Q}(\tau_p|x_0)$. This sequence of estimators can be considered as ordered observations from the distribution $Y|X = x$.

For the estimation of the extreme quantiles of $Y|X = x_0$, the extreme value index needs to be estimated. Now let $\hat{\gamma}(Y_{n-k+1,n}, \ldots, Y_{n,n})$ be an estimator for the extreme value index based on the k upper order statistics. By substituting the upper order statistics by the estimated quantile curves for a fixed point x_0 the following estimator is obtained, $\tilde{\gamma}(x_0) = \hat{\gamma}(\hat{Q}(\tau_{p-k}|x_0), \ldots, \hat{Q}(\tau_p|x_0))$ for $k < p$.

The extreme quantile estimator for $Y|X = x_0$ is now given by the following formula, with τ_n an extreme percentile,

$$\hat{Q}(\tau_n|x_0) = \left(\frac{1 - \tau_{p-k}}{1 - \tau_n}\right)^{\tilde{\gamma}(x_0)} \hat{Q}(\tau_{p-k}|x_0).$$

3.2 Intermediate quantiles via quantile regression

This section discusses two estimators where inference about the intermediate quantiles is obtained via quantile regression. First, a parametric estimator, proposed in [19], where the quantile curves are assumed to be linear is discussed together with an extension of these results, proposed in [18]. In Section 3.2.2 a non-parametric estimator is discussed, which was proposed in [2].

3.2.1 Linear quantile regression with power-transformations

In [19] a parametric estimator is proposed, which assumes the linear quantile curves in the tail of the conditional distribution $Y|X$. More formally define a high percentile level τ_c close to 1, the quantile curves of $Y|X$ are assumed to be,

$$Q(\tau|x) = \alpha(\tau) + x\beta(\tau) \text{ for } \tau \in [\tau_c, 1] \tag{3.1}$$

Note that this assumption does not pose any restrictions on the of the distribution corresponding to the percentiles in $[0, \tau_c)$. Three motivations for the development of this estimator are given. As a first motivation the authors argue that the assumption of linear quantile curves is not very restrictive in practice. Consequently, the estimator has in these practical examples better convergence rates and a higher efficiency in comparison to non-parametric methods, two of which are introduced in the following sections. In addition the threshold in this model, i.e. a high conditional linear quantile curve, is covariate dependent, which is an improvement with respect to methods in the extreme value literature, which model the tail by considering exceedances above a constant threshold. Finally, the proposed method allows for a varying extreme value

index over the covariate levels, i.e. $\gamma = \gamma(x)$, without assuming a parametric link-function, in contrast to standard approaches in extreme value literature where this is necessary. An example of such a method is discussed in Section 3.3.

To see how this estimator fits in the general framework, the sequence of percentile levels is defined as follows, $\tau_c = \tau_1 \leq \cdots \leq \tau_p$, where τ_c is a high percentile close to 1. For each of the percentiles an estimate of the corresponding intermediate quantile curve can be obtained by solving the following quantile regression problem,

$$(\hat{\alpha}, \hat{\beta}) = \text{argmin}_{\alpha,\beta} \sum_{i=1}^{n} \rho_\tau(Y_i - \alpha - \beta x_i).$$

To make this more intuitive a dataset is simulated from $Y = 1.2\exp(2.2x) + \epsilon$, where x is a randomly sampled from a uniform distribution on the interval $[-1, 1]$ and $\epsilon \sim GPD(\gamma = 0.25, \sigma = 1)$. The dataset has a size of 500 observations. The left panel of Figure 3.1 shows the observations and the true quantile curves corresponding to percentile levels $0.8, 0.85, 0.90$ and 0.95. The right panel shows the estimated linear quantile curves.

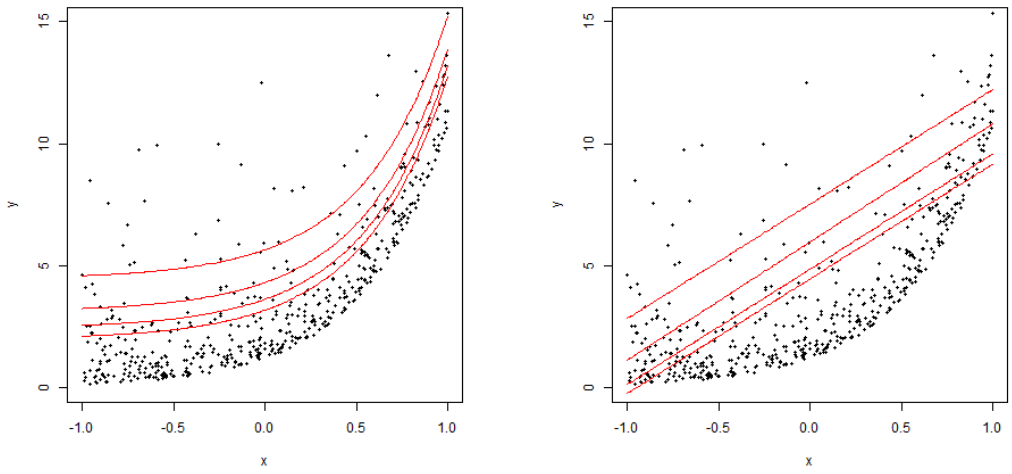

Figure 3.1: Simulated dataset of 500 observations with the exact quantile curves (left) and the estimated linear quantile curves (right). The quantile curves correspond to the percentiles $0.8, 0.85, 0.9$ and 0.95.

It is obvious that in this case the assumption of linearity of the quantile curves does not hold. However, the quantile curves are given for each of the covariate values X. This enables the use of the quantile estimates at a specific covariate level x^* to extrapolate using univariate extreme value theory on the sequence $\hat{Q}(\tau_1|X = x^*), \ldots, \hat{Q}(\tau_p|X = x^*)$. Therefore, by modelling the intermediate quantiles parametrically, the extrapolation can be done for each x^*.

Before discussing the extrapolation to the tails, first an improvement can be made on the linear quantiles by introducing power transforms. These power transforms, or Box-Cox transforms, are introduced in a follow-up paper [18].

The Box-Cox transformation was first introduced in [3] as a power transformation which maintains monotonicity of the transformed variable for every power λ. The Box-Cox power

transformation is given by,

$$\Lambda_\lambda(y) = \begin{cases} \frac{y^\lambda - x}{\lambda} & \text{if } \lambda \neq 0 \\ \log(y) & \text{if } \lambda = 0. \end{cases}$$

The power transformation is used to relax the assumption of the linear quantile curves of $Y|X = x$. Instead it is assumed there exists a power λ, such that $\Lambda_\lambda(Y)|X = x$ has linear quantile curves.

In the estimation structure as described in Section 3.1, the power transform adds an extra step, i.e. the estimation of the power λ. The power is obtained by the minimization problem stated below; for more information the reader is referred to [15] and [20].

$$\hat{\lambda} = \arg\min_{\lambda \in \mathbf{R}} \sum_{i=1}^{n} R_n(X_i, \lambda|\tau_c)^2,$$

where

$$R_n(x, \lambda|\tau) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_j \leq x} (\tau - \mathbf{1}_{\Lambda_\lambda(y_i) - X_i \hat{\beta}(\tau|\lambda) \leq 0})$$

$$\hat{\beta}(\tau|\lambda) = \arg\min_{b \in \mathbf{R}} \sum_{i=1}^{n} \rho_\tau(\Lambda_\lambda(Y_i) - X_i b).$$

Here $\hat{\beta}(\tau|\lambda)$ are the parameters of the regression quantile for the power transformed data with parameter λ. The function R_n is a measure how well the power transformed line splits the observations in τ portion below and $1-\tau$ portion above the curve for an interval $(-\infty, x]$. Then the parameter λ is the parameter that optimizes this for all intervals $(-\infty, X_i]$, $1 \leq i \leq n$.

The observations Y_i are now transformed to $Z_i = \Lambda_{\hat{\lambda}}(Y_i)$ and the linear quantile curves are estimated with $(Z_1, X_1), \ldots, (Z_n, X_n)$ resulting in $\hat{Q}_Z(\tau_1|X = x), \ldots, \hat{Q}_Z(\tau_p|X = x)$. These can thereafter be transformed back by the inverse power transformation, hence

$$\hat{Q}_Y(\tau|X = x) = \Lambda_{\hat{\lambda}}^{-1}(\hat{Q}_Z(\tau|X = x)).$$

The use of the Box-Cox power transform leads to an improved estimate of the intermediate quantiles in the simulated data example. In Figure 3.2 the left panel shows again the data with the true quantile curves and the right panel shows the estimated quantile curves fitted on the power transformed data. The estimated power transform is here given by $\hat{\lambda} = -0.7$.

An important observation from this plot is that the estimated quantiles cross around $x = 1$. This crossing first happens when the regression quantiles are estimated on the power transformed data. Transforming the regression quantiles back, will again have this crossing, as is visible in the figure, because the power transformation is monotone. This is noted in [19] and it is suggested to use the monotone rearrangements proposed in [4].

The next step in the estimation process is to extrapolate the estimates for the intermediate quantile curves to the tail. This is done in exactly the same way as described in the general structure. For a fixed value x_0 for the covariate, the estimates $\hat{Q}_i = \hat{Q}(\tau_i|X = x_0)$ are considered to be observations from the intermediate part of the tail distribution of $Y|X = x_0$. The extreme value index estimator which is used is the Hill estimator, which gives the following estimate of the extreme value index,

$$\hat{\gamma}(x_0) = \frac{1}{k} \sum_{i=1}^{k} \log \frac{\hat{Q}_{n-i+1}}{\hat{Q}_{n-k}}.$$

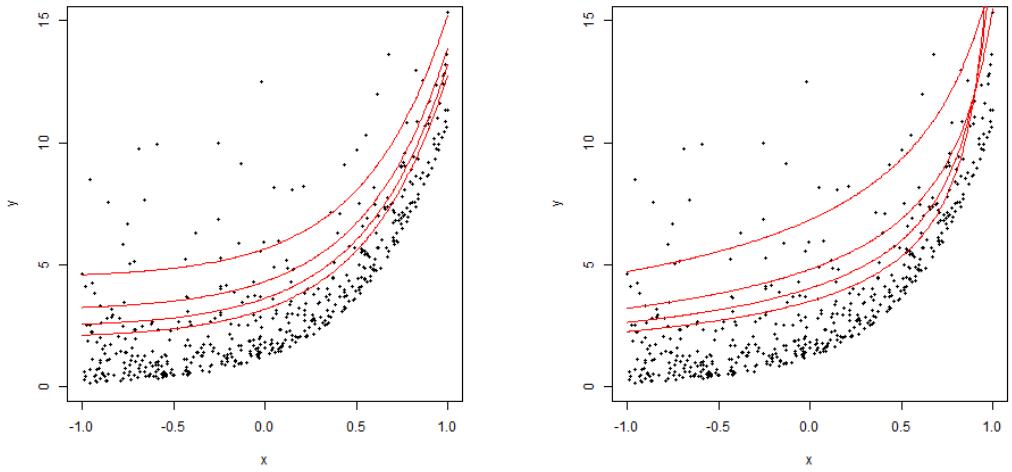

Figure 3.2: Simulated dataset of 500 observations with the exact quantile curves (left) and the power trasformed linear quantile estimator (right). The quantile curves correspond to the percentiles $0.8, 0.85, 0.9$ and 0.95 and power transform $\lambda = -0.7$

Furthermore the extreme quantile estimator for an extreme percentile τ_n is given by,

$$\hat{Q}(\tau_n | X = x_0) = \left(\frac{1 - \tau_{p-k}}{1 - \tau_n} \right)^{\hat{\gamma}(x_0)} \hat{Q}_{p-k}. \tag{3.2}$$

Before the full estimator, can be applied to our simulated data example some choices have to be made. Both the sequence of percentiles, $\tau_1, \ldots, \tau_p$, and the number k need to be determined. Note that these choices are closely related to each other and are in fact one choice. Consider first the sequence $\frac{1}{n+1}, \ldots, \frac{n}{n+1}$ and note that no extra information is gained by considering more percentiles. To see this recall that a quantile curve corresponding to percentile τ splits the observations in a τ and a $1 - \tau$ portion below and above respectively. For example the quantile curve corresponding to $\frac{2}{n+1}$ splits the data $1 : n - 1$ below and above the quantile curve and $\frac{3}{n+1}$ splits the data $2 : n - 2$ below and above the curve. The quantile curve corresponding to $\frac{2.5}{n+1}$ does not add extra information as it either splits the data $1 : n-1$ or $2 : n-2$ which makes it equivalent to the $\frac{2}{n+1}$ or $\frac{3}{n+1}$ quantile. The above percentile sequence is therefore the largest percentile sequence of interest.

Therefore by fixing k, the largest percentile sequence of interest is given by, $\frac{n-k}{n+1}, \ldots, \frac{n}{n+1}$.

The choice of k is generally considered a hard problem in extreme value theory and is in practice often obtained using heuristics. Note that if k is really small the number of observations used for the estimation of the extreme value index is small which will lead to a large variance. On the other hand when k is large the estimator is not solely based on tail observations, which will introduce bias in the estimate. To choose the right k in this bias-variance trade-off the estimate $\hat{\gamma}$ is often plotted as a function of k. The k where this function becomes stable is then chosen.

In Figure 3.3 the left side shows the function $\gamma(x)$, where for fixed x the number of upper order statistics is chosen to be $k = 37$. The horizontal dotted line indicates the true value of the extreme value index. It can be seen that the extreme value index is not estimated to be a non-constant function of x, where it should be constant. Extrapolation to the tail can now be done using the quantile estimator in Equation 3.2. Here $\hat{Q}_{n-k}$ is given by the quantile curve corresponding to the percentile $1 - \frac{k}{n} = 0.926$. For τ_n equal to $0.95, 0.99$ and 0.995 the curves

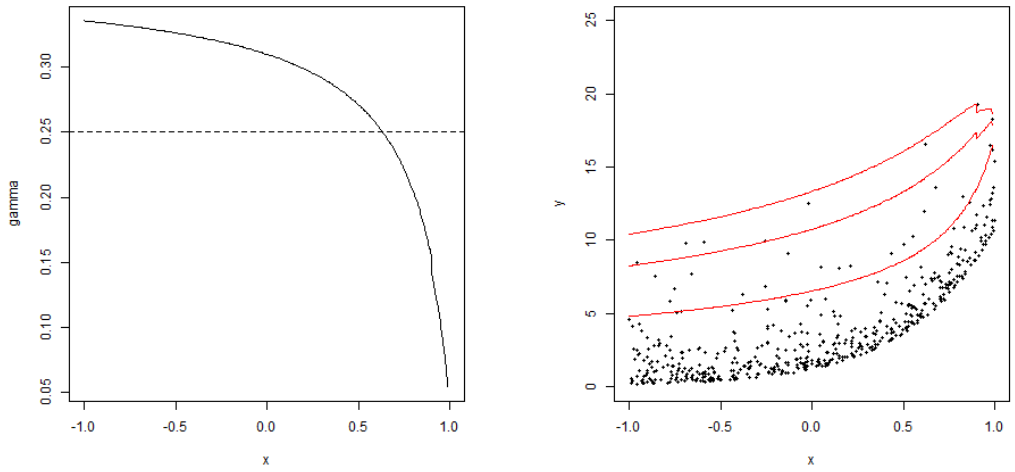

Figure 3.3: The estimated extreme value index as a function of x (left) and the extrapolated quantile curves (right) for the percentiles $0.95, 0.99$ and 0.995.

are plotted on the right side of Figure 3.3. Note that there are little jumps on the right panel of the plot. These are caused by the monotone rearrangements applied to the crossings which appear in the upper right corner in the estimated quantile curves of Figure 3.2.

3.2.2 Non-parametric extreme quantile regression

In [2], a non-parametric approach for the estimation of the conditional tail distribution is proposed. As explained in Chapter 2, non-parametric quantile regression approximates the quantile function in a neighbourhood of a fixed covariate x^* by a finite Taylor expansion of order m around x^* hence,

$$Q(\tau|x) \approx \sum_{i=0}^{m} \frac{1}{i!} \frac{\mathrm{d}^i Q(\tau|x)}{\mathrm{d}x^i} (x - x^*)^i = \sum_{i=0}^{m} \beta_i (x - x^*)^i.$$

The τ quantile at x^* is then found by solving the corresponding regression problem,

$$\arg\min_{\beta \in \mathbf{R}^{m+1}} \sum_{i=1}^{n} \mathbf{1}_{[x^*-h, x^*+h]}(x_i) \rho_\tau \left(Y_i - \sum_{j=0}^{m} \beta_j (x_i - x^*)^j \right), \tag{3.3}$$

where h is the bandwidth and $\rho_\tau(u)$ is defined in Equation 2.1.

Furthermore, note that the τ quantile estimator for a fixed point x^* is given by the intercept $\hat{\beta}_0$. The entire quantile curve can now be obtained by discretizing the range of x and creating a grid of q points, i.e $x_1^*, \ldots, x_q^*$. Performing the minimization in Equation 3.3 for every covariate level, $x_j^*, 1 \leq j \leq q$ obtains a discrete representation of the quantile curve, $\hat{\beta}_0^1, \ldots, \hat{\beta}_0^m$.

Before these quantile curves can be estimated on the simulated data example a choice for the grid on the covariate has to be made. Therefore, choose $q = 1000$ and let all covariate levels be equally spaced on the range of x, i.e. $[-1, 1]$. Let $h = 0.4$ be the bandwidth. In Figure 3.4 the non-parametric quantile curves are plotted for percentiles $0.8, 0.9$ and 0.95.

The influence of data sparseness is immediately visible from the Figure 3.4. The quantile curve corresponding to $\tau = 0.8$ looks good and resembles nicely the exact quantile curve. The quantile curves for $\tau = 0.9$ and $\tau = 0.95$ on the other hand are a lot more erratic.

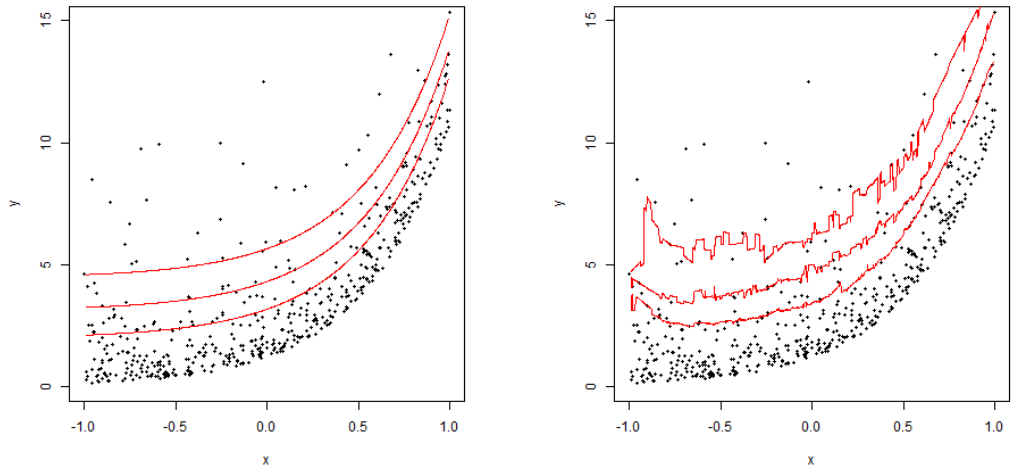

Figure 3.4: The exact quantile curves (left) and the estimated quantile curves (right) using the method of non-parametric quantile estimation. The curve correspond to the percentiles $\tau \in (0.8, 0.9, 0.95)$ and the bandwidth used is $h = 0.4$.

Let x^* be a fixed covariate level and define $\tau_1, \ldots, \tau_p$ be a percentile sequence. Remember that the number of observations used to estimate a quantile is in the non-parametric setting not equal to n, but instead to the number of observations in the windows around x^*, $n^* = \sum_{i=1}^{n} \mathbf{1}(X_i \in (x^* - h, x^* + h)$. The possible percentile sequence for the estimation of the intermediate quantiles is then given by; $\frac{1}{n^*+1}, \ldots, \frac{n^*}{n^*+1}$. In addition, the bandwidth h needs to be determined, for which the reader is referred to section 4 of [2] where a bandwidth selection method is proposed.

The intermediate quantile estimates for the above sequence are now given by $\hat{Q}\left(\frac{1}{n^*+1}|X = x^*\right), \ldots, \hat{Q}\left(\frac{n^*}{n^*+1}|X = x^*\right)$. Notice that these quantile estimates are not restricted to the tail, in which the interest lies. Therefore a number k of upper order statistics still needs to be chosen to obtain a good estimate of the extreme value index. The estimator which will be used for the extreme value index is the Hill estimator. For the data example the extreme value index as a function of x is shown on the left side of Figure 3.5.

The last step in the extreme quantile estimation is the extrapolation to the extreme quantiles. The extreme quantile estimator is now given by,

$$\hat{Q}(\tau_n|X = x^*) = \hat{Q}\left(\frac{n^* - k}{n^* + 1}\middle| X = x^*\right)\left(\frac{\frac{k}{n^*+1}}{1 - \tau_n}\right)^{\hat{\gamma}}. \tag{3.4}$$

In the right plot of Figure 3.5 the extreme quantile curves are shown for percentiles $0.95, 0.99$ and 0.995. It can be seen that the high quantile curves are really erratic. This is a result from the fact that for every point n^* depends on x^* and therefore $\frac{n^*-k}{n^*+1}$ is a different percentile for each x^*. As a result, the extreme quantile estimates give information locally rather than globally.

3.3 Estimators using distribution function

In this section, the focus will be on two methods where the intermediate quantiles are estimated from the perspective of the distribution function. Similar to the previous section, first a parametric estimator will be discussed and then a non-parametric estimator. A direct result from the usage of the distribution function to obtain inference about the intermediate quantiles is

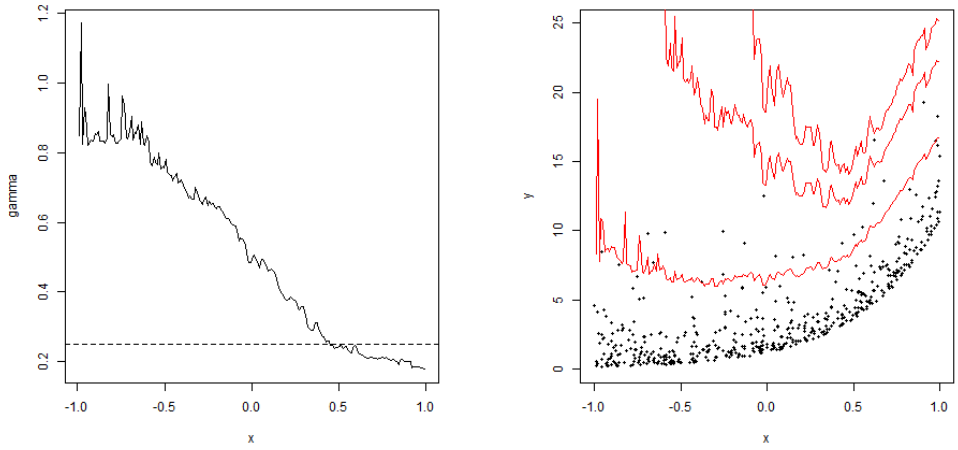

Figure 3.5: The estimated extreme value index as a function of x (left) for the non-parametric extreme quantile estimation. The extrapolated quantile curves (right) for the percentiles $0.95, 0.99, 0.995$.

that the problem of quantile crossing does not appear. This will be immediately clear from the structure of the following two methods.

3.3.1 Parametric estimation of a varying tail distribution

Parametric estimators for a distribution are quite common in statistics. As explained in Chapter 2, a parametric estimator uses a distributional assumption on the data, i.e. suppose the data is distributed according to distribution function $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$. The estimation of θ can now be obtained by maximum likelihood estimation, moment estimators or other estimators.

Our setting differs from these parametric model in two ways. First, the interest does not lie in the modelling of the entire distribution, but instead in the modelling of the tail distribution. Therefore a distributional assumption is only concerned with the tail of the distribution. Second, the parameters are functions of covariates, because the tail of the distribution is assumed to be covariate dependent.

Define a constant c to be a high threshold and let $Z_i = Y_i - c$ with $Y_i > c$, such that Z_i are the exceedances above the threshold c. The distributional assumption on the tail distribution is now given by Theorem 2.2.5, which states that the these exceedances approximately follow the generalized Pareto distribution,

$$\mathrm{P}(Y_c \leq y | Y > c) \overset{c\to\infty}{\to} F(y|\sigma,\gamma) = 1 - \left(1 + \gamma\frac{y}{\sigma}\right)^{-1/\gamma} \quad y > 0.$$

The second difference with the simple example, the dependence on the covariates, can now easily be added by assuming a parametric function for the coefficients, $\sigma = \sigma(x)$ and $\gamma = \gamma(x)$. Note that $\sigma > 0$, therefore the exponential link function is chosen, $\sigma(x) = \exp(x\beta)$. For γ the options are wider, when $\gamma > 0$ is assumed the same parametric form can be chosen, otherwise a linear parametric form could be applied, $\gamma(x) = x\beta$

In [7] an estimator is proposed for this model which is obtained by maximum likelihood estimation. The distribution function of the covariate dependent tail distribution is given by,

$$P(Y - x \leq z | X = x) = 1 - \left(1 + \gamma(x)\frac{z}{\sigma(x)}\right)^{-1/\gamma(x)} \quad z > 0.$$

Note that this method is different from the previous methods as the modelling of the intermediate quantiles is very limited, i.e. it is taken as a constant threshold, and the extrapolation process is more elaborate by estimating the tail distribution for all values of the covariates, instead of obtaining pointwise estimators.

3.3.2 Non-parametric kernel curves

Based on a univariate dataset $Y_1, \ldots, Y_n$ the ECDF given by $F_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(Y_i \leq y)$ is the most fundamental non-parametric estimator for the distribution function. In [6] and [5], an extreme quantile curve estimator is proposed, which uses a smoothed version of the ECDF to obtain inference about the intermediate quantiles of the conditional distribution $Y|X = x$.

The estimator is given as follows,

$$\hat{F}_n(y|X = x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\mathbf{1}(Y_i \leq y)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)}, \tag{3.5}$$

where $K(.)$ is a probability density function on $\mathbf{R}$. Note that by substituting $w_i(x) = \frac{K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)}$, the estimator can be seen as a weighted ECDF where the weights are with respect to the covariates.

$$\hat{F}_n(y|X = x) = \sum_{i=1}^{n} w_i(x)\mathbf{1}(Y_i \geq y) \tag{3.6}$$

An estimator for the quantile function can now be obtained by inverting the estimator, i.e.

$$\hat{Q}(\tau|X = x) = \hat{F}_n^{\leftarrow}(\tau|X = x) = \inf(t : \hat{F}_n(\tau|X = x) \leq \tau).$$

For a fixed percentile τ_0, $\hat{Q}(\tau_0|X = x)$ is a continuous function in x, because $\hat{F}_n(y|X = x)$ is continuous in x as a result of the use of the kernels. On the other hand, when considering the quantile function for a fixed $x = x^*$, it becomes a step-function with respect to τ where the steps are equal to the weights, $w_1, \ldots, w_n$. In Figure 3.6, this is demonstrated for the simulated example.

The higher quantile curves become more erratic, which motivates the use of extreme value theory to extrapolate to the extreme quantiles. Again a choice needs to be made for the percentile sequence $\tau_1, \ldots, \tau_p$. No guidelines are given for the construction of this sequence in [6]. For $h = 0.3$ there are approximately 30 observations in the interval $[x^* - h, x^* + h]$ exceeding the 0.8 quantile. This amount is random and can therefore be slightly more or less therefore the choice is made to consider 20 equally spaced percentiles between 0.8 and 0.99 for the data example. Note that when more data is available the largest percentile can be even higher. Then for the selection of k, a stable period is determined of the function $\gamma(k)$, with k the number of upper order statistics. The Hill estimator is used as an estimator for the extreme value index.

The estimate of the extreme value index as a function of x is shown on the left side of Figure 3.7, where the number of upper order statistics is chosen to be $k = 10$. Now for the sequence, $\tau_1, \ldots, \tau_p$, the corresponding quantile curve estimates are denoted with $\hat{Q}(\tau_1|X = x), \ldots, \hat{Q}(\tau_p|X = x)$. The extreme quantile estimator is now given by,

$$\hat{Q}(\tau_n|X = x_0) = \left(\frac{1-\tau_{p-k}}{1-\tau_n}\right)^{\hat{\gamma}(x_0)} \hat{Q}(\tau_{p-k}|X = x). \tag{3.7}$$

Similar to Figure 3.5, for the non-parametric extreme quantile estimation are the estimated quantiles are very erratic, and though the true value is constant the extreme value index varies heavily over the covariates.

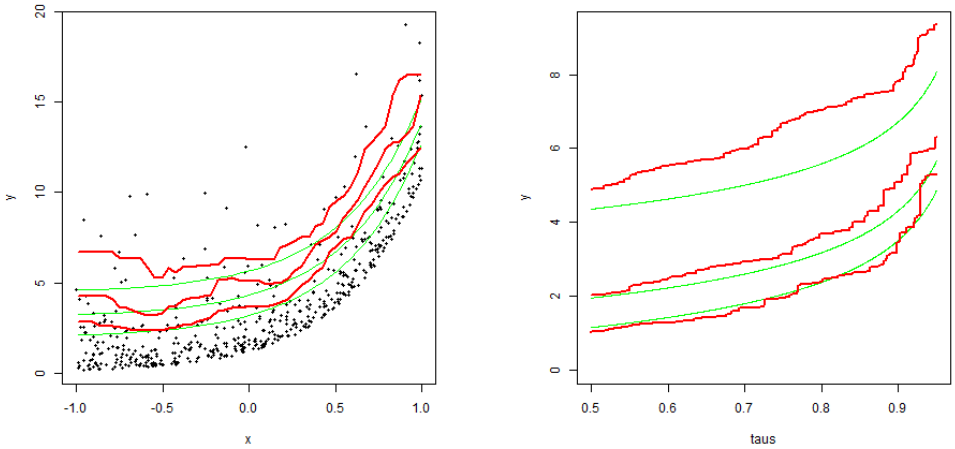

Figure 3.6: The estimated quantile curves in red and the exact quantile curves in green (left) for percentiles $0.8, 0.9, 0.95$ and estimated with the kernel estimator. The estimated distribution function in red and the exact distribution function in green (right) for the covariate values $-0.5, 0$ and 0.5.

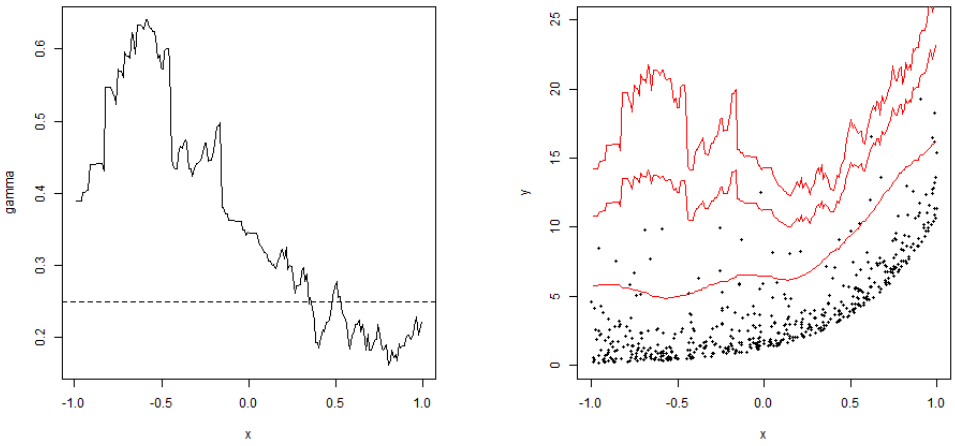

Figure 3.7: The estimated extreme value index as a function of x (left), where $k = 10$. The extreme quantile estimates (right) for the percentiles $0.95, 0.99, 0.995$ estimated with the kernel quantile estimator.

Chapter 4

The common shaped tail model

In the previous chapter, several estimators are discussed to obtain inference about the tail of the conditional distribution, $Y|X = x$. Every method makes use of a two step estimation procedure. In the first step the intermediate parts of the tail is estimated by estimating the quantile curves corresponding to a sequence of percentiles $\tau_1, \ldots \tau_p$. The second step extrapolates these intermediate quantiles, by using methods from univariate extreme value theory, which allow inference about the extreme quantiles of the conditional distribution, i.e. the quantiles that are beyond the range of the data.

The estimators, which are discussed in Chapter 3, are proposed in a general setting where the extreme value index is allowed to vary as a function of the covariates, i.e. $\gamma = \gamma(x)$. This introduces uncertainty, which is not desirable when the extreme value index can be assumed constant, as in extreme value theory the sparsity of extremes cause already large uncertainties.

Data from practice might have the feature that the extreme value index does not vary as a function of the covariates. For example, in the setting of post-processing of weather events it is natural to assume a constant extreme value index, as the distribution of a weather event given a forecast is not expected to have a varying tail index as a function of the forecast. The post-processing of precipitation is done in Chapter 6 where it is shown that the assumption of a constant extreme value index attains better validation results.

It can also be clearly seen from the Figures in previous chapter, i.e. 3.3, 3.5 and 3.5, that in the case of a stationary extreme value index this is not always reflected by the estimators of Chapter 3.

This gives the motivation to develop an estimator, that assumes a stationary extreme value index, but models the quantile curves without a distributional assumption to adequately quantify the relation between Y and X.

Therefore, a model will be introduced, which is referred to as the common shaped tail model. Though the model assumes a constant extreme value index it is still very flexible and it includes the common slope model which is introduced in [19]. By exploiting the structure of the model a quantile regression estimator is proposed that combines information from different percentile levels and thus is expected to be more efficient than the four methods introduced before, if the data fit in our framework.

It will additionally be shown that, as a by-product, the quantile crossing problem is automatically avoided in our procedure.

4.1 Combining quantile curves

As discussed in Chapter 2, regression quantiles model the τth quantile of the conditional distribution $Y|X$ as a function of X. Consider now the following model for the quantile curves,

$$Q(\tau|x) = r(x) + \epsilon(\tau). \tag{4.1}$$

Note that in such an additive model the influence of x on different quantile levels is the same, which means that all the quantile curves should be parallel. Typically, following the theory outlined in Chapter 2, $r(x)$ can be estimated by solving the following minimization problem for τ_0 such that $\epsilon(\tau_0) = 0$,

$$\arg\min_{r(x)} \sum_{i=1}^{n} \rho_{\tau_0}(Y_i - r(X_i)).$$

Though, for any percentile level τ the corresponding quantile curve is given by $r(x) + c$ with c a constant. This means that the shape of every quantile curve is exactly the same as $r(x)$, but shifted with a constant c.

Recall that a quantile curve estimate for a fixed percentile τ_0 roughly splits the data in $\tau_0 n$ observations below the curve and $(1 - \tau_0)n$ observations above the curve. Consider now a different percentile level τ_1, for which the corresponding quantile curve is assumed to have exactly the same shape as the quantile curve at τ_0. Therefore the estimate could be improved by considering τ_0 and τ_1 at the same time.

Combining information for several quantile levels will be one of the main properties of the proposed estimator. The current section will restrict to the linear version of Model 4.1 to make the explanation of combining quantile levels more clear, i.e.

$$Q(\tau|x) = \alpha(\tau) + \beta x.$$

The common shape of the quantile curves is now given by the slope β, which is the same for every percentile level τ. In [21] an estimator for the common slope was developed and is referred to as the composite quantile regression. For a sequence of percentile levels $\tau_1, \ldots \tau_p$ the composite quantile estimator is then given by the following quantile minimization problem,

$$\arg\min_{\alpha_1,\ldots,\alpha_p,\beta_c} \sum_{j=1}^{p} \sum_{i=1}^{n} \rho_{\tau_j}(Y_i - \alpha_j - \beta_c X_i).$$

The common slope estimator is shown to be more efficient in comparison with the slope estimator for a single regression quantile.

The fact that a common slope or a common shape estimator, which will be introduced in the next section , is more efficient is quite intuitive. To see this recall that a regression quantile for a single percentile τ finds the optimal line that splits the observations in a portion τ and a portion $1 - \tau$. The common slope estimator does exactly the same, but not for a sequence of percentiles. It therefore estimates one curve and p intercepts.

In the following section the linear threshold model will be discussed, i.e. the fitting of the generalized pareto distribution to the exceedances above a linear threshold. The usage of the composite quantile regression will be used to obtain a more efficient estimator of the threshold.

4.2 The common shaped tail model estimator

Consider the following model for the conditional distribution of $Y|(X = x)$,

$$Y|(X = x) = r(x) + \epsilon(U),$$

where $\epsilon(U)$ is some random variable with distribution function F_ϵ. It can be easily seen from this equation that the expression of the quantile curves is given as in Equation 4.1, where $\epsilon(\tau) = F_\epsilon^{-1}(\tau)$ is the quantile function of Z.

Note that the additivity in the model imposes some significant structure. First of all, the quantile curves as a function of x are all parallel, which implies that all the quantile curves are of the same shape and only differ by a constant. Secondly, the conditional distribution $Y|(X = x)$ only depends on X via the sift parameter, which is in the model given by $r(x)$.

The interest lies solely in the tail of the distribution, therefore the model can be stated more generally. Define τ_c to be a high percentile, then the common shaped tail model is defined as,

$$Q(\tau|X = x) = r(x) + \epsilon(\tau) \text{ for } \tau \geq \tau_c. \tag{4.2}$$

There is one issue that needs to be resolved in this setting, which is the identifiability of the model. Let $c \in \mathbf{R}$ denote a constant then the model can be rewritten as,

$$Q(\tau|X = x) = (r(x) + c) + (\epsilon(\tau) - c) \text{ for } \tau \geq \tau_c.$$

To ensure identifiability assume that $\epsilon(\tau_c) = 0$.

In a similar way as is done in the methods of Chapter 3, inference about the tail can be obtained via a two step approach, where the first step is concerned with the intermediate quantiles and the second step extrapolates to the extreme quantiles. Let $\hat{r}(x)$ denote some estimator of $r(x)$ and the residuals of this estimator are given by,

$$e_i = Y_i - \hat{r}(X_i). \tag{4.3}$$

A theoretical expression of these residuals is now given by,

$$e_i = \begin{cases} (r(X_i) - \hat{r}(X_i)) + \epsilon(u_i) & \text{if } u_i \geq \tau_c \\ Q(u_i|X_i) - \hat{r}(X_i) & \text{otherwise,} \end{cases} \tag{4.4}$$

where u_i is defined by $u_i = F(Y_i|X = X_i)$.

The following two propositions give us the tools to estimate the tail distribution of $Y|X = x$,

Proposition 4.2.1. *Assume that the quantile curves of a random variable $Y|(X = x)$ are given by,*

$$Q(\tau|X = x) = r(x) + \epsilon(\tau)\tau \in (0, 1)x \in \mathbf{R},$$

In addition assume that $Y|(X = x)$ is in the Fréchet domain of attraction. Then $Y|(X = x)$ is tail equivalent for every x.

Proof. Let x_1 and x_2 be two arbitrary fixed values, γ_1 and γ_2 the extreme value index for $Y|(X = x_1)$ and $Y|(X = x_2)$, respectively. Finally let U be a standard uniformly distributed

variable. Then it follows for $y > 0$,

$$\begin{aligned}
y^{-1/\gamma_1} &= \lim_{t\to\infty} \frac{\mathrm{P}(Y > ty|X = x_1)}{\mathrm{P}(Y > t|X = x_1)} = \lim_{t\to\infty} \frac{\mathrm{P}(r(x_1) + \epsilon(U) > ty)}{\mathrm{P}(r(x_1) + \epsilon(U) > t)} \\
&= \lim_{t\to\infty} \frac{\mathrm{P}(r(x_2) + \epsilon(U) > ty + (r(x_2) - r(x_1)))}{\mathrm{P}(r(x_2) + \epsilon(U) > t + (r(x_2) - r(x_1)))} \\
&= \lim_{s\to\infty} \frac{\mathrm{P}(r(x_2) + \epsilon(U) > sy + (1-y)(r(x_2) - r(x_1)))}{\mathrm{P}(r(x_2) + \epsilon(U) > s)} \\
&= \lim_{s\to\infty} \left(\frac{sy + (1-y)(r(x_2) - r(x_1))}{s} \right)^{-1/\gamma_2} \\
&= \lim_{s\to\infty} \left(y + \frac{(1-y)(r(x_2) - r(x_1))}{s} \right)^{-1/\gamma_2} \\
&= y^{-1/\gamma_2}.
\end{aligned}$$

In the second line a substitution is made for t by $s = t + r(x_2) - r(x_1)$. Hence, $\gamma_1 = \gamma_2$ and for all x; the tail is equivalent. □

The proposition shows that the model as presented in Equation 4.2 has constant extreme value index. The following proposition enables the estimation of the constant extreme value index from the residuals of the estimation of $r(x)$.

Proposition 4.2.2. *Let Y be distributed according to the model in Equation 4.2 with x the corresponding fixed covariate level and define γ to be the extreme value index of $Y|x$. Now let $\hat{\beta}_c$ be the estimated slope and the residuals $E_i = Y_i - \hat{r}(X_i)$. Then the tail index of E_i is given by γ*

Proof. In the following part the expression of the residuals in Equation 4.3 is used. Define U as a uniform random variable.

$$\begin{aligned}
\lim_{t\to\infty} \frac{\mathrm{P}(E_i > ty|X = x)}{\mathrm{P}(E_i > t|X = x)} &= \lim_{t\to\infty} \frac{\mathrm{P}(Q(U|X = x) - \hat{r}(x) > ty|X = x)}{\mathrm{P}(Q(U|X = x) - \hat{r}(x) > t|X = x)} \\
&= \lim_{t\to\infty} \frac{\mathrm{P}(r(x) + \epsilon(U) - \hat{r}(x) > ty|X = x)}{\mathrm{P}(r(x) + \epsilon(U) - \hat{r}(x) > t|X = x)} \\
&= \lim_{t\to\infty} \frac{\mathrm{P}(\epsilon(U) > ty - (r(x) - \hat{r}(x))|X = x)}{\mathrm{P}(\epsilon(U) > t - (r(x) - \hat{r}(x))|X = x)} \\
&= \lim_{t\to\infty} \left(\frac{ty - (r(x) - \hat{r}(x))}{t - (r(x) - \hat{r}(x))} \right)^{1/\gamma} \\
&= y^{1/\gamma}
\end{aligned}$$

□

4.3 A non-parametric common shaped tail estimator

In this section an estimator is proposed for the common shaped tail model. As explained in the previous section the estimation will consist of a two step procedure, where first the intermediate quantiles are estimated and then extrapolated to the extreme quantiles.

Recall the non-parametric quantile regression introduced in Chapter 2. The quantile curve is approximated locally with the first $m+1$ terms of Taylor expansion around a fixed point x^*,

$$\begin{aligned} Q(\tau|X=x) &\approx Q(\tau|X=x^*) + r'(x^*)(x-x^*) + \frac{r''(x^*)}{2}(x-x^*)^2 + \cdots + \frac{r^{(m)}(x^*)}{m!}(x-x^*)^m \\ &= \sum_{i=0}^{m} \beta_i (x-x^*)^i. \end{aligned}$$

Then the parameters β_i are estimated with quantile regression restricted to a neighbourhood of x^* defined by $h > 0$,

$$\underset{\beta_0,\ldots\beta_m}{\arg\min} \sum_{i=1}^{n} \mathbf{1}(X_i \in [x^*-h, x^*+h]) \rho_\tau \left(Y_i - \sum_{j=0}^{m} \beta_j (X_i - x^*)^j \right).$$

By assuming the model from Equation 4.2, all the quantile curves for $\tau \geq \tau_c$ are necessarily parallel. Let $\tau_1, \tau_2 \geq \tau_c$ and denote the estimated parameters by $(\beta_0^{(1)}, \ldots, \beta_m^{(1)})$ and $(\beta_0^{(2)}, \ldots, \beta_m^{(2)})$ respectively. In order to impose the parallel restriction $\beta_i^{(1)} = \beta_i^{(2)}$ for $i \geq 1.$, i.e all the higher order terms in the local Taylor expansion should be of a common shape as they depend on the covariates. As a result, the common shape can be estimated by borrowing strength from several quantile layers in $[\tau_c, 1]$

For a fixed covariate level x^* and bandwidth h, let the number of observations in the interval $[x^*-h, x^*+h]$ be denoted by N. Then let $p = [N(1-\tau_c)]$ denote the expected number of observations above the τ_c quantile curve. For the percentile sequence there are two requirements. First, it must contain the intermediate percentile τ_c. Second, it must use all the information available in the data, i.e. there must be one observation in between two quantile curves corresponding to consecutive percentiles. A percentile sequence can now be constructed as follows, $\tau_c, \tau_c + \frac{(1-\tau_c)}{p}, \ldots, \tau_c + \frac{(p-1)(1-\tau_c)}{p}$. This is percentile sequence which allows borrowing strength of the maximum number of percentile levels. From now on, the percentile sequence will be denoted by $\tau_1, \ldots, \tau_p$. Then the non-parametric quantile regression minimization can be done for all the percentiles at the same time,

$$\underset{\alpha_1,\ldots,\alpha_p,\beta_1,\ldots\beta_m}{\arg\min} \sum_{j=1}^{p} \sum_{i=1}^{n} \mathbf{1}(X_i \in [x^*-h, x^*+h]) \rho_{\tau_j} \left(Y_i - \alpha_j - \sum_{k=1}^{m} \beta_k (x-x^*)^k \right)$$

Here β_0 is replaced by parameter α_j corresponding to the different percentile levels. Note that $\epsilon(\tau_c) = 0$ implies $Q(\tau_c|x) = r(x)$. Therefore, it can be seen that $\hat{\alpha}_1$ is a point estimate of $\hat{r}(x^*)$. A sequence of point estimates can now be found for $r(X_1), \ldots, r(X_n)$. Note that for two covariate levels x_1^* and x_2^*, the percentile sequence is not the same but always starts at τ_c. These estimates in turn allow us to compute the residuals as indicated in Equation 4.3.

Figure 4.1 shows in the left panel the composite estimate of $r(x)$ together with the exceedances in blue for the simulated data example. In the right panel the residuals are shown.

Continuing from the definition of the residuals given in Equation 4.3, the extreme value index can be estimated from the residuals as is proven in Proposition 4.2.1. The maximum likelihood estimator is now used as estimator, $\hat{\gamma}$, for the extreme value index on the residuals.

An estimator for $\epsilon(\tau)$ is obtained by the extrapolation of the residuals,

$$\hat{\epsilon}(\tau) = e_{n-k,n} \left(\frac{1-\tau}{1-k/n} \right)^{\hat{\gamma}}, \quad \tau > \tau_c.$$

The final estimator of an extreme quantile corresponding to an extreme percentile τ_n is given by,

$$\hat{Q}(\tau_n|X=x) = \hat{r}(x) + \hat{\epsilon}(\tau_n). \tag{4.5}$$

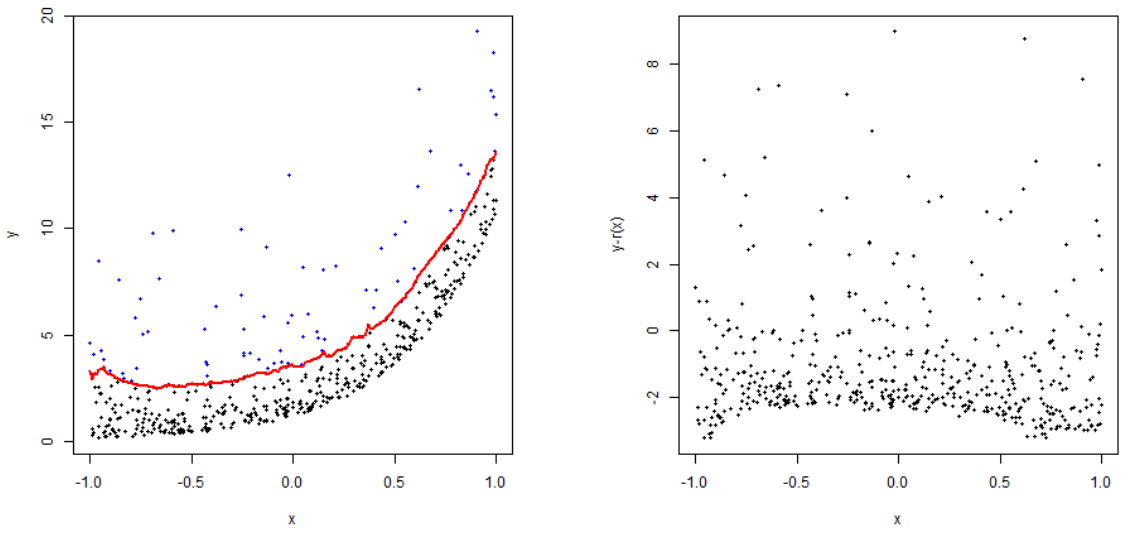

Figure 4.1: The simulated data example with the red line which is the 0.8 quantile curve(left), the exceedences are indicated in blue. The residuals from the estimation of the shape function $r(x)$ (right) where also the blue points indicate the exceedances.

In Figure 4.2 the extrapolation procedure is applied to the simulated data example. Here k is chosen to be 60. The extreme quantiles for percentiles $0.95, 0.99, 0.995$ are shown in the right panel. It is clear that the problem of quantile crossing does not appear and the extrapolation is now consistent for all the covariate levels.

4.3.1 Some remarks

Note that for every residual, the curve estimate at the corresponding X_i needs to be made. In the worst scenario all X_i are unique, which can make the estimation procedure very slow. To solve this problem one can discretize the domain of X in m points. Then for these m points the point estimates of the curve are given by $\hat{r}_1, \ldots, \hat{r}_m$. The point estimate for $r(X_i)$ can be obtained by interpolating the two closest estimates in the discretization. The interpolated point estimates $\hat{r}(X_1), \ldots \hat{r}(X_n)$ can be used in the computation of the residuals as in Equation 4.3 and these can be used for extrapolation to the tail.

The CST estimator exploits the properties of the common shaped tail model, but the estimated are not exactly parallel due to the randomness of the data. An approach is tried to force the quantile curves to be parallel, this did not yet give satisfying results. The approach is discussed in Appendix A.

Note that in general any estimator can be used for the extreme value index. Initially the Hill estimator was used for the estimation of the extreme value index. Though the Hill estimator is not shift invariant, i.e.

$$\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) \neq \log\left(\frac{Y_{n-i,n}+a}{Y_{n-k,n}+a}\right).$$

For identifiability of the model, a percentile level is chosen for $r(x)$, which in this case works as a shift. Because this threshold is chosen large, the shift that is applied in the procedure is substantial, which influences the estimator severely. Therefore the maximum likelihood estimator was chosen, because this estimator is shift invariant.

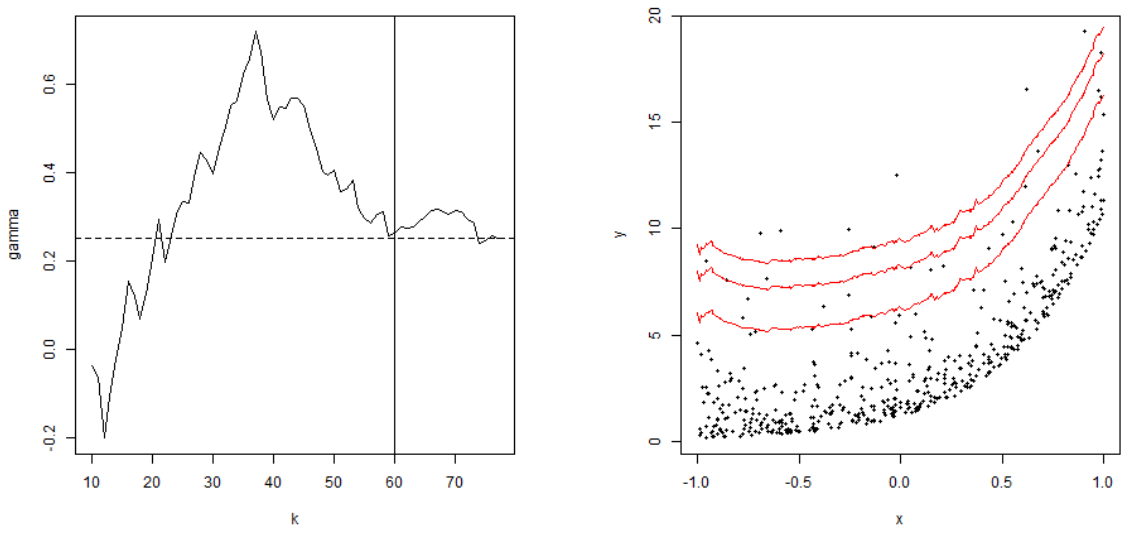

Figure 4.2: The extreme value index as a function of the number of upper order statistics of the exceedences (left). The vertical line indicates the number of upper order statistic that is chosen, i.e $k = 60$. The extrapolated extreme quantile curves (right) for the percentile levels $0.95, 0.99$ and 0.995.

Chapter 5

Simulation study of current methods

In this chapter, the finite sample performance of the non-parametric CST estimator is tested and compared to the existing methods from literature, that were discussed in Chapter 3. In the first section the simulation set-up will be discussed. The set-up is chosen to compare the estimators based on a wide range of possible distributions and shapes. The second section is concerned with the non-parametric CST estimator defined in Equation 4.5. By means of a simulation study the optimal choices for the threshold level, τ_c and the bandwidth h were determined. The selection criteria are based on the estimation procedure of both $r(x)$ and $\epsilon(\tau)$. Finally, in the last section, a comparison study was done by comparing the non-parametric CST estimator with the power transformed linear quantile estimator in Equation 3.2 and the kernel estimator in Equation 3.7.

5.1 The simulation set-up

The simulation set-up will be the same for all the simulations in this chapter. The simulated data was generated according to the model in Equation 5.1:

$$Q(\tau|X = x) = r(x) + \epsilon(\tau), \tag{5.1}$$

where $r(x) = x, \exp(x), sin(\pi x)$ and $\epsilon(\tau)$ the quantile function for $GPD(\gamma = 0.25), GPD(\gamma = 0.5), student - t(df = 1)$. The estimation of the extreme quantiles consists of the stages, that are explained in the previous two chapters. Both stages influence the estimate of the extreme quantile estimator. Recall that the extreme quantile estimator for our model was given by,

$$\hat{Q}(\tau_n|X = x) = \hat{r}(x) + \hat{Q}_\epsilon\left(1 - \frac{k}{n}\right)\left(\frac{1 - \frac{k}{n}}{1 - \tau_n}\right)^{\hat{\gamma}}$$

The first stage, introduces direct uncertainty in the extreme quantile estimator, but also indirectly as the extrapolation is based on the estimator via the residuals. The three different shape functions are chosen to compare the estimators based on a linear shape, a monotone shape and a non-monotonic shape. These three functions will enable us to test the performance of the first stage of the estimation procedure in a broad way.

In the second stage of the estimation procedure, the estimation of $\epsilon(\tau)$, the three different tail distributions are chosen to test for varying tail heaviness. The distributions have an extreme value index of 0.25, 0.5 and 1, respectively. The student-t distribution is chosen in order to see how the performance of the estimator is when the distribution is in the domain of an extreme value distribution, but not exactly equal to the generalized Pareto distribution.

Finally, three different sample sizes are considered to compare the performance for small($n = 100$), medium sized ($n = 500$) and large samples ($n = 1000$).

The covariate realizations are simulated from a uniform distribution between -1 and 1. The uniform distribution was chosen, because for every neighbourhood around x^* there are approximately the same number of observations. With $\epsilon(\tau)$ the quantile distribution of the error distribution, the response variables Y, are simulated with,

$$Y = r(X) + \epsilon(U)$$

Here U is a stand uniformly distributed random variable.

The results for extreme quantiles and the extreme value index will be compared with the mean squared error (MSE) and the bias, defined by,

$$\text{MSE}(\hat{\theta}) = \frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i - \theta)^2$$
$$\text{bias}(\hat{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\hat{\theta}_i - \theta,$$

where θ the exact value, $\hat{\theta}$ the estimate and $m = 500$ the number of re-samplings. However for $r(x)$, the interest does not lie in the performance of $\hat{r}(.)$ at a specific point x^*, but in the performance in the entire range of the covariates. Therefore an adjusted MSE is used for the estimator $\hat{r}(x)$,

$$\text{MSE}(\hat{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{1} 0(\hat{r}_i(x_j^*) - r(x_j^*))^2$$
$$\text{bias}(\hat{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{1} 0(\hat{r}_i(x_j^*) - r(x_j^*)),$$

where x_j^* for $1 \leq j \leq 10$ is an equally spaced sequence of covariates on the range of the covariate.

5.2 Choices for the non-parametric CST estimator

In this section, an extensive simulation study is done for the non-parametric common shaped tail estimator introduced in Chapter 4. Before the estimator can be applied on data choice need to be made for the hyper-parameters. For the CST estimator the choice needs to be made for the percentile level τ_c. For the estimation of $r(x)$ it generally holds that a lower τ_c results in smaller errors, due to the sparseness of tail observations. However, τ_c must be sufficiently large, such that the positive residuals follow approximately a generalized Pareto distribution.

In order to make these choices, both the estimation of $r(x)$ and γ are addressed in this section. The choice of h, the bandwidth, is generally done by a bandwidth selection method. The cross-validation method is tried, but did not yet give satisfying results. Instead the bandwidth that minimizes the MSE is chosen in this simulation study.

Before presenting the simulations consider Figure 5.1. This shows the quantile curve estimated with the non-parametric composite quantile estimator for percentile level $\tau_c = 0.8$ for a dataset of size $n = 100$, with the model with $r(x) = \sin(x\pi)$ and $\epsilon(\tau) \sim GPD(\gamma = 0.25)$.

It can be seen that at the boundary the estimates decrease rapidly. This is caused by the asymmetric distribution in the neighbourhood of a boundary point. This is quite unfortunate, because this rapid decrease may cause a severe overestimation of the exceedances in the boundary region and consequently, this will cause errors in the extrapolation. Although these boundary problems have a potential high influence on the estimates, solving these problem lies beyond the scope of this thesis.

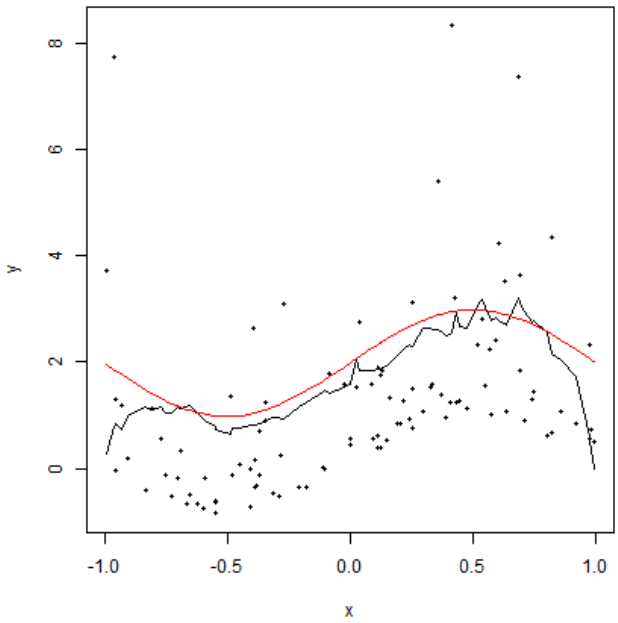

Figure 5.1: The plot shows the threshold estimate for data from the model of Equation 5.1 for $n = 100$, $r(x) = \sin(x\pi)$ and $\epsilon(\tau) \sim GPD(0.25)$. The estimate is shown as the black line and the exact threshold is shown as the red line.

To verify the influence of the boundary problems on the estimation of $r(x)$, two simulation settings will be considered. This will be the normal setting as presented above, but also the setting where the covariates are simulated uniformly between -1.5 and 1.5. The number of observations will then be increased by 50% and the shape function $r(x)$ will be calculated for covariate values between -1 and 1. This second setting removes the boundary problems, as in this setting there is no asymmetry of observations in a neighbourhood at the boundary. The sample sizes small, medium and large are for this setting defined by $n = 150$, $n = 750$ and $n = 1500$, respectively.

For the threshold, two percentiles are considered, $\tau_c^1 = 0.8$ and $\tau_c^2 = 0.9$. The bandwidth chosen for the simulations is chosen as the bandwidth that has the minimal mean squared error.

In Table 5.1 the results for the performance of the composite quantile estimator for $r(x)$ are shown. As a first observation it is clear that the MSE for the 0.9 threshold is always higher in comparison to the MSE for the 0.8 threshold. This is because the 0.9 threshold is in a more sparse region of the distribution, which makes estimating the quantile much harder. Comparing the simulation where the boundary problem arises and the part where the boundary problem is eliminated, it is clear that the MSE decreases significantly in almost every case. This gives a clear motivation for future research to deal with the boundary problem, as for the extrapolation part, it is essential that the estimation at every point of the covariate goes well as these are directly related to the residuals.

The bandwidth that is chosen for each of the simulation scenarios is displayed in Table 5.2. The optimal bandwidth was chosen out of an equally spaced sequence of 50 bandwidth values between the minimum h_1, such that there are at least two points in the neighbourhood of $x^* = 0$ and the maximum $h_{50} = 1$, which corresponds to the estimation over the entire interval.

From Table 5.2 it can be seen that for $r(x) = x$ averaging over the entire area of the covariate

is the best option, which is quite natural as the approximation is local linear. Additionally, it can be seen that for both values of τ_c the exponential quantile curves can best be approximated linearly when the sample size is small and the sine quantile curves are best linearly approximated for the 0.9 threshold. Finally the observation can be made that although the tail of the student-t distribution is more heavy then the generalized Pareto distribution with shape parameter $\gamma = 0.5$ it allows generally for a smaller bandwidth. This can be explained with a picture of the density functions in Figure 5.2. The red dots indicate that at the 0.8 quantile the density of the student-t distribution is higher than the density of the generalized Pareto distribution. Therefore the data is less sparse for the student-t distribution at this percentile level.

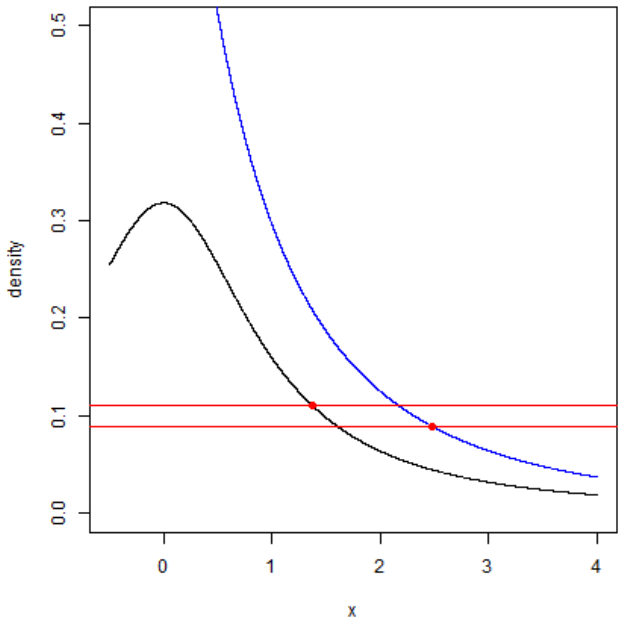

Figure 5.2: The plot shows in black the density of the student-t distribution with $df = 1$ and in blue the density of the generalized pareto distribution with shape parameter $\gamma = 0.5$. The red dots on the density curves indicate the 0.8 quantiles.

From the previous simulation it is clear that the 0.8 percentile is best for the estimation of the threshold. This is obvious as the data at the 0.8 quantile is less sparse than the data at the 0.9 quantile. However, as already stated, this does not imply that this is the optimal percentile level from the extrapolation perspective in the second stage.

In order to see whether the threshold is better set at 0.8 or 0.9 the MSE and the bias of the extreme value index will be compared, which is the most important parameter related to the tail of the distribution.

For τ_c three levels will be chosen 0.8, 0.85 and 0.9, such that the behaviour of the extreme value index can be observed as a function of an increasing threshold. The selection of k, i.e. the number of upper order statistics, also needs to be addressed. A few simulations have been made and k is selected by visual inspection of the estimate as a function of k from which the first stable period determines the choice of k. This is done for every scenario of the simulation. Due to the fact that the maximum likelihood estimator was used for the estimation, there were some convergence issues in some of the simulation scenarios. In these cases a slightly larger k value was chosen and the simulation was restarted.

$r(x)$	dist	With boundary problem				Without boundary problem			
		$\tau_c = 0.8$		$\tau_c = 0.9$		$\tau_c = 0.8$		$\tau_c = 0.9$	
					$n = 100$				
x	GPD(0.25)	2.54	(-1.49)	6.16	(-3.00)	1.78	(-0.18)	4.00	(-1.51)
	GPD(0.5)	5.68	(-0.94)	17.93	(-3.14)	3.64	(-1.12)	13.19	(-3.20)
	t(1)	13.05	(1.87)	50.74	(-0.62)	2.65	(1.01)	34.92	(3.78)
$\exp(x)$	GPD(0.25)	2.56	(-0.74)	5.88	(-2.35)	1.82	(-0.24)	3.92	(-1.17)
	GPD(0.5)	5.57	(-0.29)	19.61	(-1.80)	3.74	(-0.50)	13.09	(-2.13)
	t(1)	5.46	(-1.24)	8153.31	(7.92)	3.62	(0.20)	21.55	(0.59)
$\sin(x\pi)$	GPD(0.25)	2.90	(-0.11)	7.01	(-1.38)	2.00	(-0.48)	5.11	(-1.50)
	GPD(0.5)	6.18	(-1.15)	20.28	(-2.80)	4.67	(-0.67)	14.02	(-2.79)
	t(1)	526.48	(6.17)	14757.30	(10.98)	5.71	(-1.39)	31.61	(4.69)
					$n = 500$				
x	GPD(0.25)	0.62	(-0.46)	1.53	(-1.33)	0.43	(-0.43)	1.06	(-0.42)
	GPD(0.5)	1.27	(-0.31)	4.77	(-1.51)	0.87	(-0.50)	3.50	(-0.33)
	t(1)	0.94	(-0.03)	11.75	(0.64)	0.63	(-0.09)	18.76	(-0.28)
$\exp(x)$	GPD(0.25)	0.58	(-0.23)	1.41	(-0.34)	0.42	(-0.21)	1.08	(-0.15)
	GPD(0.5)	1.29	(-0.03)	4.53	(-1.33)	0.86	(-0.19)	3.35	(-0.31)
	t(1)	0.97	(0.22)	11.49	(0.01)	0.59	(0.08)	3.00	(1.42)
$\sin(x\pi)$	GPD(0.25)	0.87	(-0.41)	1.79	(-0.58)	0.64	(0.05)	1.38	(-0.49)
	GPD(0.5)	1.55	(-0.19)	5.71	(-0.58)	1.19	(-0.37)	3.81	(-0.45)
	t(1)	3.83	(-1.60)	5.63	(-0.25)	0.90	(-0.15)	4.09	(-0.33)
					$n = 1000$				
x	GPD(0.25)	0.34	(-0.41)	0.81	(-0.39)	0.27	(-0.51)	0.55	(-0.27)
	GPD(0.5)	0.68	(-0.58)	2.61	(-0.84)	0.51	(-0.57)	1.78	(-0.69)
	t(1)	0.48	(-0.19)	602.73	(2.25)	0.32	(-0.28)	1.67	(-0.78)
$\exp(x)$	GPD(0.25)	0.33	(-0.18)	0.81	(-0.15)	0.26	(-0.20)	0.55	(-0.15)
	GPD(0.5)	0.68	(-0.20)	2.65	(-0.13)	0.55	(-0.24)	1.81	(-0.32)
	t(1)	0.45	(0.05)	186.86	(0.15)	0.34	(0.01)	1.62	(-0.27)
$\sin(x\pi)$	GPD(0.25)	0.50	(0.03)	1.11	(-0.30)	0.33	(0.02)	0.86	(-0.31)
	GPD(0.5)	1.01	(-0.38)	2.91	(-0.72)	0.69	(0.17)	2.16	(-0.62)
	t(1)	0.84	(0.41)	2367.07	(3.59)	0.64	(-0.05)	120.84	(0.10)

Table 5.1: The table shows MSE and (bias) results for the simulation of the shape $r(x)$. The left part shows the results in the case when the boundary problems are present and the right part shows the results for the case when the boundary problems are absent. The MSE is the sum the the MSE of quantiles corresponding to 21 covariate levels equally spaced over $[-1, 1]$.

$r(x)$	x			$\exp(x)$			$\sin(x\pi)$		
	GPD(0.25)	GPD(0.5)	student(1)	GPD(0.25)	GPD(0.5)	student(1)	GPD(0.25)	GPD(0.5)	student(1)
					$n = 100$				
$\tau_c = 0.8$	1.00	1.00	1.00	0.89	1.00	1.00	0.51	0.64	0.64
$\tau_c = 0.9$	1.00	1.00	1.00	0.64	1.00	1.00	0.54	1.00	0.58
					$n = 500$				
$\tau_c = 0.8$	0.85	0.89	0.94	0.63	0.80	0.67	0.34	0.43	0.39
$\tau_c = 0.9$	1.00	1.00	1.00	0.32	0.82	0.69	0.32	0.49	0.10
					$n = 1000$				
$\tau_c = 0.8$	0.76	0.83	0.85	0.49	0.65	0.58	0.30	0.34	0.32
$\tau_c = 0.9$	1.00	1.00	0.61	0.27	0.43	0.61	0.30	0.47	0.47

Table 5.2: The table shows the bandwidth that is chosen for each of the simulation for the shape function. The bandwidth is chosen as the one that minimizes the MSE.

The simulation results are shown in Table 5.3. For the small sample sizes the likelihood maximization did not converge, due to a lack of observations. Therefore these results are not reported. It can be seen that a higher threshold results in a larger MSE in all cases. Therefore it is clear that for both the estimation of $r(x)$ and the estimation of $\epsilon(\tau)$ the threshold is best set at 0.8.

Also observe that the estimates of the extreme value index are quite consistent over the different shape functions, i.e. the mean squared error does not differ much between the different functions of $r(x)$. The bias disappears in the case of large sample sizes, but for small and medium sample sizes the bias is negative for almost all the scenarios. Furthermore, the absolute value of the bias increases with the threshold. This is counter intuitive as a higher threshold should imply a better approximation of the GPD. This shows that the errors of $\hat{r}(x)$ have a large influence on the residuals.

In the next section simulations of the extreme quantile estimates by the non-parametric CST estimator are compared with the extreme quantile estimates of the other methods, introduced in Chapter 3 to see how well it performs compared to the other estimators.

$r(x)$	$\epsilon(\tau)$	$\tau_c = 0.8$		$\tau_c = 0.85$		$\tau_c = 0.9$	
				$n = 100$			
x	GPD(0.25)	2.44	(-1.88)	NA	NA	NA	NA
	GPD(0.5)	2.79	(-1.80)	NA	NA	NA	NA
	student(1)	3.85	(-2.01)	NA	NA	NA	NA
$\exp(x)$	GPD(0.25)	2.92	(-2.35)	NA	NA	NA	NA
	GPD(0.5)	2.35	(-1.96)	NA	NA	NA	NA
	student(1)	3.60	(-2.69)	NA	NA	NA	NA
$\sin(x\pi)$	GPD(0.25)	2.23	(-1.74)	NA	NA	NA	NA
	GPD(0.5)	4.43	(-1.93)	NA	NA	NA	NA
	student(1)	3.80	(-1.95)	NA	NA	NA	NA
				$n = 500$			
x	GPD(0.25)	0.24	(-0.29)	0.35	(-0.31)	0.53	(-0.68)
	GPD(0.5)	0.33	(-0.21)	0.46	(-0.03)	0.81	(-0.41)
	student(1)	0.58	(-0.50)	0.75	(-0.24)	1.23	(-0.17)
$\exp(x)$	GPD(0.25)	0.24	(-0.32)	0.33	(-0.33)	0.51	(-0.53)
	GPD(0.5)	0.39	(-0.31)	0.49	(-0.14)	0.75	(-0.44)
	student(1)	0.57	(-0.61)	0.72	(-0.19)	1.17	(0.02)
$\sin(x\pi)$	GPD(0.25)	0.26	(-0.26)	0.39	(-0.25)	0.54	(-0.60)
	GPD(0.5)	0.34	(-0.00)	0.46	(-0.13)	0.72	(-0.27)
	student(1)	0.54	(-0.55)	0.77	(-0.42)	1.20	(0.05)
				$n = 1000$			
x	GPD(0.25)	0.10	(-0.13)	0.14	(-0.09)	0.24	(-0.07)
	GPD(0.5)	0.16	(0.05)	0.22	(0.08)	0.33	(0.05)
	student(1)	0.27	(-0.30)	0.37	(-0.20)	0.59	(0.15)
$\exp(x)$	GPD(0.25)	0.11	(-0.04)	0.15	(-0.09)	0.23	(-0.13)
	GPD(0.5)	0.16	(0.02)	0.21	(0.00)	0.30	(-0.08)
	student(1)	0.27	(-0.27)	0.37	(0.03)	0.53	(0.12)
$\sin(x\pi)$	GPD(0.25)	0.12	(-0.16)	0.16	(-0.14)	0.25	(-0.14)
	GPD(0.5)	0.16	(0.07)	0.24	(0.03)	0.32	(-0.09)
	student(1)	0.28	(-0.00)	0.35	(-0.13)	0.64	(0.15)

Table 5.3: Simulation results for the extreme value index. The MSE and the (bias) are multiplied by 10. For the threshold at $\tau = 0.8$ and for the threshold at $\tau = 0.9$.

5.3 Comparison of extreme quantile estimators

In this section the non-parametric CST estimator, from Equation 4.5 will be compared with the power transformed linear quantile curve estimator, from 3.2, and the kernel estimator for extreme quantile curves, from Equation 3.7. The comparison will be based on the MSE and the bias for the extreme quantiles corresponding to the percentiles $\tau = (0.99, 0.995)$

The simulation set-up is the same as in Section 5.1. The results are split up over three tables where each table corresponds to a distribution of $\epsilon(\tau)$.

For the non-parametric CST estimator τ_c is chosen to be 0.8, the bandwidth is chosen as the bandwidth from $(0.2, 0.4, 0.6, 0.8)$ as the one that minimizes the error in the threshold. The k upper order statistics are chosen in the same way as is done for the simulation of the extreme value index.

For the power transformed linear quantile curves the number of upper order statistics is set at $k = 3n^{1/3}$, which is supposed to be a point where the method becomes stable. There is however a problem in the method for the student-t distribution. The implementation in R that is used requires positive data. In order to achieve this, a shift can be performed on all data such that the shifted data is positive. This is done by defining $a > 0$, such that $Y_i + a > 0$, $\forall 1 \leq i \leq n$. This procedure was required because the log transform can only handle positive data. Because of the very heavy left tail, the shift was too large, which resulted in information loss. Instead the power of the transformation is estimated on the positive part of the distribution where the log transform was not allowed. Then this power transform was applied to the data and the procedure was continued from here. This worked in some cases, but unfortunately not for all. When no results were obtained the table is filled with NA.

For the kernel estimator a percentile sequence is chosen of length 20, equally spaced between 0.8 and 0.99. The number of upper order statistics was chosen by plotting the estimate of $\gamma_k(x = 0)$ as a function of k and choosing k at the point where this function becomes stable. Finally the bandwidth was chosen by a cross validation method which was proposed in [6].

The tables below show the results of the simulation study; Table 5.4 shows the results for the $GPD(\gamma = 0.25)$ error distribution, Table 5.5 for the $GPD(\gamma = 0.5)$ error distribution and Table 5.6 for the Student-t. In the Figures 5.3, 5.4 and 5.5 the MSE for the $\tau_n = 0.995$ are plotted for better comparison.

The results for $GPD(\gamma = 0.25)$ in Table 5.4 and Figure 5.3 show clearly that the power transformed linear curves estimator is the best. The CST estimator is only better one time for $n = 100$. It can be clearly from the figure that relative to the other estimators, the CST estimator becomes worse when the sample size increases. The kernel curves estimator is worse than the other estimators, except for the large sample sizes where it is better than the CST estimator. It can be seen that the MSE as a function of x is almost constant for the different scenarios, which is not the case for the other estimators.

The results for $GPD(\gamma = 0.5)$ in Table 5.5 and Figure 5.4 and the student-t distribution in Table 5.6 and Figure 5.5 are comparable with the results with the $GPD(\gamma = 0.25)$ distribution. A final observation can be made for the bias. For the CST estimator the bias is in many cases substantially larger then the bias for the other estimators. Especially for the large sample sizes, i.e. $n = 1000$, the bias of the CST estimator is many times larger then the biases of the other two estimators.

For the student-t distribution the kernel distribution performs extremely bad for small sample sizes. Although the R-code was changed for the linear quantile curves and the student-t distribution, it still fails some of the times, which makes comparison hard. However, when the linear estimator works, it can be seen that in these parts the CST estimator performs better for smaller sample sizes.

		$x=-0.5$				$x=0$				$x=0.5$			
$r(x)$	Method	$\tau=0.995$		$\tau=0.999$		$\tau=0.995$		$\tau=0.999$		$\tau=0.995$		$\tau=0.999$	
						$n=100$							
x	CST	**11.17**	(-2.14)	**35.07**	(-2.85)	10.94	(-2.12)	34.62	(-2.83)	11.01	(-2.14)	34.72	(-2.85)
	linear	15.23	(0.23)	69.32	(2.06)	**6.29**	(0.04)	**22.07**	(1.29)	**9.15**	(-0.73)	**25.02**	(-0.31)
	kernel	139.14	(4.36)	976.42	(11.06)	124.72	(3.96)	721.26	(9.13)	173.60	(3.62)	1310.78	(8.33)
$\exp(x)$	CST	12.64	(-2.14)	46.44	(-2.76)	12.48	(-2.07)	46.28	(-2.69)	12.18	(-2.06)	45.62	(-2.68)
	linear	**12.07**	(-0.54)	**37.07**	(-0.02)	**6.72**	(-0.35)	**18.04**	(0.03)	**8.46**	(-0.93)	**19.78**	(-1.00)
	kernel	1082.81	(5.19)	23447.52	(15.20)	123.71	(3.46)	587.46	(7.30)	222.12	(3.82)	1470.44	(7.94)
$\sin(x\pi)$	CST	13.02	(-2.06)	59.12	(-2.73)	14.30	(-2.33)	60.91	(-3.00)	15.65	(-2.58)	62.83	(-3.25)
	linear	**10.80**	(0.22)	**40.65**	(1.45)	**6.73**	(-0.29)	**20.95**	(0.64)	**11.47**	(-1.21)	**29.53**	(-0.83)
	kernel	376.22	(5.16)	3860.68	(14.72)	1243.49	(6.10)	17912.60	(17.51)	310.46	(4.79)	2142.82	(11.28)
						$n=500$							
x	CST	4.97	(-2.02)	11.18	(-2.90)	4.88	(-2.01)	11.04	(-2.89)	4.88	(-2.01)	11.04	(-2.89)
	linear	**2.11**	(-0.46)	**5.82**	(-0.19)	**1.54**	(-0.37)	**4.09**	(-0.09)	**2.62**	(-0.49)	**6.88**	(-0.43)
	kernel	11.88	(1.54)	52.17	(3.94)	8.54	(1.04)	32.04	(2.68)	8.23	(0.81)	28.84	(2.08)
$\exp(x)$	CST	4.94	(-2.03)	11.09	(-2.92)	4.99	(-2.03)	11.17	(-2.93)	4.92	(-2.01)	11.08	(-2.91)
	linear	**2.02**	(-0.50)	**5.18**	(-0.51)	**1.37**	(-0.26)	**3.63**	(-0.21)	**2.32**	(-0.47)	**5.95**	(-0.59)
	kernel	9.84	(1.03)	33.40	(2.40)	9.33	(0.98)	30.30	(2.16)	8.77	(0.90)	25.29	(1.79)
$\sin(x\pi)$	CST	3.98	(-1.74)	9.77	(-2.61)	4.94	(-1.99)	11.18	(-2.86)	6.00	(-2.25)	12.66	(-3.12)
	linear	**2.28**	(0.04)	**7.23**	(0.30)	**1.55**	(-0.38)	**4.26**	(-0.13)	**3.23**	(-0.94)	**7.34**	(-0.87)
	kernel	16.50	(2.07)	81.37	(5.39)	20.89	(1.74)	85.24	(4.08)	16.76	(1.34)	64.48	(2.96)
						$n=1000$							
x	CST	4.36	(-1.98)	9.61	(-2.86)	4.37	(-1.98)	9.63	(-2.86)	4.40	(-1.99)	9.68	(-2.87)
	linear	**1.12**	(-0.40)	**2.98**	(-0.39)	**0.77**	(-0.34)	**1.98**	(-0.30)	**1.35**	(-0.36)	**3.55**	(-0.44)
	kernel	4.96	(0.80)	21.33	(2.49)	4.30	(0.55)	16.04	(1.75)	3.90	(0.33)	12.59	(1.20)
$\exp(x)$	CST	4.46	(-2.01)	9.73	(-2.91)	4.41	(-2.00)	9.65	(-2.90)	4.29	(-1.97)	9.48	(-2.87)
	linear	**1.16**	(-0.36)	**2.99**	(-0.46)	**0.73**	(-0.17)	**1.96**	(-0.21)	**1.13**	(-0.32)	**2.93**	(-0.51)
	kernel	4.94	(0.59)	16.20	(1.62)	3.57	(0.29)	10.51	(0.97)	4.04	(0.28)	10.95	(0.78)
$\sin(x\pi)$	CST	5.02	(-2.14)	10.62	(-3.06)	5.25	(-2.20)	10.93	(-3.11)	5.57	(-2.27)	11.41	(-3.18)
	linear	**0.98**	(0.08)	**3.03**	(0.06)	**0.80**	(-0.35)	**2.13**	(-0.32)	**1.88**	(-0.84)	**4.12**	(-0.91)
	kernel	12.83	(1.67)	65.18	(4.71)	7.63	(0.92)	27.95	(2.44)	5.52	(0.42)	16.52	(1.24)

Table 5.4: MSE and (bias) for simulations done for $Q(\tau_n|x)$ with residual distribution $GPD(\gamma = 0.25)$, where $\tau_n = 0.995$ and $\tau_n = 0.999$

In the next Chapter the methods will be compared based on real dataset, with precipitation data as response and predicted precipitation as covariates.

		$x=-0.5$				$x=0$				$x=0.5$			
$r(x)$	Method	$\tau=0.995$		$\tau=0.999$		$\tau=0.995$		$\tau=0.999$		$\tau=0.995$		$\tau=0.999$	
						$n=100$							
x	CST	0.99	(-0.08)	**4.36**	(-0.11)	1.00	(-0.08)	4.39	(-0.11)	1.00	(-0.08)	4.40	(-0.11)
	linear	**0.98**	(-0.01)	4.71	(0.01)	**0.63**	(-0.01)	**2.49**	(-0.00)	**0.76**	(-0.02)	**2.66**	(-0.02)
	kernel	13.56	(0.13)	147.04	(0.36)	8.84	(0.09)	75.27	(0.25)	16.61	(0.13)	145.12	(0.34)
$\exp(x)$	CST	1.04	(-0.08)	4.15	(-0.12)	1.03	(-0.08)	4.13	(-0.12)	1.03	(-0.08)	4.13	(-0.12)
	linear	**0.68**	(-0.03)	**2.21**	(-0.03)	**0.51**	(-0.02)	**1.64**	(-0.03)	**0.68**	(-0.03)	**2.12**	(-0.05)
	kernel	22.20	(0.16)	201.57	(0.42)	12.92	(0.11)	105.65	(0.28)	21.85	(0.14)	170.95	(0.36)
$\sin(x\pi)$	CST	**0.87**	(-0.07)	**3.42**	(-0.10)	0.94	(-0.08)	3.50	(-0.10)	**1.00**	(-0.08)	**3.59**	(-0.11)
	linear	1.12	(-0.01)	6.21	(0.02)	**0.68**	(-0.02)	**2.85**	(-0.01)	1.26	(-0.04)	6.64	(-0.04)
	kernel	1114.72	(0.37)	67863.72	(1.98)	158.73	(0.24)	4389.14	(0.99)	196.84	(0.22)	4971.62	(0.89)
						$n=500$							
x	CST	0.75	(-0.08)	1.88	(-0.13)	0.75	(-0.08)	1.88	(-0.13)	0.75	(-0.08)	1.87	(-0.13)
	linear	**0.19**	(-0.02)	**0.70**	(-0.02)	**0.13**	(-0.01)	**0.47**	(-0.02)	**0.21**	(-0.02)	**0.75**	(-0.02)
	kernel	0.81	(0.03)	4.21	(0.09)	0.82	(0.03)	3.91	(0.08)	0.81	(0.03)	3.73	(0.07)
$\exp(x)$	CST	0.75	(-0.08)	1.90	(-0.13)	0.75	(-0.08)	1.90	(-0.13)	0.75	(-0.08)	1.89	(-0.13)
	linear	**0.20**	(-0.01)	**0.70**	(-0.02)	**0.14**	(-0.01)	**0.50**	(-0.02)	**0.23**	(-0.01)	**0.82**	(-0.02)
	kernel	0.48	(0.01)	1.92	(0.03)	0.55	(0.01)	2.23	(0.03)	0.41	(0.01)	1.49	(0.02)
$\sin(x\pi)$	CST	0.71	(-0.08)	1.83	(-0.13)	0.75	(-0.08)	1.89	(-0.13)	0.80	(-0.09)	1.96	(-0.13)
	linear	**0.17**	(-0.01)	**0.62**	(-0.02)	**0.14**	(-0.02)	**0.46**	(-0.02)	**0.23**	(-0.02)	**0.76**	(-0.03)
	kernel	8.32	(0.11)	79.99	(0.33)	7.11	(0.09)	56.44	(0.24)	3.46	(0.07)	17.59	(0.15)
						$n=1000$							
x	CST	0.73	(-0.08)	1.83	(-0.13)	0.73	(-0.08)	1.83	(-0.13)	0.73	(-0.08)	1.83	(-1.33)
	linear	**0.11**	(-0.01)	**0.38**	(-0.02)	**0.07**	(-0.01)	**0.25**	(-0.02)	**0.10**	(-0.01)	**0.35**	(-0.22)
	kernel	0.31	(0.01)	1.48	(0.03)	0.29	(0.00)	1.19	(0.02)	0.23	(0.00)	0.86	(0.14)
$\exp(x)$	CST	0.73	(-0.08)	1.82	(-0.13)	0.73	(-0.08)	1.82	(-0.13)	0.72	(-0.08)	1.81	(-1.31)
	linear	**0.10**	(-0.01)	**0.37**	(-0.02)	**0.07**	(-0.01)	**0.25**	(-0.02)	**0.10**	(-0.01)	**0.39**	(-0.23)
	kernel	0.61	(0.02)	2.71	(0.05)	0.80	(0.02)	3.29	(0.05)	0.74	(0.02)	2.87	(0.38)
$\sin(x\pi)$	CST	0.69	(-0.08)	1.76	(-0.13)	0.73	(-0.08)	1.82	(-0.13)	0.77	(-0.09)	1.89	(-1.35)
	linear	**0.09**	(-0.01)	**0.36**	(-0.02)	**0.07**	(-0.01)	**0.23**	(-0.02)	**0.12**	(-0.02)	**0.37**	(-0.28)
	kernel	1.11	(0.04)	6.93	(0.13)	1.51	(0.04)	8.61	(0.11)	0.96	(0.02)	4.14	(0.60)

Table 5.5: MSE $\times 10^{-1}$ and (bias) $\times 10^{-1}$ for simulations done for $Q(\tau_n|x)$ with residual distribution $GPD(\gamma = 0.5)$, where $\tau_n = 0.995$ and $\tau_n = 0.999$.

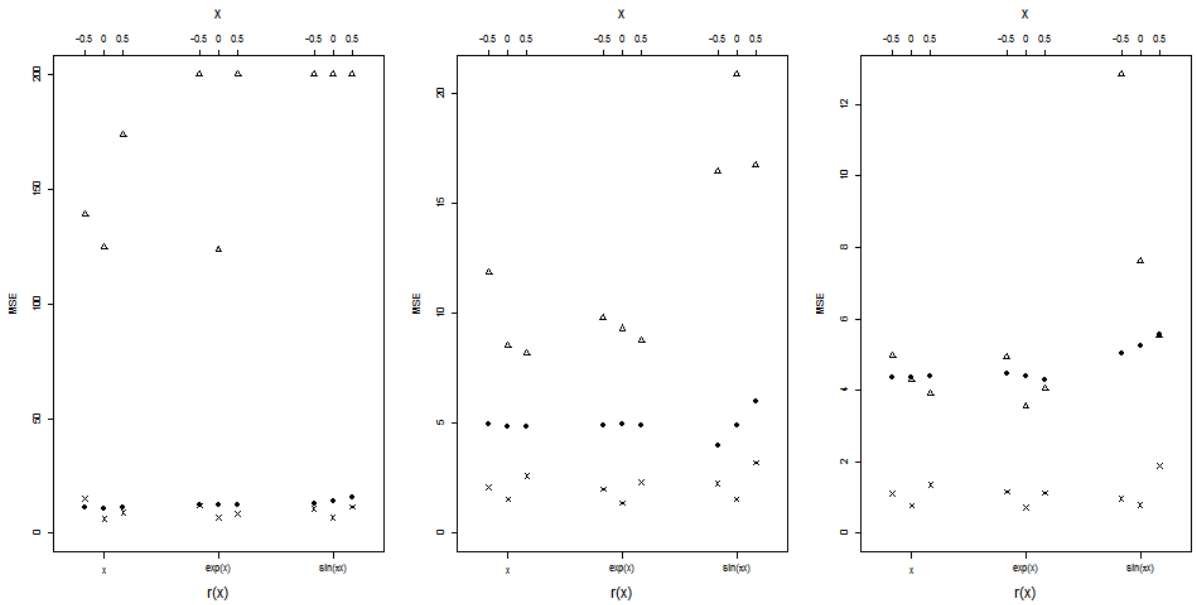

Figure 5.3: The MSE of the simulation results in Table 5.4 for $\tau_n = 0.995$, with $n = 100$ (left), $n = 500$ (middle) and $n = 1000$ (right). The scores for the CST estimator, the linear estimator and the kernel curves estimator are plotted with the filled dots, the crosses and triangles respectively.

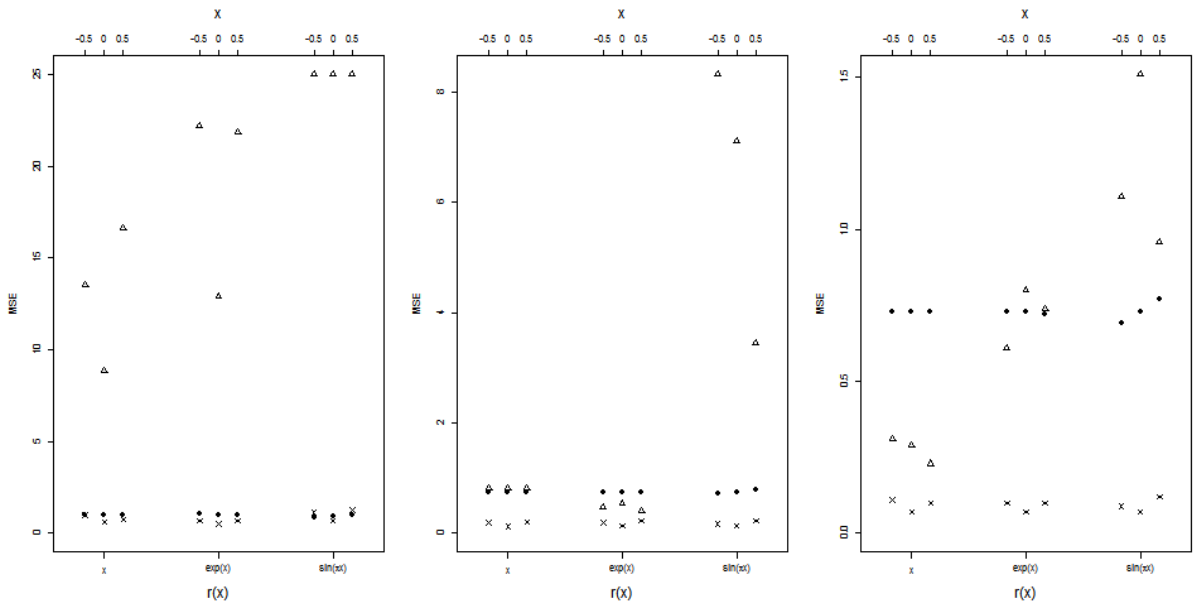

Figure 5.4: The MSE of the simulation results in Table 5.5 for $\tau_n = 0.995$, with $n = 100$ (left), $n = 500$ (middle) and $n = 1000$ (right). The scores for the CST estimator, the linear estimator and the kernel curves estimator are plotted with the filled dots, the crosses and triangles respectively.

		$x=-0.5$				$x=0$				$x=0.5$			
$r(x)$	Method	$\tau=0.995$		$\tau=0.999$		$\tau=0.995$		$\tau=0.999$		$\tau=0.995$		$\tau=0.999$	
						$n=100$							
x	CST	**0.51**	(-1.37)	**4.86**	(-2.12)	**0.51**	(-1.37)	**4.86**	(-2.11)	**0.51**	(-1.38)	**4.86**	(-2.12)
	linear	1.88	(-0.22)	26.60	(0.92)	1.22	(-0.55)	14.35	(-0.64)	NA	NA	NA	NA
	kernel	493.86	(15.12)	26812.27	(89.11)	99.10	(7.90)	3733.43	(37.65)	66.55	(7.82)	1920.71	(34.22)
$\exp(x)$	CST	**0.59**	(-1.24)	**7.72**	(-1.65)	0.59	(-1.23)	7.72	(-1.64)	0.59	(-1.23)	7.72	(-1.64)
	linear	NA	NA	NA	NA	**0.35**	(-1.06)	**1.86**	(-2.63)	**0.46**	(-1.28)	**2.41**	(-3.13)
	kernel	7376.44	(19.97)	1070301.61	(183.85)	15.97	(4.16)	253.32	(14.25)	90.46	(8.35)	2371.23	(34.24)
$\sin(x\pi)$	CST	**108.76**	(0.34)	**20355.43**	(19.52)	108.76	(0.30)	20355.17	(19.48)	**108.74**	(0.25)	**20354.88**	(19.42)
	linear	NA	NA	NA	NA	**0.45**	(-0.72)	**3.01**	(-1.37)	NA	NA	NA	NA
	kernel	NA	NA	NA	NA	74290687.98	(1380.63)	380207753935.19	(90202.43)	7249.71	(40.64)	1069267.33	(350.97)
						$n=500$							
x	CST	0.27	(-1.52)	**1.29**	(-3.08)	0.27	(-1.52)	1.29	(-3.08)	0.27	(-1.52)	1.29	(-3.08)
	linear	**0.24**	(-0.60)	2.15	(-1.26)	**0.15**	(-0.60)	**1.09**	(-1.45)	**0.15**	(-0.62)	**0.97**	(-1.64)
	kernel	2.10	(0.98)	46.55	(4.13)	0.66	(0.55)	6.10	(1.89)	0.99	(0.64)	10.43	(1.93)
$\exp(x)$	CST	0.28	(-1.56)	1.35	(-3.16)	0.28	(-1.55)	1.35	(-3.15)	**0.28**	(-1.55)	**1.35**	(-3.15)
	linear	**0.17**	(-0.76)	**1.17**	(-1.98)	**0.13**	(-0.77)	**0.90**	(-2.11)	NA	NA	NA	NA
	kernel	12.00	(3.18)	228.60	(11.77)	21.85	(3.55)	571.30	(13.88)	11.29	(2.67)	201.00	(8.89)
$\sin(x\pi)$	CST	0.27	(-1.51)	**1.38**	(-3.05)	0.28	(-1.53)	1.39	(-3.08)	0.29	(-1.56)	1.41	(-3.10)
	linear	**0.21**	(-0.53)	1.53	(-1.27)	**0.15**	(-0.65)	**1.11**	(-1.56)	**0.18**	(-0.79)	**1.25**	(-1.89)
	kernel	4.28	(1.95)	98.39	(8.39)	2.38	(1.40)	34.14	(4.88)	1.32	(0.67)	12.46	(2.03)
						$n=1000$							
x	CST	1.08	(-2.94)	40.41	(-12.89)	1.08	(-2.93)	40.40	(-12.89)	1.08	(-2.94)	40.41	(-12.89)
	linear	**0.14**	(-0.41)	**1.08**	(-1.06)	**0.09**	(-0.49)	**0.73**	(-1.34)	**0.09**	(-0.54)	**0.74**	(-1.57)
	kernel	0.92	(0.87)	10.88	(3.43)	0.69	(0.64)	6.52	(2.16)	0.94	(0.62)	9.79	(1.89)
$\exp(x)$	CST	0.26	(-1.56)	1.20	(-3.23)	0.26	(-1.56)	1.20	(-3.23)	0.26	(-1.56)	1.20	(-3.22)
	linear	NA	NA	NA	NA	**0.09**	(-0.63)	**0.66**	(-1.90)	**0.09**	(-0.65)	**0.72**	(-1.99)
	kernel	**0.16**	(-0.30)	**1.07**	(-0.81)	0.15	(-0.44)	0.92	(-1.28)	0.16	(-0.61)	0.98	(-1.83)
$\sin(x\pi)$	CST	3.36	(-1.20)	224.57	(-0.96)	3.37	(-1.25)	224.58	(-1.01)	3.39	(-1.30)	224.60	(-1.06)
	linear	**0.13**	(-0.42)	**1.02**	(-1.18)	**0.10**	(-0.53)	**0.74**	(-1.45)	**0.11**	(-0.62)	**0.80**	(-1.70)
	kernel	0.59	(0.76)	6.64	(3.31)	0.45	(0.30)	3.82	(1.12)	0.31	(0.01)	2.05	(-0.08)

Table 5.6: MSE $\times 10^{-3}$ and (bias) $\times 10^{-1}$ for simulations done for $Q(\tau_n|x)$ with residual distribution $student-t(1)$, where $\tau_n = 0.995$ and $\tau_n = 0.999$.

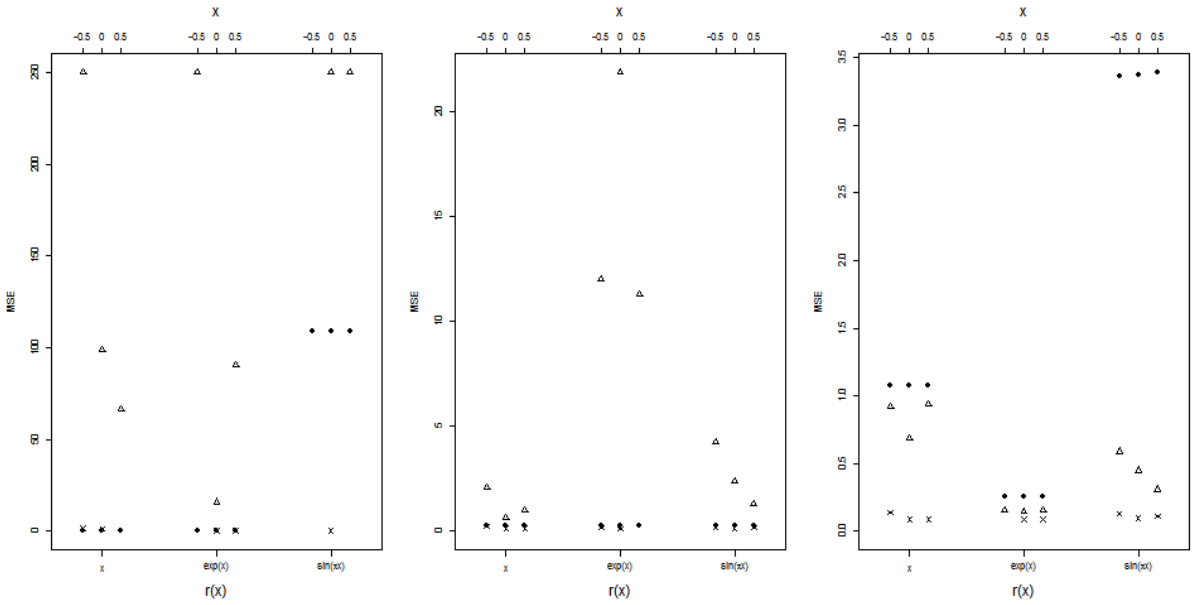

Figure 5.5: The MSE of the simulation results in Table 5.6 for $\tau_n = 0.995$, with $n = 100$ (left), $n = 500$ (middle) and $n = 1000$ (right). The scores for the CST estimator, the linear estimator and the kernel curves estimator are plotted with the filled dots, the crosses and triangles respectively.

Chapter 6

Application on precipitation data

6.1 Introduction to numerical weather prediction models

Quantitative precipitation forecasting (QPF), is one of the most difficult problems in the field of meteorology. This is due to large uncertainties in both the spatial and the temporal component. Numerical weather prediction models (NWP) models approximate the solution to the physical differential equations describing the atmospheric circulation pattern changes. This does not lead to perfect forecasts, as these changes are highly dependent on the initial conditions. A small change in these conditions can have a severe influence on the forecast, due to the non-linearity of the differential equations.

The large uncertainties are intuitively quite obvious as a slight change in the wind direction could cause a rain cloud to pass by instead of passing over a specific location. The large uncertainties in the forecast are not only a result of the uncertainty in the initial conditions. As a model is always an approximation of the real world there is also uncertainty introduced by parametrizations of non-resolved processes like cloud formation. The non-resolved processes are the processes, which are not described by the partial differential equations and therefore estimated on a small scale by parametrizing them. The parametrization is a simplification of the real world and therefore introduces uncertainty in the forecast. These two types of uncertainty will be referred to as uncertainty of the initial conditions and uncertainty of parametrizations.

A quantification of the uncertainty due to both of these causes needs to be made. An approach to quantify that uncertainty is called an ensemble. In the ensemble from ECMWF, instead of making one forecast using the 'best' initial condition, 50 more forecasts are made. Each of these forecasts has a perturbed initial condition and a slightly altered parametrization of the non-resolved processes. Therefore, instead of a single prediction, an empirical distribution is created, which corresponds to the uncertainty in both the initial condition and the parametrizations. These 50 perturbed members are referred to as ensemble members. Denote these 50 ensemble members by $P_1, \dots P_{50}$. The estimates of the quantiles of the forecast distribution can now be estimated by $\hat{Q}\left(\frac{i}{51}\right) = P_{i,50}$. Note that approximating the solution to these non-linear differential equations is a computationally heavy task. Therefore approximating the solutions for an entire ensemble requires the use of a super computer to obtain the forecasts in time.

In Figure 6.1 a visualization is made of an ensemble forecast. Here the green lines represents the forecasts of the 50 ensemble members. The blue line indicates the prediction with the 'best' initial condition and the red line is a reference forecast using the same model and initial conditions, but with a higher resolution.

The figure shows clearly that the first few days will be dry days with no rain, but as time progresses the forecasts start to have more uncertainty. A slight change in the initial conditions and the parametrizations at this time may cause an entirely different weather forecast for next week.

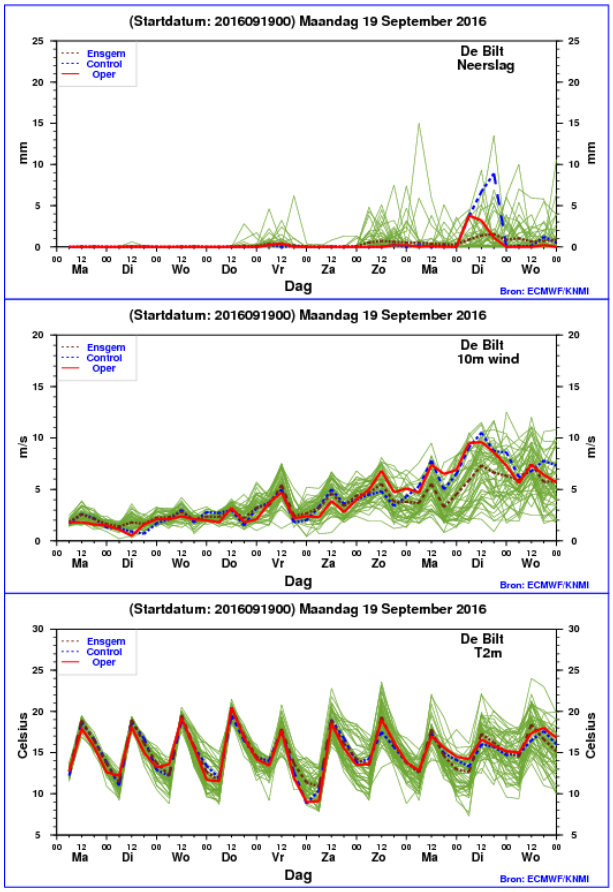

Figure 6.1: The visualization of an ensemble for precipitation (top), wind (middle) and temperature (bottom). Source: https://www.knmi.nl/nederland-nu/weer/waarschuwingen-en-verwachtingen/ensemble-verwachting-detail.

Statistical methods can be used to post-process the forecasts from the NWP models by using historical data. Because the predictions are calibrated with real data the uncertainty and the potential bias can be quantified and systematic errors can be corrected.

Wetterskip Fryslân is the organization responsible for the water management in the province of Friesland. Therefore it is important for them to have accurate forecasts concerning extreme precipitation, such that proper precautions can be taken in case of a possible overloading of the water drainage infrastructure. In the warm half year (April-October), there can be heavy precipitation showers especially in the late afternoon and evening, (15-21 UTC).

In this chapter a dataset of precipitation forecasts and observations for the years 2010,2011 and 2013 in the 15-21 UTC period is considered. An attempt is made to post-process precipitation forecasts from the NWP model HARMONIE, which is a relatively new high-resolution model used by KNMI. The methods that are compared are exactly the same as the ones in the simulation study of Chapter 5.

In the following two sections the gathering of the precipitation observations and the model output will be discussed. Then in the last two sections, the methods from Chapter 3 and Chapter 4 are applied to the dataset and compared using the quantile verification skill score as is used in [10].

6.2 Precipitation observations

Precipitation is measured with the radar system of KNMI. KNMI has radars in de Bilt and in Den Helder which are able to measure precipitation amounts for every location in the Netherlands. Due to the property of radar data to cover the entire area it is an easy operation to average the amount of precipitation over a specific area. However radar data is subject to a lot of noise, which affects the quality of the measurements.

To understand where the noise in the radar data comes from, the working of the radar will be briefly explained. A radar transmits pulses in all directions and these pulses are reflected by the drops of precipitation. The radar measures the reflections and these reflections are translated to precipitation measurements. In the case of heavy rainfall the amount of water drops in the air increases, which results in more reflections.

The radar measurements are displayed for a height of 1500 meters, but sometimes clouds are located lower to the ground. As a result it could happen that in a rain event the radar is not able to measure it.

Overestimation of the amount of precipitation is also possible. This happens when the air is really dry and precipitation evaporates before it reaches the ground. Additionally, in some circumstances the radar pulses could be bent towards the ground. When these radar pulses touch the ground, water or buildings cause reflections which distort the measurements.

The quality of the radar measurements can be improved by calibrating the data with point measurements of accumulated precipitation. KNMI has in addition to the radar data, a rain gauge network. This network consists of approximately 300 stations where the amount of precipitation of the previous 24 hours is measured every day at 8 UTC and 35 automatic stations where precipitation is measured continuously. These measurements are of significantly better quality, but the spatial coverage is of course worse.

See [16] for further discussion of the radar calibration method with the rain gauge system. This method will not be described here as it is beyond the scope of this thesis.

These calibrated radar data are used and area averaged to obtain the accumulated precipitation in Friesland between 15 and 21 UTC.

6.3 Predictor variables from the HARMONIE NWP model

The predictive variables are from a NWP model called HARMONIE. As explained already, a NWP model approximates the solution to the physical differential equations. By laying a grid over the area of interest the solution is approximated at each grid point. A full description of the HARMONIE model is not discussed as this is beyond the scope of this thesis. A full description of a similar model used in France, i.e. the AROME model, is given in [17].

The main difference between HARMONIE and its predecessors is that it has a finer grid and it resolves deep convection, which means that heavy showers can be resolved by the model. The grid size for the HARMONIE model is 2.5 km in comparison to its predecessor HIRLAM, which has a grid size of 11 km. Naturally, the computational complexity increases when the grid size is smaller. The total size of the area is restricted to western Europe. The boundary conditions are taken from the global ECMWF model, which has a lower resolution than both HIRLAM and HARMONIE.

As a covariate variable the area-averaged accumulated precipitation in Friesland between 15 and 21 UTC is used. Four different run times are considered. These consist of the forecasts started at 12 UTC, 6 UTC, 0 UTC and 18 UTC. The predicted periods all concern the 15-21 UTC interval, but will be denoted by the time interval with respect to the lead time, i.e 3-9 h, 9-15 h, 15-21 h and 21-27 h respectively.

The averaging over the area of Friesland is done by averaging all the predictions at the grid points which are in the smallest square covering Friesland. This square is defined by the following coordinates: 5.35-6.45 E, 52.8-53.4 N.

In Figure 6.2 scatter diagrams of the precipitation observations and the forecasts from the HARMONIE model are shown for the four different lead times. A problem with precipitation observations is that a lot of observations are 0 as on a substantial number of days it does not rain. In this chapter the focus on the conditional distribution of precipitation, given that it will rain. Therefore only the days where there was precipitation observed were included, in total this were 298 days for 2011,2012 and 2013. When used in practice also a forecast is needed for the probability of precipitation. In post-processing often logistic regression is used for estimating these probabilities.

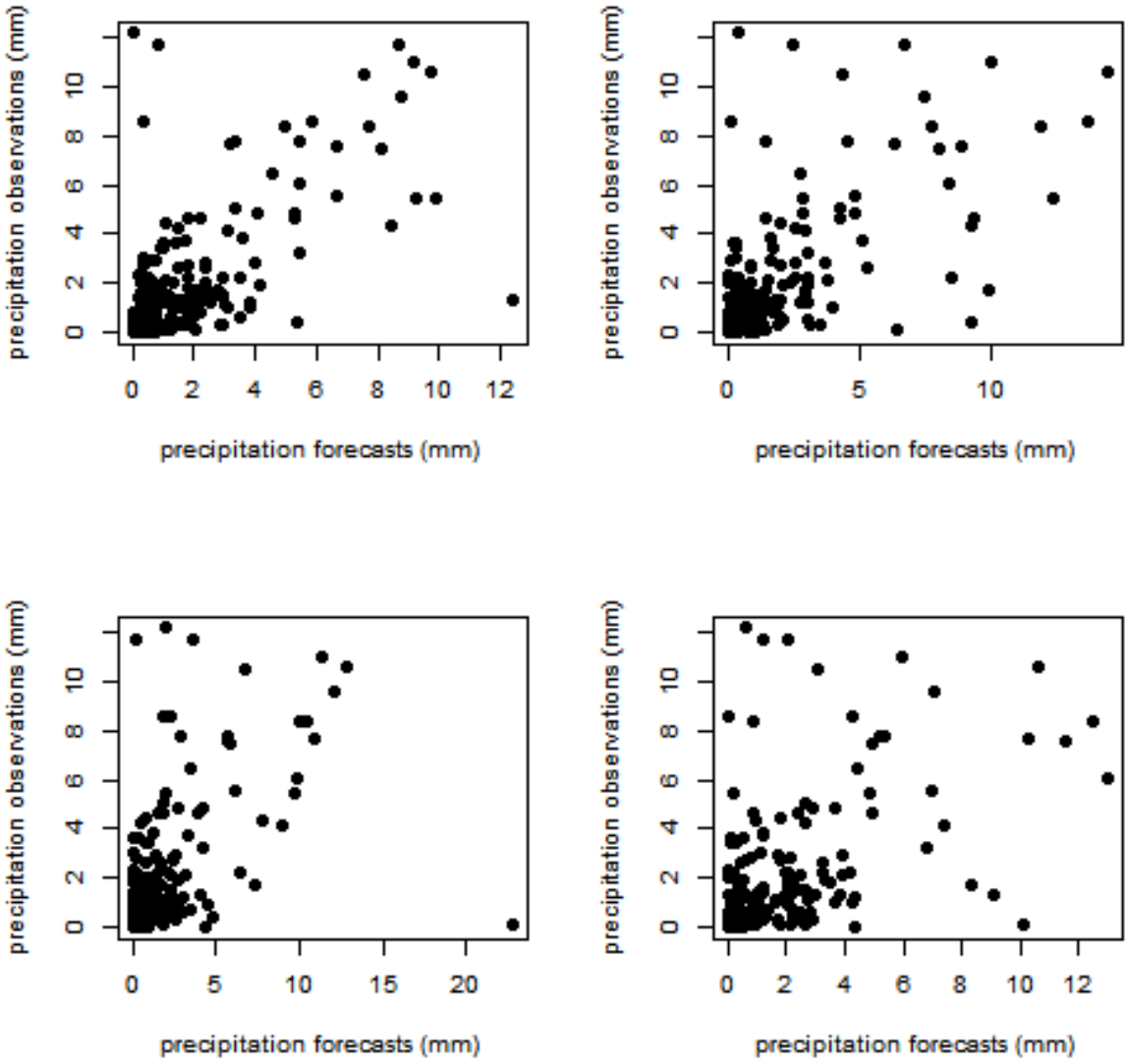

Figure 6.2: Scattered diagrams of area-averaged precipitation against the area-averaged forecast precipitation by HARMONIE for the 15-21 UTC period in Friesland. From left to right and top to bottom the lead times are 3-9, 9-15, 15-21, and 21-27 h respectively.

6.4 Data Analysis

Throughout this thesis, the distribution is assumed to be fat tailed. Therefore it is important to check whether this assumption is justified for the precipitation data. In the non-parametric CST estimator from Chapter 4, the Hill estimator is used for the extrapolation to the tail distribution. It can be problematic when the assumption is not checked, as the Hill estimator is by definition positive. Therefore, if data is from a short tailed distribution, the estimates will have large errors.

There are several tools available in the literature which enable to derive the correct class of tail distributions of the data. In this section three methods will be used to check for the assumption of a heavy tailed distribution. First consider the QQ-plot, or quantile-quantile plot, this plot is often used to check the validity of a distributional assumption.

Consider an ordered dataset $Y_{1,n}, \ldots, Y_{n,n}$ to construct a QQ-plot to visually check whether the sample Y_i for $1 \leq i \leq n$ follows a distribution F. Assume now that the dataset is indeed from the distribution F, in that case ordered observations, or empirical quantiles, should be close to the exact quantiles. For example, in this case it can be expected that the smallest observation $Y_{1,n}$ corresponding to percentile $p_1 = \frac{0.5}{n}$ should be close to $F^{-1}(p_1)$. A general observation $Y_{i,n}$ is expected to lie close to $F^{-1}\left(\frac{i-0.5}{n}\right)$. In a QQ-plot the set of points $\left(Y_{i,n}, F^{-1}\left(\frac{i-0.5}{n}\right)\right)$ are shown. Note that the points lie on the line $y = x$ when the data is from the distribution F.

This result can be formulated a bit more general, by noticing that when the dataset is from $F\left(\frac{x-\mu}{\sigma}\right)$ for some scale $\sigma > 0$ and some location μ the plotted points should still follow a linear line.

In extreme value theory the QQ-plot is useful to determine whether the data is indeed from a heavy tailed distribution, a light tailed distribution or a short tailed distribution. This is determined best by comparing the distribution of the data with an exponential distribution, which is light tailed. An exponential QQ-plot shows then a concave behaviour for heavy tailed data, convex behaviour for short tailed data, and these two behaviours are shown in Figure 6.3.

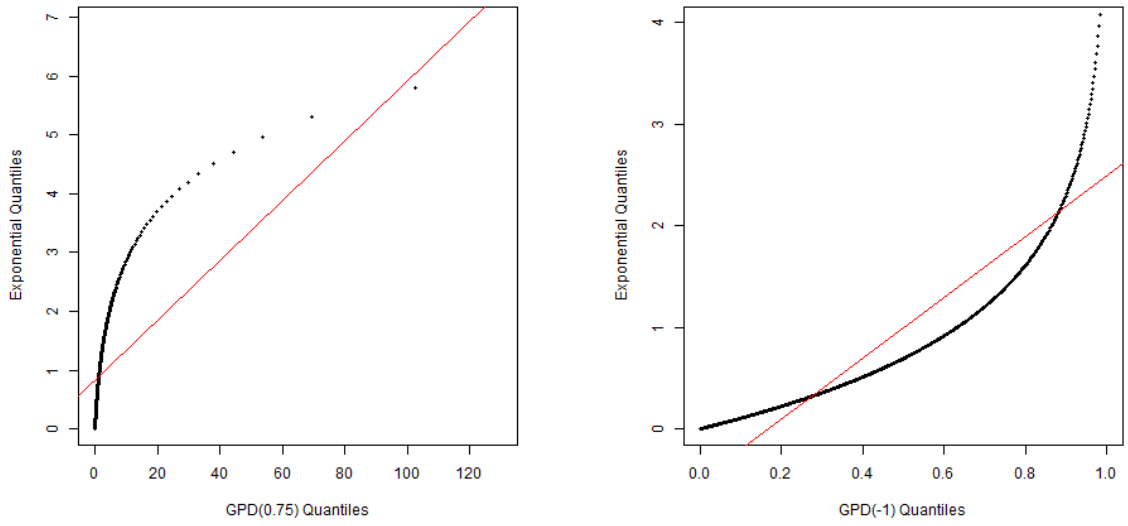

Figure 6.3: The exact QQ-plot with a heavy tailed GPD(0.75), showing concave behaviour (left) and a short tailed GPD(-1), showing convex behaviour (right).

For all the precipitation observations the QQ-Plot is shown in Figure 6.4 on the left side. It

shows the concave behaviour for low values, but for high values there is some convex behaviour visible. Therefore, it is not a clear indication of heavy tailed data. For the non-parametric CST estimator, the heavy tailed assumptions holds for the exceedances above a non-stationary threshold. By correcting for the dependence of the covariates, in this case the NWP output, the QQ-plot could be a lot clearer. For a bandwidth $h = 3$, the 0.8 composite quantile curve is estimated. In the right panel of Figure 6.4 the QQ-plot is shown for the exceedances of the 0.8 quantile estimate; notice the clear concave behaviour in the figure.

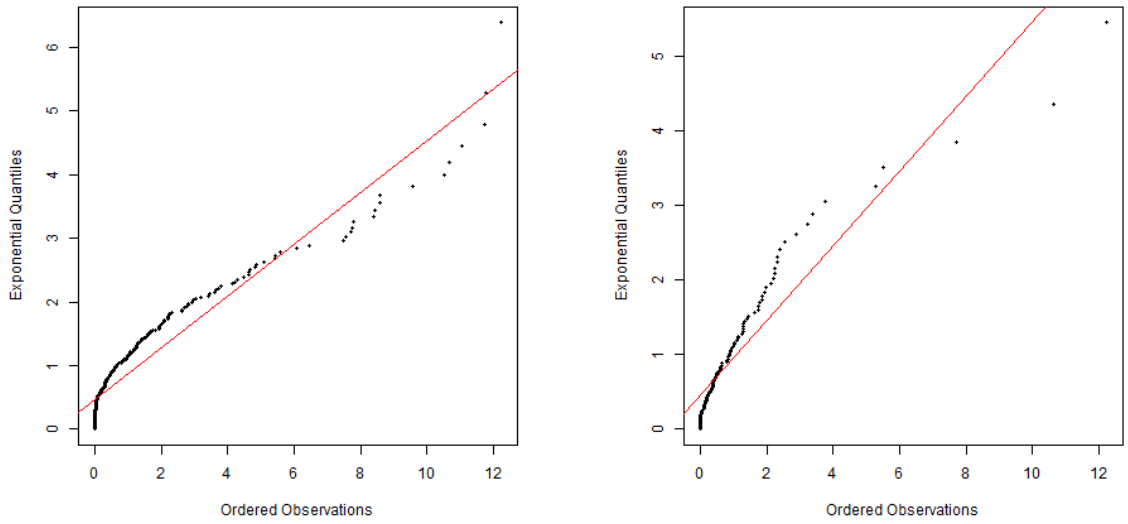

Figure 6.4: The QQ-plot for all the precipitation observations (left) and the QQ-plot for the residuals of the common shape estimator (right).

The second plot that can be considered is the Zipf plot, that is based on the behaviour of a heavy tailed distribution. As seen in Chapter 2, a heavy tailed distribution has asymptotically a power low decay with power $-\frac{1}{\gamma}$. This means that a heavy tailed distribution can be approximated by,

$$1 - F(y) = y^{-1/\gamma}.$$

Taking the log on both sides and plugging in the empirical cumulative distribution function gives the following expression:

$$\log(1 - \hat{F}_n(y)) = -\frac{1}{\gamma} \log y.$$

This relation shows that by plotting the log observations against the log ECDF should result in a linear decreasing line in the case of heavy tailed distributions. In Figure 6.5 the Zipf plot is shown for different distributions to see the behaviour of other distributions.

Figure 6.6 shows on the left side the Zipf plot on all of the precipitation observations and on the right side the exceedances as obtained for the QQ-plot. Both plots show a clear linear decay, i.e. the data has a power-law decay which strengthens the hypothesis that the data is heavy tailed.

In general, the linear decay is not always as good visible as in Figure 6.6. This is due to the fact that the Pareto decay is an asymptotic result which is not necessarily visible from a finite data set. This makes it hard to make the distinction between a power-law decay and a

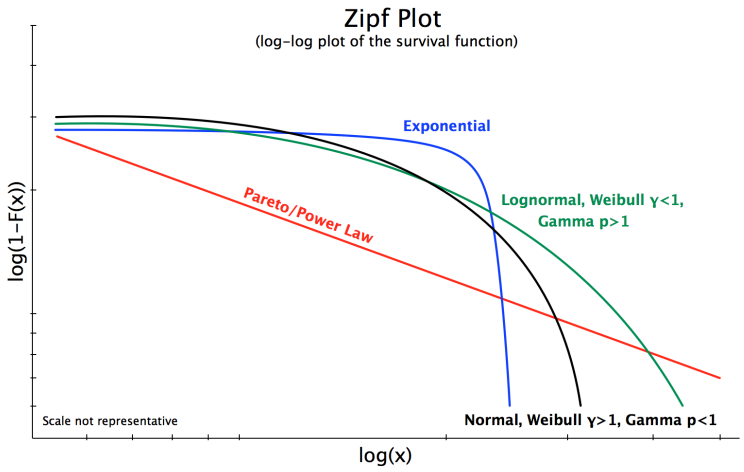

Figure 6.5: The behaviour of the Zipf plot for several distributional families (made by Dr. P. Cirillo).

lognormal distribution. In addition, due to data sparseness in the tail, the most right part of the Zipf plot is not always reliable.

The final plot to be discussed is the mean-excess plot. Let u be a threshold and consider the function which maps this threshold to the expectation of the excesses above this threshold, i.e.

$$e(u) = \mathbf{E}(Y - u|Y > u)$$

For heavy tailed distributions this function is always linear in the threshold. In Figure 6.7 the behaviour of the mean-excess plot is shown for several distributions. Notice that it is easier to reject several distributions based on this plot as the different distributions do not have all the same monotone behaviour.

In Figure 6.8 the mean-excess function is plotted on the left for all the precipitation observations and on the right for only the exceedances.

From both the left and right panel it can be seen that there is an approximately linear increase, which strengthens the hypothesis of heavy tailed data.

The three plots (Figure 6.4, Figure 6.6 and Figure 6.8) clearly justify the heavy tailed assumption of the precipitation data.

6.5 Model fitting

Similar to the simulation in Chapter 5 a comparison will be made between the Common shaped tail estimator, introduced in Chapter 4, the extreme quantile estimator with linear quantile curves by [18] and the extreme quantile curve estimator by [5]. As the true distribution of the data is not known, the comparison will be made using the quantile verification skill score (QVSS) introduced in [10].

First define the Quantile verification score of a quantile estimator $\hat{Q}(\tau|x)$ by,

$$QVS(\hat{Q}(\tau|x)) = \sum_{i=1}^{n} \rho_\tau(Y_i - \hat{Q}(\tau|X_i)),$$

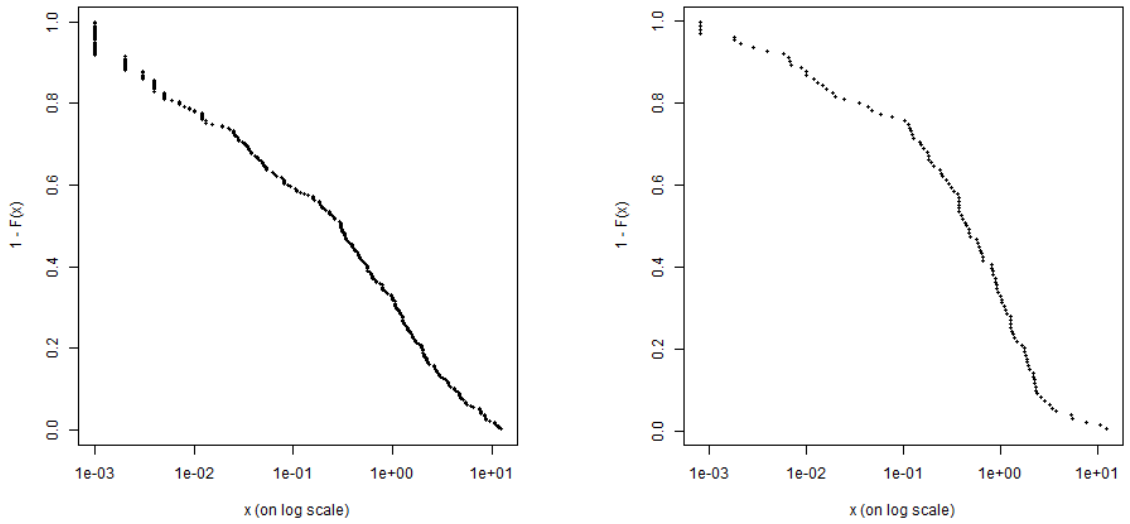

Figure 6.6: The Zipf plot for all the precipitation observations (left) and the Zipf plot for the location, $r(x)$, corrected precipitation observations (right).

where $\rho_\tau(.)$ is the check function introduced in Chapter 2. Observe, that the QVS is given by the quantile loss function which is used in the quantile regression framework. The QVSS is defined relative to the quantile verification score of the empirical quantile as a reference:

$$QVS_{ref} = \sum_{i=1}^{n} \rho_\tau(Y_i - Y_{[n\tau],n})$$

The QVSS for a quantile estimator $\hat{Q}(\tau|x)$ is given by,

$$QVSS(\hat{Q}(\tau|x)) = 1 - \frac{QVS(\hat{Q}(\tau|x))}{QVS_{ref}}$$

Note that a higher QVSS corresponds to a better fit to the data.

In order to verify on independent data, the data will be split up in a part for the model estimation and a part for the model validation, two thirds and one third respectively. As the data consists of three years of data the years 2010 and 2011 will be used to train the models and the year 2013 is used for the validation of the models.

It can be clearly seen from Figure 6.2 that the covariate is not uniformly distributed, which is different from the simulated data in Chapter 5. This causes some problems for the non-parametric methods, as they assume a single bandwidth, which is not optimal for a non-uniform covariate. Therefore an adjusting bandwidth selection method is applied for both the CST estimator in Equation 4.5 and the kernel quantile curves estimator in Equation 3.7. The bandwidth h for a covariate level x^* is chosen as the one which minimizes,

$$h(x^*) = \arg\min_{h \in \mathcal{H}} \frac{1}{\#\{X_i : X_i \in [x^* - h, x^* + h]\}} QVS_h(\hat{Q}^h(\tau_c = 0.8|x^*)),$$

where QVS_h is the quantile verification score based on the model data with the covariate in $[x^* - h, x^* + h]$. Here $\mathcal{H}$ denotes the set of bandwidths over which the expression is minimized, which in this case are 100 equally spaced bandwidth values between 0.1 and 10. This method thus, selects the bandwidth that minimizes the quantile verification score for the training period.

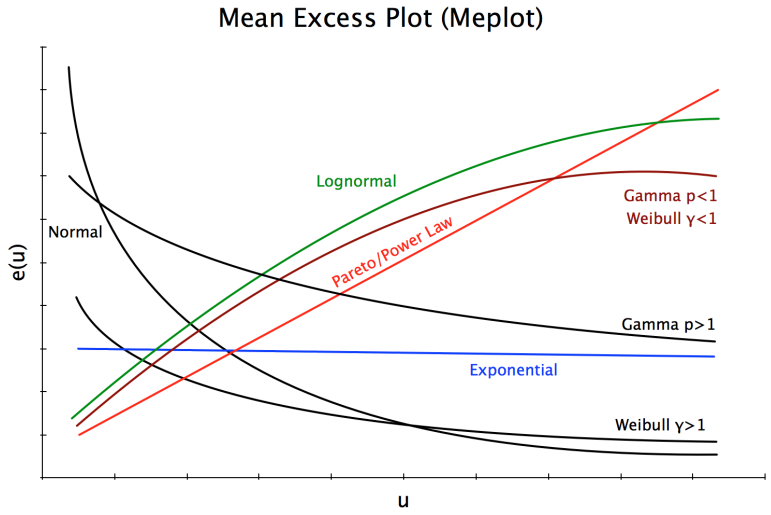

Figure 6.7: The behaviour of the mean excess plot for several distributional families (made by Dr. P. Cirillo).

For the validation the percentiles 0.9, 0.95 and 0.98 are considered. These quantiles are not as extreme as the quantiles in Chapter 5, but in order to do the validation the empirical quantiles need to exist.

The three estimators that are compared are the CST estimator from Equation 4.5, the power transformed linear quantile estimator from Equation 3.2 and the kernel quantile estimator from Equation 3.7. For the CST estimator the threshold percentile is set at $\tau_c = 0.8$, the percentile sequence for the extreme quantile curves estimator is chosen as 10 equally spaced values between 0.8 and 0.99. In all the methods the k is chosen at the point where the function $\hat{\gamma}(k)$ becomes stable.

6.6 Verification results

It can be seen from Table 6.1 that only for the 0.9 quantile the kernel curves estimator is better than the CST estimator for the $3-9$ and $9-15$h lead times. For all the other lead times and quantiles the CST has higher QVSS values. The linear quantile curves estimator is outperformed for every percentile level and lead time by both the CST estimator and the kernel curve estimator.

lead time(h)	τ	CST	linear	Kernel
3-9	0.90	0.540	0.450	**0.541**
	0.95	**0.662**	0.496	0.596
	0.98	**0.779**	0.550	0.653
9-15	0.90	0.519	0.290	**0.635**
	0.95	**0.681**	0.391	0.659
	0.98	**0.816**	0.505	0.636
15-21	0.90	**0.683**	0.278	0.510
	0.95	**0.776**	0.320	0.554
	0.98	**0.861**	0.388	0.571
21-27	0.90	**0.546**	0.200	0.340
	0.95	**0.704**	0.263	0.376
	0.98	**0.835**	0.363	0.414

Table 6.1: The table shows a comparison between the CST estimator, the linear quantile curves estimator and the kernel curves estimator based on the QVSS score. The comparison is made for the four different lead times and the percentiles 0.9, 0.95, 0.98.

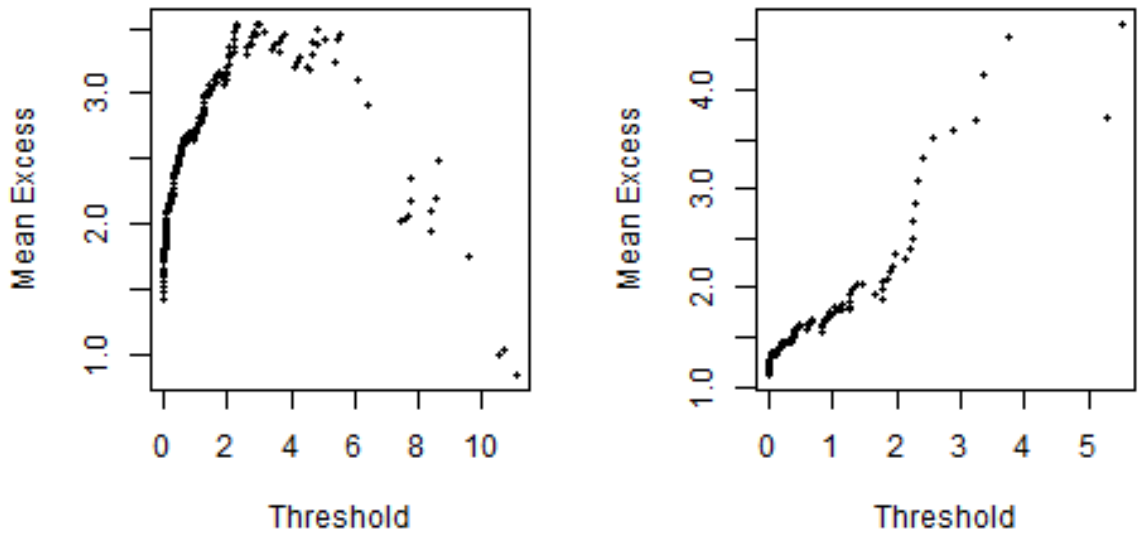

Figure 6.8: The mean-excess plot for all the precipitation observations (left) and the mean-excess plot for the location, $r(x)$, corrected precipitation observations (right).

Chapter 7

Concluding remarks

In practice data may have an extreme value index which does not vary with respect to the covariates, this is seen in Chapter 6 where precipitation observations were considered together with forecasts from a NWP model. The aim of this thesis was to obtain proper tail inference for data where the extreme value index is a constant function over the covariates. This is not a trivial problem and several estimators from the literature are studied, which all take a different approach to this problem. However, it is shown in Chapter 5 that the estimated extreme value index from these estimators varies heavily as a function of the covariates. The freedom in these estimators introduces large uncertainties in the estimations. This motivated the introduction of an estimator that assumes a constant extreme value index over the covariates.

Additionally in quantile regression the problem of quantile crossing arises. When this happens the estimated curves can no longer be interpreted as quantile curves. The main method that is applied in the literature is the monotone rearrangement method in [4], though this removes the smoothness from the quantile curves.

Introducing the common shaped model in Equation 4.1, the common shaped tail (CST) estimator is proposed to improve on the estimators from the literature for tail equivalent data. The CST estimator takes, similar to all estimators a two step approach. First it estimates the common shape, $r(x)$, at several quantiles at the same time, by exploiting the common shape property in the common shaped model. For the estimator a non-parametric version of the composite quantile estimator proposed in [21] was used to model the dependence on the covariate as flexible as possible. The second step continues the estimation process from the residuals of the common shape defined in Equation 4.3. The positive residuals of this estimation can be considered as exceedances above a high threshold, i.e. the estimated common shape. Proposition 4.2.1 and Proposition 4.2.2 show that assuming the data is from a common shaped model, the tail index can be estimated from the residuals.

In Chapter 5 an extensive simulation study is used to make the choices for the percentile level of the shape function and to compare the CST estimator with estimators from Chapter 3. The power transformed linear quantile curves from [18] and the kernel quantile curve estimator from [6] are used in the comparison. From the simulations it is clear that for small and medium sample sizes the CST performs better than the kernel estimator, but in general the power transformed estimator performs better. The bias of the CST estimator appears to be very large compared to the other estimators. This causes bad results in general, but especially for the large sample sizes where the bias dominates the mean squared error. The few times that the CST estimator performs better than both the other estimators is for the small sample sizes. This can be an indication that the variance of the CST estimator is smaller compared to the variance of the other estimators. Below some suggestions are made to reduce the bias of the estimator.

In Chapter 6 a dataset is analysed with area-averaged precipitation observations in the province Friesland in the Netherlands as response and as a covariate the area-averaged precip-

itation forecasts from the NWP model HARMONIE. First, it is shown that the data can be assumed to be from a heavy tailed distribution and then all the three methods are applied to the data to see which has better predictive ability. In order to compare the different methods, the data was split up in the years 2010 and 2011 for the model training and 2013 for the verification. The CST estimator has the best predictive abilities based on this dataset and the quantile verification skill score.

It can be concluded that the method has potential in applications where weather data is post-processed. Based on the small data example the predictive power of the CST estimator is substantially larger than the predictive power of the other estimators.

7.1 Future research

The results from Chapters 5 and 6 provide motivation for future research on the CST estimator. As stated in the conclusion, the properties of the estimator are shown by means of an extensive simulation study, but the properties still need to be mathematically proven. The setting in Equation 4.2 is general from an extreme value theory perspective, as it does not pose assumptions on the sign of the extreme value index. Though a heavy tail is assumed in this thesis, the peaks over threshold method also works in the short tailed setting, which would allow to extend to the general domain of the extreme value distribution. In the current estimator this is already implicitly assumed by the usage of the maximum likelihood estimator, which is consistent for $\gamma > -\frac{1}{2}$. The reason this estimator was chosen came from the fact that the Hill estimator was not shift invariant, which resulted in large errors in the estimation procedure.

In the conclusions it was already noted that the bias of the CST estimator was substantially larger compared to the other estimators. This caused large errors and bad performance compared to the other estimators. Here two improvements are proposed to reduce the bias in the estimator. The first improvement is suggested for the boundary issues which occur in the non-parametric composite quantile estimation. The estimates at the boundary are equally important as the estimates in the middle as all residuals are used in the extrapolation to the extreme quantiles. Future research should therefore be focussed on removing the boundary issues for the composite quantile estimator. Other improvements could be the introduction of kernel weighting of the different observations to obtain smooth quantile curve estimates or the weighting of the different quantile levels as the sparseness in the higher quantiles may influence the estimates too much.

A second improvement can be made in the extrapolation to the extreme quantiles. The maximum likelihood estimator does a reasonably good job in the estimation of the extreme value index, but the estimation setting of the two step procedure might produce better results with a different extreme value index estimator. The residuals can be considered as contaminated observations and tail estimation from a contaminated dataset might benefit from a different tail index estimator. Future research has to reveal which estimator would be best in this estimation framework.

The selection of the bandwidth is a very important part which needs further investigation. In our research a cross validation approach is tried, but did not yet give satisfying results. The extension to a framework where d dimensional covariates are considered can be obtained in a similar way.

From a meteorological point of view it would be interesting to compare the results of the CST estimator on the precipitation data from Chapter 6 with ensemble forecasts. In Chapter 6 it was explained that the ECMWF ensemble quantifies the uncertainty from the initial conditions and the parametrizations by perturbing the initial condition and parameters in 50 different ways and considering the empirical distribution of these 50 weather forecasts. A statistical post-processing approach, can correct for systematic errors in the mean and spread of the ensemble. Therefore it

should have better predictive power, maybe also compared to the larger ensemble members using the estimator proposed in this thesis. This would be an interesting topic for future research.

As this MSc thesis is a precursor to the PhD research with title "Probabilitic forecasts of extreme weather utilizing advanced methods from extreme value theory", a rough outline will be sketched of future research which can be done in the coming four years.

A first question that would be of interest is how to verify extreme quantile predictions. In Chapter 6 the verification is done for the percentiles 0.9, 0.95 and 0.98, because 0.98 is the highest possible percentile in the verification dataset ($n = 88$). This means that there should be exactly one observation above the 0.98 quantile curve. Naturally the verification based on a single observation will cause such high uncertainties that it is not reliable in practice any more. Therefore it would be interesting to develop methods that would enable the verification of extreme quantiles which are in the range of the data, i.e intermediate quantiles, or even exceed the data, i.e. extreme quantiles.

Secondly it would be interesting, when the model is expanded to d dimensional covariates, to develop variable selection methods based on the extremes. Most variable selection methods select the variables which have the largest explanatory power for the middle part of the distribution. It would therefore be interesting to develop methods, which select variables based on the explanatory power for the extremes.

Furthermore it is interesting to consider not only one response variable, but two instead. The modelling of the bivariate tail distribution conditioned on covariates will create insight in the bivariate risks of concurrent extreme weather events. An application of the modelling of these combined weather risks is the railway system. In the Netherlands the railway system is fairly complex due to the high density of cities especially in the west. The combination of snow and low temperatures results in ice on the tracks. This could potentially result in delays for the traveller if proper precautions have not been taken. Skillfull forecasting of the bivariate risks of low temperature and snowfall will enable a company as ProRail to take proper actions in time.

All the approaches above are focussed on the extremal behaviour of Y given some covariate x. However, in the example of a weather alert system that is considered in the introduction, i.e. issue an alarm if a threshold is expected to exceed with a specified probability, the probability of exceedance is generally quite large. For code orange issued by KNMI this probability is set at 60%. So given the prediction from the NWP model, it would be interesting to model the middle part of the distribution. This concerns modelling $Y|X$ in a part where both the observations Y and covariates X are sparse. Some parametric assumption on the relationship between Y and X could be assumed here, which would result in better estimates in these data sparse parts with respect to X.

Acknowledgements

This thesis is written at the Delft University of Technology in collaboration with KNMI. First, I would like to thank my daily supervisor Juan-Juan Cai for providing advice and guidance in every aspect of this thesis. I also like to thank Geurt Jongbloed for the input and clarifications during our monthly meetings. From the KNMI I thank Maurice Schmeits and Kirien Whan for providing useful feedback from the meteorological perspective and for proof reading my thesis.

Special thanks to Pim Otte, who helped me in obtaining a slightly deeper understanding of the optimization methods used in my research.

Finally, I thank Emma, Rune, Carli, Jason, Anton, Jason, Mirrelijn, Daniëlle, Rik, Hassan and all others I forgot, for providing the social aspect during my thesis in all the conversations, discussions, coffee breaks and the occasional table football matches.

Bibliography

[1] Ian Barrodale and Frank DK Roberts. An improved algorithm for discrete l_1 linear approximation. *SIAM Journal on Numerical Analysis*, 10(5):839–848, 1973.

[2] Jan Beirlant, Tertius De Wet, and Yuri Goegebeur. Nonparametric estimation of extreme conditional quantiles. *Journal of statistical computation and simulation*, 74(8):567–580, 2004.

[3] George EP Box and David R Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252, 1964.

[4] Victor Chernozhukov, Iván Fernández-Val, and Alfred Galichon. Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125, 2010.

[5] Abdelaati Daouia, Laurent Gardes, Stéphane Girard, et al. On kernel smoothing for extremal quantile regression. *Bernoulli*, 19(5B):2557–2589, 2013.

[6] Abdelaati Daouia, Laurent Gardes, Stéphane Girard, and Alexandre Lekina. Kernel estimators of extreme level curves. *Test*, 20(2):311–333, 2011.

[7] Anthony C Davison and Richard L Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 393–442, 1990.

[8] Holger Drees, Ana Ferreira, and Laurens de Haan. On maximum likelihood estimation of the extreme value index. *The Annals of Applied Probability*, 14(3):1179–1201, 08 2004.

[9] Ronald Aylmer Fisher and Leonard Henry Caleb Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press, 1928.

[10] P Friederichs and A Hense. Statistical downscaling of extreme precipitation events using censored quantile regression. *Monthly weather review*, 135(6):2365–2378, 2007.

[11] Boris Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453, 1943.

[12] Bruce M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 09 1975.

[13] Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

[14] CJ Kok, BGJ Wichers Schreur, and DHP Vogelezang. Meteorological support for anticipatory water management. *Atmospheric Research*, 100(2):285–295, 2011.

[15] Yunming Mu and Xuming He. Power transformation toward a linear regression quantile. *Journal of the American Statistical Association*, 2012.

[16] Aart Overeem, Iwan Holleman, and Adri Buishand. Derivation of a 10-year radar-based climatology of rainfall. *Journal of Applied Meteorology and Climatology*, 48(7):1448–1463, 2009.

[17] Y Seity, P Brousseau, S Malardel, G Hello, P Bénard, F Bouttier, C Lac, and V Masson. The arome-france convective-scale operational model. *Monthly Weather Review*, 139(3):976–991, 2011.

[18] Huixia Judy Wang and Deyuan Li. Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association*, 108(503):1062–1074, 2013.

[19] Huixia Judy Wang, Deyuan Li, and Xuming He. Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107(500):1453–1464, 2012.

[20] Guosheng Yin, Donglin Zeng, and Hui Li. Power-transformed linear quantile regression with censored data. *Journal of the American Statistical Association*, 103(483):1214–1224, 2008.

[21] Hui Zou and Ming Yuan. Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, pages 1108–1126, 2008.

Appendices

Appendix A

An extension to the CST estimator

The common shaped tail estimator exploits the properties of the common shaped model introduced in Chapter 4 and given by,

$$Q(\tau|x) = r(x) + \epsilon(\tau),\ \epsilon(\tau_c) = 0. \tag{A.1}$$

The main property that is exploited by the common shaped tail estimator is the fact that the quantile curves are parallel. Recall that the quantile shape estimator for $r(x)$ is based on a percentile sequence $\tau_c = \tau_1, \dots \tau_p$ and at a fixed covariate level x^* the estimate of $r(x^*)$ is given by $\hat{\alpha}_1$ in the following minimization,

$$\underset{\alpha_1,\dots,\alpha_p,\beta_1,\dots\beta_m}{\arg\min} \sum_{j=1}^{p}\sum_{i=1}^{n} \mathbf{1}(|X_i - x^*| \leq h)\rho_{\tau_j}\left(Y_i - \alpha_j - \sum_{k=1}^{m}\beta_k(x - x^*)^i\right) \tag{A.2}$$

Though the estimator exploits the fact that the quantile curves are parallel, the estimates will not be exactly parallel. Similar to the quantile crossing problem, this is caused by the fact that for two distinct covariate levels x_1^* and x_2^* the parameters are fitted independently of each other.

To clarify consider two percentile levels x_1^* and x_2^* and let $\alpha_1^1, \dots \alpha_p^1$ denote the intercepts at x_1^* and $\alpha_1^2, \dots \alpha_p^2$ at x_2^*. Rewriting these parameters, they can be considered relative to the curve $r(x)$, define $d_i^1 = \alpha_i^1 = \alpha_1^1$ and $d_i^2 = \alpha_i^2 - \alpha_1^2$. These can then be interpreted as the distances between the quantile curves of levels τ_c and τ_i. Because model A.1 assumes parallel quantile curves, it must hold that $d_i^1 = d_i^2$ for all $1 \leq i \leq p$. Note this is not is not necessarily true as the estimates are based on a different subset of the n observations.

In this appendix an idea is proposed to exploit Model A.1 even more by forcing the quantile curves to be parallel.

Parallel quantile curves estimation

The distances, defined above, can be written as the difference of the quantile function $\epsilon(\tau)$.

$$\begin{aligned} d_1 &= Q_{Y|x}(\tau_1|x) - Q_{Y|x}(\tau_1|x) &&= \epsilon(\tau_1) - \epsilon(\tau_1) \\ &\vdots && \vdots \\ d_p &= Q_{Y|x}(\tau_p|x) - Q_{Y|x}(\tau_m|x) &&= \epsilon(\tau_p) - \epsilon(\tau_1) \end{aligned}$$

In order to maintain the same distance between the curves, the optimization needs to be performed over a sequence of covariate levels $x_1^*, \dots, x_m^*$. For simplicity from now on everything

will be written for a local linear expansion, though it would be straightforward to write it in terms of pth order expansion. Hence the optimization problem can be written by,

$$\underset{r_k,d_j,\beta_k^i}{\arg\min} \sum_{k=1}^{m}\sum_{j=1}^{p}\sum_{i=1}^{n} \mathbf{1}(|X_i - x_k^*| \leq h)\rho_{\tau_j}\left(Y_i - r_k - d_j - \beta_k(X_i - x_k^*)\right) \tag{A.3}$$

For each grid point x_k^* the estimated point on $r(x)$ is given by $\hat{r}_k$ and extrapolation can be based on $d_1, \ldots, d_p$. The strength of the estimator is that it combines the infomation in a semi-parametric way. First non-parametrically it determines the local dependence between the response between Y and X and then it uses global information, i.e. the distances d_j, to improve on this local relationship by imposing the parallel constraint.

This is exactly where the difficulty of the model comes in. As can be noted, the model contains two intercepts, i.e. r_k and d_j, which makes the estimation procedure extremely hard to perform at once. In the research for this thesis a proper minimization method is not found and will be interesting for future research.

Though an iterative optimization scheme was tried, which did not give satisfying results. The taken approach estimated a first iterate for $\hat{r}(x)$ at the grid-points $x_1^*, \ldots, x_m^*$, this iterate will be denoted by $\hat{\bar{r}}_1 = (\hat{r}_1^1, \ldots, \hat{r}_m^1)$. From this estimate iteratively the parameters d_j can be estimated. Then from the residuals of $\hat{d}_j$ a new iterate $\hat{\bar{r}}_2$ can be obtained. The algorithm is given as follows,

1. Obtain an initial estimate $\hat{\bar{r}}_0 = (\hat{\alpha}_1^1, \ldots, \hat{\alpha}_1^m)$ by solving for every x_k^*, $1 \leq k \leq m$, the non-parametric composite quantile minimization problem,

$$\underset{\alpha_1,\ldots,\alpha_p,\beta_1,\ldots\beta_m}{\arg\min} \sum_{j=1}^{p}\sum_{i=1}^{n} \mathbf{1}(|X_i - x_k^*| \leq h)\rho_{\tau_j}\left(Y_i - \alpha_j^k - \beta^k(x - x^*)^i\right)$$

2. Let x_q^* be defined, such that it minimizes $|X_i - x_k^*|$ for $1 \leq k \leq m$. Then the residuals of iterate $\hat{\bar{r}}_l$ is given by,

$$E_i = Y_i - \hat{r}_q^l - \hat{\beta}_q^l(X_i - x_q^*)$$

3. estimate d_j corresponding to τ_j by calculating the τ_jth empirical quantile by minimizing

$$\underset{d_j}{\arg\min} \sum_{i=1}^{n} \rho_{\tau_j}(E_i - d_j)$$

4. Estimate a new iterate $\hat{\bar{r}}_{l+1}$ by the following minimization for each grid point x_k^*

$$\underset{r_k,\beta_k^i}{\arg\min} \sum_{j=1}^{p}\sum_{i=1}^{n} \mathbf{1}(|X_i - x_k^*| \leq h)\rho_{\tau_j}\left(Y_i - \hat{d}_j - r_k - \beta_k(X_i - x_k^*)\right)$$

5. Repeat from step 2 until step 4 until the estimates of $\hat{\bar{r}}_l$ and d_j^l converge.

Note that in step 4 the quantile curves at percentile levels $\tau_1, \ldots, \tau_p$ are estimated at the same time by estimating one curve $r(x)$ and correcting the percentiles $\tau_2, \ldots, \tau_p$ by correcting for $\hat{d}_j$. For the implementation of step 1 the composite quantile estimator in [21] is used and for steps 2 and 3 the implementation is straightforward. Though existing quantile regression methods were not sufficient for step 4. The implementation for this step was done by a simplex method which was proposed in [1].

Appendix B

R-code

Non-parametric quantile estimator using a kernel estimator

Description

Estimates the intermediate quantiles for a sequence of percentiles using the estimator in Equation 3.5

Usage

kernel.intermediate(y,x,xstar,taus,h)

Arguments

y	a column vector of response variables
x	a column vector with covariates
xstar	a column vector of covariate values for which to calculate the intermediate quantiles
taus	an increasing percentile sequence for which to estimate the quantiles
h	the bandwidth defining the width of the neighbourhood for the estimates around xstar

Output

A list with the intermediate quantiles and the corresponding percentile sequence.

Bandwidth selection for the extreme kernel curves estimator

Description
Select a bandwidth for the extreme kernel curves estimator using a cross validation approach

Usage
kernel.bandwidth(y,x)

Arguments
y	a column vector of response variables
x	a column vector with covariates

Output
The selected bandwidth h.

Non-parametric extreme quantile estimator extrapolating using the kernel quantile curve estimator

Description
Estimates the extreme quantiles by extrapolating the quantiles from the kernel.intermediate function using the estimator in Equation 3.7

Usage
kernel.intermediate(y,x,xstar,taus,h)

Arguments
y	a column vector of response variables
x	a column vector with covariates
xstar	a column vector of covariate values for which to calculate the intermediate quantiles
taus	an increasing percentile sequence for which to estimate the quantiles
h	the bandwidth defining the width of the neighbourhood for the estimates around xstar

Output
A list with the intermediate quantiles and the corresponding percentile sequence.

Non-parametric quantile estimator using quantile regression

Description

Estimates the intermediate quantiles using the non-parametric quantile estimator from Equation 3.3

Usage

nonpar.quant.intermediate(y,x,xstar,h)

Arguments

y	a column vector of response variables
x	a column vector with covariates
xstar	the covariate values for which to calculate the quantiles.
h	the bandwidth defining the width of the neighbourhood for the estimates around xstar

Output

A list with the estimated quantiles and the estimated percentiles for every covariate xstar.

Non-parametric extreme quantile estimator using quantile regression

Description

Estimates the extreme quantiles by extrapolating the estimated quantiles from the function nonpar.quant.intermediate using the estimator in Equation 3.4

Usage

nonpar.quant.extrapolate(k,quant,tau,tau.e)

Arguments

k	the number of upper order statistics used for the extrapolation
quant	a list where each element consists of the estimated quantiles from the nonpar.quant.intermediate function
tau	a list where each element corresponds to the percentile sequence for the elements in quant
tau.e	a column vector of extreme percentiles

Output

A list with the estimated extreme value index and the extreme quantile estimates.

Bandwidth selection method for the non-parametric quantile regression estimator

Description

Selects the optimal bandwidth for the non-parametric quantile estimator in Equation 3.4 using a bootstrapped approach

Usage

nonpar.quant.bandwidth(y,x,xstar,tau.e)

Arguments

y	a column vector of response variables
x	a column vector with covariates
xstar	a covariate value for which to select the bandwidth
tau.e	the extreme percentile for which to select the bandwidth

Output

The selected bandwidth h.

Point estimator for the shape estimator of the CST estimator

Description

Point estimator for $r(x)$ in the Common Shaped Tail estimator of Equation 4.5

Usage

local.quantile(y,x,h,xstar,tau.c)

Arguments

y	a column vector of response variables
x	a column vector with covariates
h	the bandwidth defining the neighbourhood of estimation around xstar
xstar	the value of the covariate for which a point estimate is wanted at the tau.c quantile
tau.c	the percentile levels at which to obtain a quantile estimate

Output

A list with a column vector containing the quantiles and the corresponding percentiles tau.c and above.

The CST estimator

Description

Estimates an extreme quantile for using the CST estimator specified in Equation 4.5

Usage

CST.fit(y,x,h,tau.c,tau.e,k,xstar=NULL)

Arguments

y	a column vector of response variables
x	a column vector with covariates
h	the bandwidth defining the width of the neighbourhood for the estimates around xstar
tau.c	the percentile levels at which to obtain a quantile estimate
tau.e	a column vector with extreme percentiles bigger than tau.c
k	number of upper order statistics that need to be used for the extrapolation
xstar	the covariate values for which to calculate the quantiles. If not specified only the shape function for the values of x are calculated

Output

A list with the shape function for x and xstar and the extreme quantile estimates.