ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master thesis Econometrics

A beta kernel scedasis estimator for extreme quantile prediction in the heteroscedastic extremes model*

Author

M. van der Linden

(441844)

Supervisor

A.J. Koning

Second assessor

C. Zhou

April 26, 2022

Abstract This thesis explores the heteroscedastic extremes model for independent but non-identically distributed observations in which the so-called scedasis function governs the relative frequency of extremes over time. We propose a non-parametric beta kernel scedasis estimator, of which the compact support matches that of the scedasis function, and which has no boundary bias problem. Our focus lies on extreme quantile prediction in the heavy-tailed case and we prove, under intuitively plausible assumptions, asymptotic normality of the scedasis estimator and the prediction, which is verified by a simulation study. While our estimator generally does not outperform a convolution-type boundary kernel estimator, it proves to be valid and potentially useful. We further demonstrate both estimators on financial loss data of the S&P 500 leading up to the financial crisis of 2008.

*The content of this thesis is the sole responsibility of the author and does not reflect the view of the supervisor, second assessor, Erasmus School of Economics or Erasmus University.

Contents

List of symbols and abbreviations

$\sim$	$a(x) \sim b(x)$ as $x \to y$ means $\lim_{x \to y} a(x)/b(x) = 1$
O	of the same asymptotic order
o	of strictly smaller asymptotic order
O_p	bounded in probability with
o_p	converges in probability with
γ	extreme value index
$\hat{c}_{n,b}$	beta kernel scedasis estimator
$\hat{c}_{n,h}$	convolution-type kernel scedasis estimator
$\mathbb{1}_A$	indicator function of set or event A
$\mathbb{E}$	expectation
$\mathbb{P}$	probability
$\mathbb{V}$	variance
b	smoothing bandwidth for beta kernel estimator
C	integrated scedasis function
c	scedasis function
F	cumulative distribution function
$F_{n,i}$	cumulative distribution function of observation i
h	smoothing bandwidth for convolution-type kernel estimator
k	number of upper order statistics used in extreme value statistics
K_b	beta kernel function $K_{1/b+1,1}$ at the boundary
U	left-continuous generalised inverse of $(1/(1-F))$: the upper tail quantile function
$U_{n,i}$	upper tail quantile function of observation i
$X_{n,j}$	j-th order statistic
EVT	extreme value theory
MDA	maximum domain of attraction

1 Introduction

The impact of the highly improbable, according to best selling author Nassim Nicholas Taleb, is what we at the individual level, as society, or even as a species should pay particular attention to (Taleb, 2007). It stresses the importance of protection against rare but extreme events. For instance, the Dutch government requires that the height of dikes is such that the probability of a flood in a given year is once in 10,000 (de Haan and Ferreira, 2006), protecting against a highly improbable but consequential event. In the context of financial risk management, the regulatory capital for the market risk of banks is determined in such a way that a loss is not exceeded once in 100 (McNeil et al., 2015; Nolde and Zhou, 2021). Extreme value theory (EVT) provides a sound framework for statistical inference on such extremely low probability events in the tail region of a distribution.

Classical extreme value theory assumes that the observations are independent and identically distributed (iid). However, this assumption might be violated in many applications. For example, the observations may be serially correlated or cluster with respect to their variance, as in many financial time-series. A discussion of such situations is given by, among others, Leadbetter et al. (1983), Hsing (1991), Kearns and Pagan (1997), Drees (2000), and a more recent review by McElroy (2016). In general, these studies conclude that most classical EVT statistics are still applicable, but the standard asymptotic statistics can be misleading.

The case of independent but non-identically distributed observations has gained more attention in recent literature. The probability distributions of extremes can be characterised by a location, scale, and shape parameter. Smith (1989) and Davison and Smith (1990) model covariates explicitly in the scale and shape parameters of these distributions, which is a pure parametric approach. However, the shape parameter, referred to as the extreme value index, is often considered as the most relevant in governing the tail behavior. This allows for semi-parametric approaches specifically focused on the extreme value index. Therefore, for instance, Wang and Tsai (2009), Gardes and Girard (2010), and Goegebeur et al. (2014) propose conditional models to relate the extreme value index to covariates, whereas de Haan and Zhou (2021) develop a methodology to model a non-parametric trend in this parameter.

Einmahl et al. (2016) introduce the concept of heteroscedastic extremes to model non-identically distributed observations. In this semi-parametric framework, the proportionality of the tail distributions of different observations is governed over time by a deterministic function, called the scedasis function, while the extreme value index remains fixed. The scedasis function is left unspecified under the assumption that it has support $[0, 1]$ and integrates to one, which allows it to be interpreted as the relative frequency of extreme events over time. A uniform scedasis function then resembles classical iid EVT,

or homoscedastic extremes. For a positive extreme value index, which is for heavy-tailed distributions, Einmahl et al. (2016) propose estimators for the scedasis function, extreme value index, and for extreme quantile prediction. They establish the asymptotic behavior of these statistics and propose tests for the scedasis function and for the stability of the extreme value index over time. To increase the testing power for specific scedasis trends, Mefleh et al. (2020) extend the model by specifying parametric densities for the scedasis function in terms of so-called exceedance times. These exceedance times are shown to be asymptotically independent of the values of the exceedances, shedding new light on the results by Einmahl et al. (2016). Further, de Haan et al. (2015) impose a similar tail proportionality condition, but the corresponding parametric scedasis trend functions are not restricted to be densities.

To estimate the scedasis function, Einmahl et al. (2016) propose an adaptation of the non-parametric, standard convolution-type kernel estimator, as discussed, for instance, in Silverman (1986) and Wand and Jones (1994). Proper estimation of the scedasis function at the most recent observation, i.e. at its right endpoint, is essential for extreme quantile prediction. However, standard symmetric convolution-type kernel estimators are inconsistent in the boundary region for densities with compact support: a problem often referred to as 'boundary bias' or 'boundary effects'. Therefore, Einmahl et al. (2016) use a boundary kernel, as introduced by Jones (1993), in this case.

Chen (1999) suggest using a beta density as kernel function for densities with compact support. This density estimator is adaptive and asymmetric, as the shape and therefore the amount of smoothing of its kernel varies over the support without explicit adjustment of the smoothing bandwidth. Chen (1999) show that this estimator has no boundary bias problem. Further, because the support of the beta kernel matches that of the density to be estimated, the effective sample size is larger. Chen (1999) find that this can lead to smaller finite-sample variance than the convolution-type boundary kernel estimators by Jones (1993) and by Jones and Foster (1996).

In our research, we propose a beta kernel estimator for the scedasis function in the heavy-tailed heteroscedastic extremes model. Given the positive characteristics of beta kernel estimators outlined above, we are interested in its properties for extreme quantile prediction, both individually and in comparison to the convolution-type boundary kernel scedasis estimator.

Our asymptotic analysis establishes the asymptotic normality of the beta kernel scedasis estimator at the right boundary point and of the corresponding extreme quantile prediction under intuitively plausible assumptions. It further suggests an optimal mean squared error (MSE), balancing bias and variance with respect to the smoothing bandwidth parameter of our estimator, which is of larger asymptotic order than that of the convolution-type boundary kernel estimator. A Monte Carlo simulation study verifies our asymptotic results. It also shows that our estimator generally does not outperform

a convolution-type biweight boundary kernel estimator in terms of its optimised MSE for various sample sizes and data generating processes: the larger effective sample size does not lead to important finite-sample performance gains. Finally, we apply the beta and convolution-type kernel scedasis estimators on a sample of financial loss data of the S&P 500, which Einmahl et al. (2016) argue to be suitable for heteroscedastic extremes, progressively zooming into the emerging financial crisis of 2008. The results by both estimators are generally close together and suggest higher extreme quantile predictions than the traditional homoscedastic model and the empirical estimator.

Our research differs from existing literature in the following ways. Whereas Chen (1999) derive the asymptotic bias and variance of the beta kernel density estimator, which are similar to our results, we focus on scedasis estimation solely at the right boundary point and additionally provide the asymptotic distribution. Further, in contrast to the convolution-type boundary kernel considered by Einmahl et al. (2016), which has similar asymptotic results to ours, the beta kernel is non-negative and matches the support of the scedasis function. Moreover, rather than giving a formal proof, we assume and give intuitive arguments for asymptotic independence of the scedasis and extreme value index estimators to establish asymptotic normality of the extreme quantile prediction. Contrasted to Chen (1999) and Einmahl et al. (2016), we also provide a discussion on the local optimal bandwidth and convergence. Furthermore, while our simulation design is similar to Einmahl et al. (2016), it considers a greater variety of sample sizes and data generating processes to compare the scedasis estimators. Finally, our empirical application, while considering the same data as Einmahl et al. (2016), additionally focuses on extreme quantile prediction and on subsamples of the data.

In conclusion, our beta kernel scedasis estimator has clear asymptotic statistical properties for extreme quantile prediction which are in accordance with existing literature. While its performance is generally not better than that of the convolution-type boundary kernel estimator, it is often similar in finite-sample applications. We suggest that research into data-driven selection of the bandwidth parameter and number of upper order statistics for this non-parametric approach to the heteroscedastic extremes model may provide further, likely more realistic insights for its potential to applications in practice.

Our report proceeds as follows. We first provide an overview of the existing theory related to our research, where Section 2 discusses classical EVT, Section 3 the heteroscedastic extremes model, and Section 4 kernel density estimation. While these subjects deserve extensive studies on their own, our overview may still be more elaborate than strictly necessary. However, we believe that this may provide greater understanding of the subject matter, which is unfamiliar to many. In Section 5 we introduce our beta kernel scedasis estimator and its asymptotic properties, together with a discussion on practical issues and the proofs. Section 6 and 7 present the Monte Carlo simulation and empirical application, respectively. We end with our concluding remarks in Section 8.

2 Classical extreme value theory

In this section we discuss classical EVT, in which observations are assumed to be independent and identically distributed (iid), and some of the corresponding statistics. It closely follows Embrechts et al. (2013) and de Haan and Ferreira (2006), to which we refer for extensive reviews on the subject matter.

2.1 Setup

Consider the $n \in \mathbb{N}$ iid random variables X_i, for $i = 1, \ldots, n$, with cumulative distribution function F and right endpoint $x^* \coloneqq \{x \in \mathbb{R} : F(x) < 1\}$. These random variables could for instance, in the context of financial risk management, correspond to losses, and in the context of water management, to sea levels.

For the sums $S_n \coloneqq X_1 + \ldots + X_n$, when properly centered and normalised, it is known that the α-stable laws, which include the Gaussian distribution, are the only possible limiting distributions as $n \to \infty$. Similar to sums of random variables, the maxima $M_n \coloneqq \max(X_1, \ldots, X_n)$, when properly centered and normalised, can only have one of three limiting distributions: the extreme value distributions. For iid random variables it follows that

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, \ldots, X_n \leq x) = F^n(x), \quad \text{for } x \in \mathbb{R},$$

where $\mathbb{P}$ denotes the probability operator. Then, defining the sequences of constants $c_n > 0$ and $d_n \in \mathbb{R}$, for all n, we get

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - d_n}{c_n} \leq x\right) = \lim_{n\to\infty} F^n(c_n x + d_n) \coloneqq H(x), \tag{2.1}$$

where H is some non-degenerate distribution. When (2.1) holds for some H, F is said to belong to the maximum domain of attraction (MDA) of H, or $F \in \mathrm{MDA}(H)$. The Fisher-Tippett-Gnedenko theorem states that if $F \in \mathrm{MDA}(H)$, H must be a generalised extreme value (GEV) distribution with standard distribution function

$$H_\gamma(x) \coloneqq \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right) & \text{if } \gamma \neq 0, \\ \exp\left(-\exp(-x)\right) & \text{if } \gamma = 0, \end{cases} \tag{2.2}$$

where $1 + \gamma x > 0$. To include a location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$, one could define the three-parameter location-scale family $H_{\gamma,\mu,\sigma}(x) \coloneqq H_\gamma((x-\mu)/\sigma)$.

The parameter γ is a shape parameter, often called the extreme value index, which allows to distinguish the three types of extreme value distributions: Gumbel for $\gamma = 0$,

Fréchet for $\gamma > 0$, and Weibull for $\gamma < 0$. For the Weibull distribution, the right endpoint x^* is finite, while for the other two extreme value distributions $x^* = \infty$. Further, the decay of the (right) tail of the Fréchet distribution is much slower than that of the Gumbel distribution. Therefore, if F belongs to the Fréchet MDA, F is said to be a heavy-tailed distribution.

The following basic result in EVT gives the essential information on the MDA: $F \in \text{MDA}(H_\gamma)$ if and only if there exists some positive scale function a such that

$$\lim_{t\to\infty} \frac{U(ty) - U(t)}{a(t)} = \frac{y^\gamma - 1}{\gamma}, \quad \text{for } y > 0, \tag{2.3}$$

where

$$U(y) \coloneqq \left(\frac{1}{1-F}\right)^{\leftarrow}(y) = \inf\left\{x \in \mathbb{R} : F(x) \geq 1 - \frac{1}{y}\right\} \tag{2.4}$$

is the quantile function for the right tail probability $p = 1/y$ and with $^{\leftarrow}$ denoting the left-continuous generalised inverse. This result is important for statistical inference in EVT and an important contribution from de Haan (1984) and Dekkers and de Haan (1993).

A different approach in EVT is from the perspective of threshold exceedances rather than maxima. That is, by considering the distribution of outcomes which exceed a high level. The corresponding excess distribution over some high threshold u is given by

$$F_u(x) \coloneqq \mathbb{P}\big(X - u \leq x | X > u\big) = \frac{F(x+u) - F(u)}{1 - F(u)}, \tag{2.5}$$

for $0 \leq x < x^* - u$. The Pickands-Balkema-de Haan theorem states, roughly speaking, that the generalised Pareto distribution (GPD) can be used to model the excess distribution for a high u (Balkema and De Haan, 1974; Pickands III, 1975). The GPD has a distribution function given by

$$G_{\gamma,\beta}(x) \coloneqq \begin{cases} 1 - \left(1 + \dfrac{\gamma x}{\beta}\right)^{-1/\gamma} & \text{if } \gamma \neq 0, \\ 1 - \exp\left(-\dfrac{x}{\beta}\right) & \text{if } \gamma = 0, \end{cases} \tag{2.6}$$

where $\beta > 0$ is a scale parameter, and $x \geq 0$ if $\gamma \geq 0$ and $0 \leq x \leq -\beta/\gamma$ if $\gamma < 0$ and γ a is shape parameter equivalent to that of the GEV distribution in (2.2). That is, γ in the GPD corresponds to the same parameter as that in the GEV distributions. For $\gamma > 0$, G_γ, β reduces to an ordinary Pareto distribution, for $\gamma = 0$ to an exponential distribution, and for $\gamma < 0$ a short-tailed Pareto type II distribution.

In accordance with Einmahl et al. (2016), our focus lies on the Fréchet MDA, with $\gamma >$

0, because of its application to financial data. See, among others, the empirical studies by Mandelbrot (1967), Koedijk et al. (1990), Jansen and de Vries (1991), and Kearns and Pagan (1997). Mandelbrot (1967) first noticed that the tails of the distribution of certain asset returns decay like a power law, which can be related to the Fréchet MDA as follows. In correspondence with (2.1) and (2.2) but for different choices of the constants c_n and d_n, one could write the standard Fréchet distribution as $\exp\big(-x^{-\alpha}\big)$, with $\alpha = 1/\gamma > 0$. Then, following Embrechts et al. (2013), by a Taylor expansion,

$$1 - \exp\big(-x^{-\alpha}\big) \sim x^{-\alpha}, \quad \text{as } x \to \infty,$$

which shows that the right tail of the Fréchet distribution decreases as a power law.

This can be formalised using the definition of slowly and regularly varying functions. Following Embrechts et al. (2013), a positive function L on $(0, \infty)$ is said to be slowly varying at ∞ if it holds that

$$\lim_{t\to\infty} \frac{L(tx)}{L(t)} = 1, \quad \text{for } x > 0. \tag{2.7}$$

Similarly, a positive function h on $(0, \infty)$ is said to be regularly varying of index $\alpha \in \mathbb{R}$ at ∞ if it satisfies

$$\lim_{t\to\infty} \frac{h(tx)}{h(t)} = x^{\alpha}, \quad \text{for } x > 0. \tag{2.8}$$

Using these definitions, F is heavy-tailed if and only if

$$1 - F(x) = x^{-1/\gamma} L(x), \tag{2.9}$$

with L slowly varying at ∞. Equivalently, one could characterise the Fréchet MDA by means of excess probability ratios as

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{X}{t} > x \mid X > t\right) = \lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \quad \text{for } x > 0, \tag{2.10}$$

or, in correspondence with (2.3), by

$$\lim_{t\to\infty} \frac{U(ty)}{U(t)} = y^{\gamma}, \quad \text{for } y > 0. \tag{2.11}$$

Thus, distributions in the Fréchet MDA have regularly varying tails with a negative index of variation. This means that the tails decay like a power function with rate $\alpha = 1/\gamma$. The parameter α is therefore often called the tail index of the distribution. An important property of heavy-tailed variables X is that the population moments $\mathbb{E}(X^r) < \infty$ if and only if $\alpha > r$, resulting in important implications for standard statistics which rely on

the existence of population moments of order higher than r. The characteristics of the Fréchet MDA in (2.9), (2.10), (2.11) turn out to be useful for statistical inference, as the well-known Hill estimator for γ explained in Section 2.2 results from these.

2.2 Statistics

We now consider statistical methods for inference on extremal events based on a sample of iid observations $X_1, \ldots, X_n$, with the corresponding order statistics $X_{n,j}$, for $j = 1, \ldots, n$, such that $X_{n,1} \leq \ldots \leq X_{n,n}$. Estimation is achieved from three perspectives, of which the first two result in fully parametric estimation, while the last is of semi-parametric nature. A general assumption is that the observed sample comes from a distribution which belongs to the MDA of one of the three extreme value distributions, which holds for essentially all common continuous distributions in statistics (de Haan and Ferreira, 2006; Embrechts et al., 2013; McNeil et al., 2015).

The first perspective is from the n-block maxima $M_n = \max(X_1, \ldots, X_n)$ of an observed sample. The Fisher-Tippett-Gnedenko theory implies that, for sufficiently large n, one could approximate the true distribution of M_n by the location-scale family of GEV distributions $H_{\gamma,\mu,\sigma}$ as defined in (2.2). Dividing a sample into m disjoint time periods, or blocks, of the n-block maxima M_n, one could use the observed maxima M_n to fit $H_{\gamma,\mu,\sigma}$. The division is often natural: for instance, in blocks of one year in the application in hydrology (Embrechts et al., 2013; McNeil et al., 2015). However, this is not the case in financial applications (Nolde and Zhou, 2021). The GEV distribution can be fitted using maximum likelihood estimation and the method of probability-weighted moments. Typically, n should be large to assure (sufficient) independence of the observed block maxima, even when the observations X_i are not. A disadvantage of the block maxima method is that one only retains the m maxima to fit the GEV distribution, leading to a large loss of potentially useful data. Moreover, one has to decide on the number of blocks m: something that can greatly influence the quality of the estimators.

The second estimation method is called peaks-over-threshold (POT), which is based on the perspective and theoretical results of the threshold exceedances (2.5) and (2.6). This method makes use of the data more efficiently than the block maxima, using all observations that exceed some high threshold (McNeil et al., 2015). Following the Pickand-Balkema-de Haan theorem, we can approximate the excess distribution F_u in (2.5) by the GPD $G_{\gamma,\beta}$ in (2.6) for some sufficiently high threshold u. The model is then estimated by transforming the observations which exceed u to their corresponding excesses and using maximum likelihood or probability-weighted moments. Similar to the block maxima approach, POT requires choosing a proper threshold u.

The last approach to estimate the tail distribution is based on the conditions associated with each MDA and is considered mainly for heavy-tailed distributions and in

Einmahl et al. (2016). Therefore, we continue with a more detailed discussion specific to this case, in which we are mainly interested in estimating extreme value index $\gamma > 0$. In accordance with (2.9), estimating the tail of the distribution then reduces to a semi-parametric approach, as the slowly varying function L is not of interest in the model. This becomes more clear when we move to extreme quantile estimation at the end of this section. Further, we are only interested in the k upper order statistics of the sample for estimation purposes. Then the following assumptions are standard in the EVT literature: for a proper intermediate sequence $k \coloneqq k_n$,

$$\lim_{n\to\infty} k = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{k}{n} = 0. \tag{2.12}$$

This essentially says that we need a sufficiently large number of upper order statistics, but only the ones in the tails.

A well-known and extensively studied estimation method for γ in the Fréchet MDA is the Hill approach, as introduced by Hill (1975). The Hill estimator, for a k as defined in (2.12), is given by

$$\hat{\gamma}_n \coloneqq \frac{1}{k} \sum_{j=1}^{k} \big(\log X_{n,n-j+1} - \log X_{n,n-k} \big). \tag{2.13}$$

This estimator is also part of the key results in Einmahl et al. (2016) and therefore also in our research. Asymptotic normality of the Hill estimator requires the following second-order conditions. First, in addition to (2.11),

$$\lim_{t\to\infty} \frac{U(ty)\,/U(t) - y^{\gamma}}{A(t)} = y^{\gamma} \frac{y^{\rho\gamma} - 1}{\rho\gamma}, \quad \text{for } y > 0, \tag{2.14}$$

for some function A of constant sign and $\lim_{t\to\infty} A(t) = 0$ and second-order parameter $\rho < 0$ governing the rate of convergence. This condition is, however, impossible to check in practice (Embrechts et al., 2013). Secondly, under assumption (2.12), it must hold that

$$\lim_{n\to\infty} \sqrt{k} A\left(\frac{n}{k}\right) = \kappa \in \mathbb{R}. \tag{2.15}$$

Then, following de Haan and Peng (1998), under (2.11), (2.12), (2.14), and (2.15), as $n \to \infty$,

$$\sqrt{k}(\hat{\gamma}_n - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\kappa}{1-\rho\gamma}, \gamma^2 \right). \tag{2.16}$$

Conditions (2.10) and (2.11), together with (2.12), imply a natural estimator for the

quantile function in the tail region (2.4). That is, for tail probability p sufficiently small,

$$\hat{U}\left(\frac{1}{p}\right) \coloneqq X_{n,n-k}\left(\frac{k}{np}\right)^{\hat{\gamma}_n}, \tag{2.17}$$

as introduced by Weissman (1978). Clearly, γ is here the only relevant parameter to estimate. This extreme quantile estimator plays a central role in our study with its extension to the heteroscedastic extremes framework.

In all approaches, the quality of the extreme value statistics is subject to proper selection of a subsample of extreme observations. Generally, it is desirable for an estimator to have minimal bias and variance, which follows, for instance, from the mean squared error (MSE) criterion. That is, for the estimator $\hat{\theta}$ of some parameter θ, $\text{MSE}(\hat{\theta}) \coloneqq \mathbb{E}((\hat{\theta}-\theta)^2) = (\mathbb{E}(\hat{\theta})-\theta)^2 + \mathbb{V}(\hat{\theta})$, where $\mathbb{E}$ and $\mathbb{V}$ denote the mean and variance operators, respectively, decomposing into a (squared) bias and variance term. In extreme value statistics, a relatively larger subsample trivially leads to less estimation variance, but this comes at the cost of potentially using observations which may not qualify as extreme, inducing bias. This represents a typical bias-variance trade-off problem: the subsample should be selected in such a way that the bias and variance are optimally balanced, minimising the MSE (or some other relevant criterion).

In the block maxima approach, the bias-variance trade-off is reflected in terms of m: a relatively larger sample of m n-block maxima leads to lower variance, but may introduce bias by also including observations which cannot be regarded as maxima. Similarly, in the POT method and Fréchet MDA conditions approach, respectively, threshold u and the number of upper order statistics k balance the bias and variance. For example, if k is too large (i.e. u too low), observations which are not in the tail of the distribution are considered, resulting in estimation bias. However, if k is too small (i.e. u too high), the estimation variance tends to be high as there are few observations. This is illustrated by conditions (2.12) and (2.15) and in result (2.16) for the Hill estimator: for $k \to \infty$ sufficiently slowly we get $\kappa = 0$ in (2.15) and therefore an asymptotically unbiased estimator for γ.

As the MSE is unknown in practice, one needs data-driven methods for proper selection of m, u, and k. Focusing on k, a popular pragmatic solution is by means of a so-called Hill plot: selecting k as a point, e.g. the midpoint, in the first stable part in a plot of $\hat{\gamma}_n$ against k, aiming to balance the bias and variance. This can also be done in a plot of extreme quantile estimates against k, as these are often the most relevant quantities in an application. These heuristics are often suggested in the literature, for instance, in Embrechts et al. (2013) and Einmahl et al. (2016), and considered in our empirical application. Further, various (semi-)automatic approaches are proposed, for example, in Danielsson and de Vries (1997) and Danielsson et al. (2001).

3 Heteroscedastic extremes

We now elaborate on the heteroscedastic extremes framework by Einmahl et al. (2016), where the observations are independent but non-identically distributed. This constitutes a non-parametric approach to model differences in the tails across observations, similar to heteroscedasticity in the sense of different variances across observations.

3.1 Setup

Consider the $n \in \mathbb{N}$ independent but non-identically distributed observations X_i, for time points $i = 1, \ldots, n$, which follow the various continuous distribution functions $F_{n,i}$ with common endpoint x^*. Further, let there be a continuous distribution function F with the same endpoint and a continuous positive scaling function $c : [0, 1] \to (0, \infty)$ such that

$$\lim_{x \to x^*} \frac{1 - F_{n,i}(x)}{1 - F(x)} = c\left(\frac{i}{n}\right), \tag{3.1}$$

uniformly for all n and i. Instead of being tail equivalent, the distributions F and $F_{n,i}$ are said to be tail comparable. In case $c = 1$, the set-up reduces to that of classical EVT in the iid case described in Section 2, or homoscedastic extremes.

The function c is called the scedasis function, on which Einmahl et al. (2016) impose that

$$C(1) \coloneqq \int_0^1 c(s)ds = 1 \tag{3.2}$$

and where the function C on $[0, 1]$ is referred to as the integrated scedasis function. Condition (3.2) assures that c is both uniquely defined and can be interpreted as the relative frequency of extremes. Mefleh et al. (2020) show that the scedasis function is asymptotically the probability density function of the relative times in the sample at which extreme events occur.

Additionally, the distribution function F is assumed to belong to the MDA of a GEV distribution, such that (2.3) holds. Together with the tail comparability relation in (3.1), this implies the following MDA condition in the heteroscedastic extremes framework:

$$\lim_{t \to \infty} \frac{U_{n,i}(ty) - U_{n,i}(t)}{a(t)(c(i/n))^\gamma} = \frac{y^\gamma - 1}{\gamma}, \quad \text{for } y > 0. \tag{3.3}$$

Here, $U_{n,i}$ are quantile functions similar to (2.4), but for $F_{n,i}$ instead of F. Condition (3.3) implies that all $F_{n,i}$ belong to the same MDA with the same extreme value index γ, which is a key assumption.

As discussed in Section 2 and in accordance with Einmahl et al. (2016), we focus on

the Fréchet MDA, such that $\gamma > 0$ and $x^* = \infty$. Then, analogously to condition (2.11), (3.3) becomes

$$\lim_{t\to\infty} \frac{U_{n,i}(ty)}{U(t)(c(i/n))^\gamma} = y^\gamma, \quad \text{for } y > 0. \tag{3.4}$$

Combining (2.9) and (3.1) shows that in this case, the scedasis function can also be thought of as a deterministic functional time trend scaling the extreme events:

$$1 - F_{n,i}(x) \sim c\left(\frac{i}{n}\right) L(x) x^{-1/\gamma}, \quad \text{as } x \to \infty. \tag{3.5}$$

3.2 Statistics

We now discuss the statistical methods in the heteroscedastic extremes framework and their asymptotic properties. In general, condition (2.12) on the sequence $k = k_n$ is assumed to hold. Einmahl et al. (2016) suggest to estimate the extreme value index γ by the Hill estimator (2.13) and pose the standard second-order conditions (2.14) and (2.15), with $\kappa = 0$, for the speed of convergence of (2.11). To estimate the integrated scedasis function $C(s) = \int_0^1 c(s)ds$, they propose the sequential empirical survival function

$$\hat{C}_n(s) \coloneqq \frac{1}{k} \sum_{i=1}^{[ns]} \mathbb{1}_{\{X_i > X_{n,n-k}\}}, \tag{3.6}$$

where $\mathbb{1}$ denotes the indicator function, which equals one if its argument is true and zero otherwise. Intuitively, (3.6) is proportional to the number of extreme observations within the first $[ns]$ observations of the sample. To estimate the scedasis function, Einmahl et al. (2016) introduce an adaptation of the standard convolution-type kernel density estimator, which we discuss in Section 4. That is,

$$\hat{c}_{n,h}(s) \coloneqq \frac{1}{kh} \sum_{i=1}^{n} \mathbb{1}_{\{X_i > X_{n,n-k}\}} K\left(\frac{s - i/n}{h}\right), \tag{3.7}$$

where $kh \to \infty$, as $n \to \infty$, and K is a symmetric kernel function as in (4.1), with the additional condition that $K(u) = 0$ for $|u| > 1$. In their simulation and empirical application, Einmahl et al. (2016) consider the biweight kernel function (4.2). They further extend the extreme quantile estimator (2.17) by Weissman (1978) to the heteroscedastic extremes model:

$$\widehat{U}_{n,i}\left(\frac{1}{p}\right) \coloneqq X_{n,n-k} \left(\frac{\hat{c}_{n,h}(i/n)k}{np}\right)^{\hat{\gamma}_n}. \tag{3.8}$$

Then, when extending the support of c to $[0, 1+\epsilon]$ for some $\epsilon > 0$ and including $i = n+1$ to the limit in tail comparability relation (3.1), they propose

$$\widehat{U}_{n,n+1}\left(\frac{1}{p}\right) \coloneqq X_{n,n-k}\left(\frac{\hat{c}_{n,h}(1)k}{np}\right)^{\hat{\gamma}_n} \tag{3.9}$$

as an estimator for the extreme quantile of the unobserved random variable X_{n+1}, that is, an extreme quantile prediction, which plays a key role in our study. Clearly, (3.9) requires proper estimation of the scedasis function at the right boundary point, corresponding to the most recent observation. Therefore,

$$\hat{c}_{n,h}(1) = \frac{1}{kh}\sum_{i=1}^{n} \mathbb{1}_{\{X_i > X_{n,n-k}\}} \tilde{K}\left(\frac{1-i/n}{h}\right), \tag{3.10}$$

where $\tilde{K}$ is the boundary kernel (4.4) to eliminate the bias associated with symmetric convolution-type kernels at the boundaries of densities with compact support. We discuss this more thoroughly in Section 4.

Einmahl et al. (2016) provide additional second-order conditions on the speed of convergence of the tail comparability relation (3.1). That is, for some positive, eventually decreasing function $\tilde{A}$, with $\lim_{t\to\infty} \tilde{A}(t) = 0$,

$$\sup_{n\in\mathbb{N}} \max_{i\in[1,n]} \left| \frac{1-F_{n,i}(x)}{1-F(x)} - c\left(\frac{i}{n}\right) \right| = O\left(\tilde{A}\left(\frac{1}{1-F(x)}\right)\right), \quad \text{as } x \to \infty, \tag{3.11}$$

and

$$\lim_{n\to\infty} \sqrt{k}\tilde{A}\left(\frac{n}{2k}\right) \to 0. \tag{3.12}$$

Further, to assure sufficient smoothness of the scedasis function,

$$\lim_{n\to\infty} \sqrt{k} \sup_{|u-v|\le 1/n} \left|c(u) - c(v)\right| = 0. \tag{3.13}$$

That is, c should be Lipschitz continuous of order at least $\frac{1}{2}$.

Under the assumptions outlined above, Einmahl et al. (2016) prove that, under a Skorokhod construction, as $n \to \infty$,

$$\sup_{0\le s\le 1} \left| \sqrt{k}\big(\hat{C}_n(s) - C(s)\big) - B\big(C(s)\big) \right| \to 0 \quad \text{almost surely} \tag{3.14}$$

and

$$\sqrt{k}(\hat{\gamma}_n - \gamma) \to \gamma Z \quad \text{almost surely,} \tag{3.15}$$

where B is a standard Brownian bridge and Z a standard normal random variable independent of B. These results mean that, asymptotically, the (integrated) scedasis function and extreme value index can be estimated independently. Result (2.16) and (3.15), for $\kappa = 0$, are therefore identical: this classical extreme value statistics result remains, asymptotically, valid. The proofs for these results are based on the sequential tail empirical process (STEP), a normalised sequential empirical distribution function for the tail region introduced by Einmahl et al. (2016).

In extension to these results, Mefleh et al. (2020) show that the times at which extremes occur and the size of these extremes are asymptotically independent. The asymptotic distribution of these exceedance times T_j, defined such that $X_{n,n-j+1} = X_{nT_j}$, has a density equal to the scedasis function, while the values of the exceedances are asymptotically Pareto distributed.

Further, in addition to the assumptions above, supposing $\left|c''(1)\right| < \infty$, $\int_{-1}^{1} \left|K''(u)\right| du < \infty$ and, as $n \to \infty$, the bandwidth h satisfies $kh \to \infty$ and $k^{1/5}h \to \lambda_h \in [0, \infty)$, Einmahl et al. (2016) prove that, as $n \to \infty$,

$$\sqrt{kh}\big(\hat{c}_{n,h}(1) - c(1)\big) \xrightarrow{d} \mathcal{N}\left(\lambda_h^{5/2} \frac{c''(1)}{2} \int_0^1 u^2 \tilde{K}(u)du, c(1) \int_0^1 \tilde{K}^2(u)du\right). \tag{3.16}$$

This is similar to the results for the standard convolution-type kernel density estimators discussed in Section 4. For instance, the mean is similar to expression (4.3), if corrected for boundary effects. As $n \to \infty$, together with result (3.15) and letting the small tail probability $p = p_n$ satisfy $np/k \to 0$ and $\sqrt{h}\log(k/(np)) \to \beta_h \in [0, \infty)$, they further show that

$$\sqrt{kh}\left(\frac{\hat{U}_{n,n+1}(1/p)}{U_{n,n+1}(1/p)} - 1\right) \xrightarrow{d} \mathcal{N}\left(\lambda_h^{5/2} \frac{\gamma c''(1)}{2c(1)} \int_0^1 u^2 \tilde{K}(u)du, \gamma^2 \left(\frac{\int_0^1 \tilde{K}^2(u)du}{c(1)} + \beta_h^2\right)\right). \tag{3.17}$$

Results (3.16) and (3.17) are the convolution-type kernel estimator counterparts to our key results given in Section 5.1.

Einmahl et al. (2016) further suggest test statistics for the heteroscedasticity of the extremes and specific parametric scedasis functions. They also propose statistics to test whether γ is fixed over the observations, which is a key assumption for the validity of the model. Last, Mefleh et al. (2020) consider parametric models for the scedasis function in terms of exceedance times T_j to test for a trend with more power.

4 Kernel density estimation

Kernel density estimation is a non-parametric approach to estimating the probability density function f underlying a sample of observations $X_1, \ldots, X_n$. The classical convolution-type kernel density estimator for f at a point x in its support is

$$\hat{f}_{n,h}(x) \coloneqq \frac{1}{nh} \sum_{i=1}^{n} K\bigg(\frac{x - X_i}{h}\bigg), \tag{4.1}$$

for some smoothing bandwidth $h > 0$ such that $h \to 0$ and $nh \to \infty$, as $n \to \infty$, and K a kernel function. The estimator is such that the kernel gives the relative contribution of each observation X_i to the density estimate at x. Typically, K is a symmetric and unimodal density function: $K \geq 0$, $\int K(u)du = 1$, and $K(-u) = K(u)$. For instance, Einmahl et al. (2016) consider the biweight, or quartic kernel function

$$K(u) = \frac{15}{16}(1 - u^2)^2, \quad \text{for } u \in [-1, 1], \tag{4.2}$$

and $K(u) = 0$ otherwise, in their scedasis function estimator (3.7). Other examples are the standard Gaussian, uniform, triangular, and Epanechnikov kernel functions. The bandwidth h determines the degree of smoothness of the estimator and thus plays a central role in its bias-variance trade-off: a concept introduced in Section 2.2 in relation to extreme value statistics. Relatively larger values of h result in a smoother density curve estimate, and thus in less variance, but higher bias, as observations X_i further away from x contribute relatively more to $\hat{f}_{n,h}(x)$. We more formally discuss this trade-off by means of the statistical properties of the beta kernel estimator in (4.8) and (4.9).

When the density to be estimated has compact support, the symmetric convolution-type kernel estimator is generally inconsistent in the regions near the boundaries, as it allocates weight outside the theoretical range of the data. We call this 'boundary bias', in correspondence with Chen (1999). Along the lines of the discussion in Jones (1993), but for a density f with support $[0, 1]$, it follows that

$$\begin{aligned} \mathbb{E}\Big(\hat{f}_{n,h}(x)\Big) = f(x) \int_{(x-1)/h}^{x/h} K(u)du \; - hf'(x) \int_{(x-1)/h}^{x/h} uK(u)du \\ + h^2 \frac{f''(x)}{2} \int_{(x-1)/h}^{x/h} u^2 K(u)du + o\big(h^2\big)\,. \end{aligned} \tag{4.3}$$

Now, for simplicity and in accordance with Jones (1993) and Einmahl et al. (2016), consider a symmetric kernel with support $[-1, 1]$, e.g. (4.2). Then, because $\int_{(x-1)/h}^{x/h} K(u)du = 1$ only if the limits of integration reduce to the limits of support of the kernel, the estimator is inconsistent in the boundary regions $x \in [0, h] \cup [1 - h, 1]$. For instance, specific

to our discussion for estimation at the right boundary, as $n \to 0$, we have

$$\begin{aligned}\mathbb{E}\Big(\hat{f}_{n,h}(1)\Big) &\sim f(1)\int_0^{1/h} K(u)du - hf'(1)\int_0^{1/h} uK(u)du \\ &\sim \frac{1}{2}f(1) - hf'(1)\int_0^1 uK(u)du,\end{aligned}$$

implying a bias of $-f(1)/2 + O(h)$. In contrast, the bias is $O(h^2)$ in the interior region, $x \in (h, 1-h)$, leaving the third term of (4.3) as leading, which results in a consistent estimator.

One of the solutions to this problem is the linear correction generalised jackknife boundary kernel proposed by Jones (1993). This boundary kernel is also considered by Einmahl et al. (2016) for the boundary scedasis estimator (3.10) and by Chen (1999) for comparison to their beta kernel estimator and is therefore central in our study. Specifically, for a density with compact support $[0, 1]$, the boundary kernel for the upper boundary region $x \in [1-h, 1]$ is given by

$$\tilde{K}\left(\frac{x - X_i}{h}\right) \coloneqq \frac{\int_{(x-1)/h}^1 u^2K(u)du - ((x - X_i)/h)\int_{(x-1)/h}^1 uK(u)du}{\int_{(x-1)/h}^1 K(u)du \int_{(x-1)/h}^1 u^2K(u)du - \left(\int_{(x-1)/h}^1 uK(u)du\right)^2} K\left(\frac{x - X_i}{h}\right), \tag{4.4}$$

where K is a symmetric kernel as in (4.1) with support $[-1, 1]$. A similar kernel is defined for $x \in [0, h]$, which can be found explicitly in Jones (1993). This corrected kernel estimator ensures consistency in the boundary regions, though it has the disadvantage that it can take on negative values and does not result in unit integral density estimators (Jones, 1993). Jones and Foster (1996) give a solution to the former problem.

Opposed to a convolution-type kernel estimator, Chen (1999) propose to use a beta density as a kernel function for estimating densities with compact support. That is, let

$$K_{\alpha,\beta}(u) \coloneqq \frac{1}{\mathcal{B}(\alpha, \beta)} u^{\alpha-1}(1-u)^{\beta-1}, \quad \text{for } u \in [0, 1], \tag{4.5}$$

denote the beta density function of a Beta(α, β) variable with shape parameters $\alpha, \beta > 0$, where

$$\mathcal{B}(\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \tag{4.6}$$

denotes the beta function with Γ the gamma function. Then, for a density with support

$[0, 1]$ and a point x in this support, the beta kernel density estimator is

$$\hat{f}_{n,b}(x) \coloneqq \frac{1}{n} \sum_{i=1}^{n} K_{x/b+1,(1-x)/b+1}(X_i), \tag{4.7}$$

where b is a smoothing bandwidth parameter such that $b \to 0$ and $nb \to \infty$ as $n \to \infty$. In symmetric convolution-type kernel estimators, the bandwidth h and value x can be seen as the scale and location parameters of the kernel density function, respectively. In contrast, in the beta kernel estimator, b and x control the shape parameters of the beta density function, such that x is its mode. For a fixed b, the beta kernel then varies in shape over x, resulting in an adaptive amount of smoothing. Therefore, it is an asymmetric and adaptive kernel density estimator. Moreover, because the kernel and the density underlying the data share the same compact support, the estimator allocates no weight outside the range of the data. Therefore, the beta kernel density estimator has no boundary bias problem and has a larger effective sample size. Chen (1999) find that the estimator can have smaller finite-sample variance than the boundary kernel estimator (4.4) and its successor by Jones and Foster (1996).

Chen (1999) show, as $n \to \infty$, assuming $x \in [0, 1]$, $\left|f''(x)\right| < 0$, and $b \to 0$, that

$$\mathbb{E}\Big(\hat{f}_{n,b}(x)\Big) = f(x) + b\bigg((1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x)\bigg) + o(b) \tag{4.8}$$

and

$$\mathbb{V}\Big(\hat{f}_{n,b}(x)\Big) = \begin{cases} \dfrac{1}{2n\sqrt{\pi b x(1-x)}}\big(f(x) + o(1)\big) & \text{if } x \in (0,1), \\ \dfrac{1}{2nb}\big(f(x) + o(1)\big) & \text{if } x = 0 \text{ or } x = 1. \end{cases} \tag{4.9}$$

In contrast to (4.3), (4.8) clearly shows that the bias of the estimator is $O(b)$, irrespective of proximity to a boundary. However, $b \to 0$ should be at a rate slow enough to have $n\sqrt{b} \to \infty$ and $nb \to \infty$, respectively, for the variances of the interior and the boundaries in (4.9) to shrink to zero. This presents the typical trade-off between bias and variance governed by the bandwidth parameter, which is similar for convolution-type kernel estimators.

We must note that Chen (1999) additionally propose a slightly different estimator than presented in (4.7), which eliminates the unsatisfactory dependence on $f'(x)$ in the bias for the interior region. However, as we are mainly interested in estimation at the right boundary point of the scedasis function, this estimator is irrelevant in our research as it merely complicates the expressions and derivations to follow.

Due to the bias-variance trade-off, the choice of the bandwidth is generally considered as more important for the performance of kernel estimators than the choice of the kernel

(Chen, 1999). Globally, over their full support, density estimators are typically evaluated in terms of the mean integrated squared error (MISE). That is, for some estimator $\hat{f}_n$ for density f, $\text{MISE}(\hat{f}_n) = \int \text{MSE}(\hat{f}_n(x))dx$, with the MSE as discussed in Section 2.2. The global optimal bandwidth is then the minimiser of this criterion, balancing the bias and variance of the estimator over its full support. Both the beta and convolution-type kernel estimators asymptotically achieve an optimal rate of MISE convergence of $O(1/n^{4/5})$ for their global optimal bandwidths $h^* = O(1/n^{1/5})$ and $b^* = O(1/n^{2/5})$, respectively. We discuss the local optimal bandwidth and convergence, i.e. in terms of the MSE at a specific point x, for the kernel estimators specific to scedasis estimation at the right boundary, in Section 5.2.

Among the many different data-based techniques to select a proper bandwidth, with respect to the (approximate) MISE, are rule-of-thumb, least-squares and biased cross-validation, solve-the-equation plug-in, and smoothed bootstrap. For a review, see, among others, Jones et al. (1996) and, more recently, Heidenreich et al. (2013) and Zambom and Ronaldo (2013). These are generally specific to convolution-type kernel estimators, as literature for beta kernel estimators is scarce.

5 The beta kernel scedasis estimator

In this section we discuss the beta kernel estimator for the scedasis function defined in (3.1). We provide the asymptotic distribution of the estimator at the right boundary point and of the corresponding extreme quantile prediction, together with a brief analysis of the local optimal MSE and bandwidth, in Section 5.1. Some practical considerations with respect to selecting the bandwidth and the number of upper order statistics are discussed in Section 5.2 and the proofs are in Section 5.3. Generally, the notation is adopted from the previous sections.

In correspondence with the scedasis estimator (3.7) by Einmahl et al. (2016) and the beta kernel density estimator (4.7) by Chen (1999), our beta kernel scedasis estimator is

$$\hat{c}_{n,b}(s) \coloneqq \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\{X_i > X_{n,n-k}\}} K_{s/b+1,(1-s)/b+1}\left(\frac{i}{n}\right), \tag{5.1}$$

for which we assume $b \to 0$ and $kb \to \infty$ as $n \to \infty$. We continue to distinguish this estimator from (3.7) and (3.10) by the subscript b instead of h, which correspond to their specific bandwidth parameters. Because of our focus on extreme quantile prediction, we

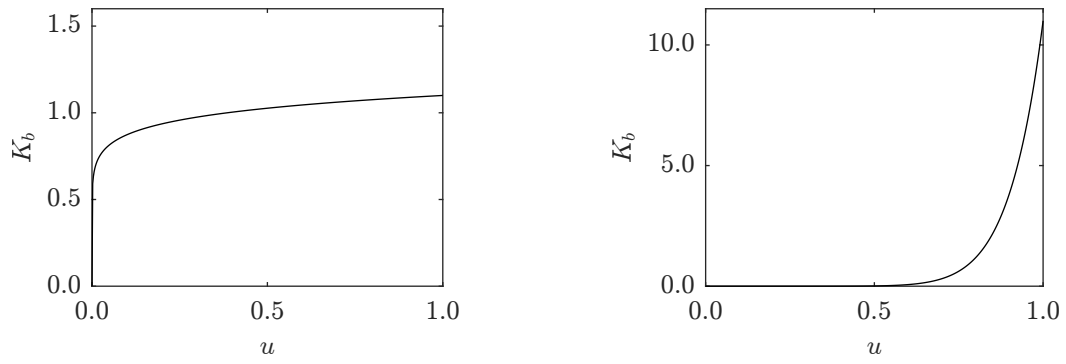


Figure 5.1: Beta kernel $K_b = K_{1/b+1,1}$ at the right boundary for bandwidths $b = 10$ (left) and $b = 0.1$ (right).

are mainly interested in estimating $c(1)$. The estimator then reduces to

$$\begin{aligned} \hat{c}_{n,b}(1) &= \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\{X_i > X_{n,n-k}\}} K_{1/b+1,1}\left(\frac{i}{n}\right) \\ &= \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\{X_i > X_{n,n-k}\}} \left(\frac{1}{b}+1\right)\left(\frac{i}{n}\right)^{1/b}. \end{aligned} \tag{5.2}$$

For notational convenience, we denote $K_b \coloneqq K_{1/b+1,1}$ in the remainder of this section. Note that in case $b \to \infty$, K_b reduces to a standard uniform kernel, and in case $b \to 0$, to a degenerate distribution concentrated at its right boundary, which is demonstrated in Figure 5.1. Further, the extreme quantile estimator (3.8) and prediction (3.9) extend in a trivial way to the beta scedasis estimator by replacing $\hat{c}_{n,h}$ by $\hat{c}_{n,b}$, In particular,

$$\widehat{U}_{n,n+1}\left(\frac{1}{p}\right) \coloneqq X_{n,n-k}\left(\frac{\hat{c}_{n,b}(1)\,k}{np}\right)^{\hat{\gamma}_n}. \tag{5.3}$$

5.1 Asymptotic theory

We first provide a theorem on the asymptotic distribution of $\hat{c}_{n,b}(1)$. It is in close correspondence with result (3.16) for the boundary kernel estimator by Einmahl et al. (2016) and with results (4.8) and (4.9) by Chen (1999). However, in contrast to Einmahl et al. (2016), the beta kernel scedasis estimator is non-negative and matches the compact support of the scedasis function and we explicitly prove normality via the Lindeberg-Feller theorem. Moreover, opposed to Chen (1999), we focus solely on the kernel estimator at the right boundary point and establish its asymptotic distribution.

Theorem 1. *Let the scedasis function be as defined in* (3.1) *and satisfy* (3.2) *and* $\left|c'(1)\right| < \infty$. *Further, let* $\hat{c}_{n,b}(1)$ *be the beta kernel scedasis estimator* (5.2) *for* $c(1)$. *Assume that the conditions in* (2.12), (3.11), (3.12), *and* (3.13) *hold and that* $b \to 0$, $kb \to \infty$, *and*

$k^{1/3}b \to \lambda_b \in [0, \infty)$ *as* $n \to \infty$. *Then, as* $n \to \infty$,

$$\sqrt{kb}\big(\hat{c}_{n,b}(1) - c(1)\big) \xrightarrow{d} \mathcal{N}\left(-\lambda_b^{3/2} c'(1), \frac{1}{2}c(1)\right).$$

Proof. See Section 5.3 □

We further present a theorem on the asymptotic distribution of the extreme quantile prediction based on the beta kernel scedasis estimator. This is is similar to result (3.17) by Einmahl et al. (2016), but explicitly in terms of our beta kernel scedasis estimator and under the additional assumption of asymptotic independence of $\hat{c}_{n,b}(1)$ and $\hat{\gamma}_n$.

Theorem 2. *Adopt the assumptions and definitions in Theorem 1. Additionally, let the scedasis function be defined on* $[0, 1+\epsilon]$ *for some* $\epsilon > 0$ *and assume* (3.11) *holds with* $i = n+1$ *included. Further, let* $\hat{\gamma}_n$ *be the estimator* (2.13) *for* γ, *which we assume is asymptotically independent of* $\hat{c}_{n,b}(1)$, *and assume that conditions* (2.14) *and* (2.15) *for* $\kappa = 0$ *hold. Let* $\hat{U}_{n,n+1}(1/p)$ *be the extreme quantile prediction* (5.3). *Then, as* $n \to \infty$, *for* $p = p_n$ *such that* $np/k \to 0$ *and* $\sqrt{b}\log(k/(np)) \to \beta_b \in [0, \infty)$,

$$\sqrt{kb}\left(\frac{\hat{U}_{n,n+1}(1/p)}{U_{n,n+1}(1/p)} - 1\right) \xrightarrow{d} \mathcal{N}\left(-\lambda_b^{3/2}\frac{\gamma c'(1)}{c(1)}, \gamma^2\left(\frac{1}{2c(1)} + \beta_b^2\right)\right).$$

Proof. See Section 5.3 □

The additional assumptions (2.14) and (2.15) are necessary for result (3.15) on the normality of the Hill estimator. Further, we need the asymptotic independence of $\hat{\gamma}_n$ and $\hat{c}_{n,b}(1)$ to establish their joint normality. This is something we do not formally prove and cannot directly extend from the proof by Einmahl et al. (2016). In contrast to their boundary kernel, the beta kernel always has support over the full range of the data, independent of the choice of its bandwidth parameter. However, as $b \to 0$, $\hat{c}_{n,b}(1)$ increasingly concentrates on observations close to the right endpoint, as shown in Figure 5.1. As $\hat{\gamma}_n$ is based on the full sample, we argue intuitively that their dependence becomes asymptotically negligible. Further, Mefleh et al. (2020) prove asymptotic independence of the exceedance times and corresponding exceedance values, which are, respectively, associated with the estimation of c and γ. A more elaborate discussion on the independence assumption can be found in Appendix A.2.

The biases of the estimators are $O(b)$, implying that there is no boundary bias problem. However, as they include $-c'(1)$, there is a tendency to underestimate increasing scedasis at the boundary in finite-sample applications. This is potentially worrisome in the application to risk management, as it would lead to extreme quantile predictions which are too low in times of increasingly more extremes. Zhang and Karunamuni (2010) even suggest that if the so-called shoulder condition $c'(1) = 0$ is not satisfied, the beta

kernel estimator has a severe boundary problem. For comparison, the convolution-type boundary kernel estimator (3.10) tends to underestimate peaks as its bias depends on $c''(1)$. The variances are similar to the results by Einmahl et al. (2016): as $b \to 0$, setting $h = 1$ and replacing the boundary kernel by K_b gives the same results.

There is a typical bias-variance trade-off governed implicitly in λ_b by the bandwidth parameter, as discussed in Section 4. Therefore, we briefly discuss the asymptotically optimal properties in terms of the MSE. Theorem 1 implies, as $n \to \infty$,

$$\begin{aligned} \mathrm{MSE}\big(\hat{c}_{n,b}(1)\,\big) &\coloneqq \Big[\mathbb{E}\big(\hat{c}_{n,b}(1)\big) - c(1)\Big]^2 + \mathbb{V}\big(\hat{c}_{n,b}(1)\big) \\ &\sim b^2\big(c'(1)\big)^2 + \frac{1}{2kb}c(1). \end{aligned} \tag{5.4}$$

Following straightforward minimisation of (5.4) with respect to b, the local optimal bandwidth choice, i.e. at the boundary point, is

$$b^* \coloneqq \left(\frac{c(1)}{4k(c'(1))^2}\right)^{1/3} = O\left(\frac{1}{k^{1/3}}\right). \tag{5.5}$$

Combining (5.4) and (5.5) gives, as $n \to \infty$,

$$\mathrm{MSE}\big(\hat{c}_{n,b^*}(1)\big) \sim \frac{1}{k^{2/3}}\left[\big(c(1)c'(1)\big)^{2/3}\left(\frac{1}{4^{2/3}} + \frac{4^{3/2}}{2}\right)\right] = O\left(\frac{1}{k^{2/3}}\right). \tag{5.6}$$

Similarly, following (3.16), it can be shown that the local optimal bandwidth of the convolution-type boundary kernel estimator is $h^* = O(1/k^{1/5})$, with an optimal MSE of $O(1/k^{4/5})$ as $n \to \infty$. Therefore, in terms of k, the convergence of the optimal MSE of the beta kernel scedasis estimator at the boundary is slower than that of the convolution-type estimator and that of its own global MISE counterpart as discussed in Section 4. However, Chen (1999) highlight the potential finite-sample advantages of a beta kernel estimator, which we examine in our simulation study in Section 6.

5.2 Practical considerations

We briefly discuss some practical considerations for selection of the bandwidth and the number of upper order statistics in kernel scedasis estimation in general.

Most of the existing literature on kernel estimation discuss data-driven methods for global bandwidth selection. In the same light, we derive a simple cross-validation criterion for global bandwidth selection for the convolution-type kernel scedasis estimator (3.7), which can be found in Appendix A.1. This traditional criterion has no explicit extension to a beta kernel and would require numerical integration. In agreement with existing literature, for instance, Jones et al. (1996), we find that the criterion results

in bandwidths which are too small. That is, the resulting scedasis estimates have too much variation, both objectively in preliminary simulation results and subjectively in application to our financial loss data. Standard rule-of-thumb bandwidths are neither appropriate, as the scedasis functions may deviate substantially from a Gaussian density. More generally, as becomes apparent in our empirical application, the scedasis function may often substantially differ from standard density functions. Therefore, we believe that familiar global bandwidth selection methods are often not satisfactory for scedasis estimation. However, local bandwidth selection requires additional complicated bandwidth choices (Jones et al., 1996), for instance, for the second-order derivative estimator in the solve-the-equation plug-in method by Thombs and Sheather (1992).

To narrow the scope of our study, we leave the subject of data-driven bandwidth selection methods for kernel scedasis estimation to future research. We consider pragmatic solutions in our simulation and application, for instance, by taking only the optimal order of the bandwidths into consideration. That is, as discussed in Section 5.1, $b^* = O(1/k^{1/3}) = O((h^*)^{5/3})$ at the boundary and $b^* = O(1/k^{2/5}) = O((h^*)^2)$ globally.

Further, the optimal bandwidth of the kernel scedasis estimator depends on k. Vice versa, the extreme quantile prediction, which depends on the bandwidth choice via the scedasis estimator, is an important consideration in selecting k. Therefore, the selection of the bandwidths and k are interconnected. However, results (3.14) and (3.15) imply that estimating the extreme value index and the scedasis can be done independently of each other. Therefore, we consider a sequential approach: first determining k, as discussed in Section 2.2, by a Hill plot rather than a quantile plot, and subsequently selecting the bandwidth.

5.3 Proofs

Proof of Theorem 1

Let $t_n \coloneqq (n/k)(1 - F(X_{n,n-k}))$ and Z denote a standard normal random variable. Then Einmahl et al. (2016) prove that, under a Skorokhod construction, as $n \to \infty$,

$$\left|\sqrt{k}(t_n - 1) + Z\right| \to 0 \quad \text{almost surely.}$$

Almost sure convergence implies convergence in probability, such that, as $n \to \infty$,

$$\mathbb{P}\left(\left|\sqrt{k}(t_n - 1) + Z\right| > \varepsilon\right) \to 0, \quad \text{for all } \varepsilon > 0,$$

or, equivalently, $\sqrt{k}(t_n - 1) + Z = o_p(1)$. Because Z is a random variable independent of n, it follows that it is by definition bounded in probability of order 1: $Z = O_p(1)$. Thus, $\sqrt{k}(t_n - 1) + Z = \sqrt{k}(t_n - 1) + O_p(1) = o_p(1)$. Then, following van der Vaart (2000),

using the properties that $O_p(1)+o_p(1)=O_p(1)$ and that o_p implies O_p, it must hold that $\sqrt{k}(t_n-1)=O_p(1)$. Now define the intermediate sequence $\delta \coloneqq \delta_n$ such that $\delta \to 0$ as $n \to \infty$. Then, $\delta\sqrt{k}(t_n-1)=O_p(\delta)=o_p(1)$, as $n \to \infty$:

$$\mathbb{P}\Big(\delta\sqrt{k}|t_n-1|>\varepsilon\Big) \to 0, \quad \text{for all } \varepsilon>0$$

Equivalently,

$$\mathbb{P}\Big(\delta\sqrt{k}|t_n-1|\leq 1\Big) = \mathbb{P}\big(t^- \leq t_n \leq t^+\big) \to 1, \tag{5.7}$$

where we let $t^{\pm} \coloneqq 1 \pm 1/(\delta\sqrt{k})$. Einmahl et al. (2016) use $\delta = h^{1/4}$, in terms of the bandwidth of boundary kernel scedasis estimator (3.10), but (5.7) would also hold for $\delta=\sqrt{b}$, as long as $\delta \to 0$ as $n \to \infty$. Now, because

$$U\left(\frac{n}{kt_n}\right) = U\left(\frac{1}{1-F\big(X_{n,n-k}\big)}\right) = X_{n,n-k},$$

we define, in correspondence with (5.2),

$$\tilde{c}_{n,b}(u) \coloneqq \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\left\{X_i > U\left(n/(ku)\right)\right\}} K_b\left(\frac{i}{n}\right), \tag{5.8}$$

such that $\hat{c}_{n,b}(1)=\tilde{c}_{n,b}(t_n)$ almost surely and, in case (5.7) happens,

$$\tilde{c}_{n,b}\big(t^-\big) \leq \tilde{c}_{n,b}(t_n) \leq \tilde{c}_{n,b}\big(t^+\big)\,. \tag{5.9}$$

Note that, in contrast to the proof in Einmahl et al. (2016), who make a distinction between the negative and positive parts of the boundary kernel, the beta kernel is always non-negative. Thus, we have to prove that, as $n \to \infty$,

$$\sqrt{kb}\big(\tilde{c}_{n,b}\big(t^{\pm}\big)-c(1)\big) \xrightarrow{d} \mathcal{N}\left(-\lambda^{3/2}c'(1), \frac{1}{2}c(1)\right), \tag{5.10}$$

in order to prove the main theorem.

We first need the following proposition.

Proposition 1. *Under the assumptions stated in Theorem 1, as $n \to \infty$,*

$$\mathbb{E}\Big(\tilde{c}_{n,b}\big(t^{\pm}\big)\Big) = c(1) - bc'(1) + o\left(\frac{1}{\sqrt{kb}}\right)$$

and

$$\mathbb{V}\Big(\tilde{c}_{n,b}\big(t^{\pm}\big)\Big) = \frac{1}{2kb}c(1) + o\bigg(\frac{1}{kb}\bigg).$$

Proof. To prove Proposition 1, we need two lemmas, which are slight adaptations of proofs by Chen (2000).

Lemma 1. *Let* $T_1 = T_{1/b+1,1}$ *and* $T_2 = T_{2/b+1,1}$ *be Beta*$(1/b+1,1)$ *and Beta*$(2/b+1,1)$ *random variables, respectively. Then, as* $b \to 0$*, their means and variances are, respectively,*

$$\mathbb{E}(T_1) = \mathbb{E}(T_2) = 1 - b + O(b^2) \quad \textit{and} \quad \mathbb{V}(T_1) = \mathbb{V}(T_2) = O(b^2).$$

Proof. The mean and variance of beta random variable $T \sim \text{Beta}(\alpha, \beta)$ are

$$\mathbb{E}(T) = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \mathbb{V}(T) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)},$$

respectively. Therefore,

$$\mu_1(b) = \mathbb{E}(T_1) = \frac{1+b}{1+2b}, \quad \mu_2(b) = \mathbb{E}(T_2) = \frac{2+b}{2+2b},$$

$$\sigma_1^2(b) = \mathbb{V}(T_1) = \frac{1/b+1}{(1/b+2)^2(1/b+3)}, \quad \text{and} \quad \sigma_2^2(b) = \mathbb{V}(T_2) = \frac{2/b+1}{(2/b+2)^2(2/b+3)}.$$

Then by considering a second-order Taylor expansion around 0, as $b \to 0$ by assumption, we can write

$$\mu_1(b) = \mu_2(b) = 1 - b + O(b^2) \quad \text{and} \quad \sigma_1^2(b) = \sigma_2^2(b) = O(b^2),$$

which proves the lemma.

□

Lemma 2. *Let* $T_1 = T_{1/b+1,1}$ *and* $T_2 = T_{2/b+1,1}$ *be Beta*$(1/b+1,1)$ *and Beta*$(2/b+1,1)$ *random variables, respectively, and let* c *be a continuous function with* $\big|c''(1)\big| < \infty$*. Then, as* $b \to 0$*,*

$$\mathbb{E}\big(c(T_1)\big) = \mathbb{E}\big(c(T_2)\big) = c(1) - bc'(1) + o(b).$$

Proof. Using a second order Taylor expansion around 1, following the proof by Chen

(2000), we can write

$$c(T_1) = c(1) + c'(1)(T_1 - 1) + \frac{1}{2}c''(1)(T_1 - 1)^2 + r(T_1 - 1),$$

where r denotes the remainder of the expansion. Then by taking the expectation on both sides and using Lemma 1,

$$\begin{aligned}\mathbb{E}\big(c(T_1)\big) &= c(1) + c'(1)\mathbb{E}(T_1 - 1) + \frac{1}{2}c''(1)\mathbb{E}\big((T_1 - 1)^2\big) + \mathbb{E}\big(r(T_1 - 1)\big) \\ &= c(1) - bc'(1) + \frac{1}{2}c''(1)\mathbb{V}(T_1 - b) + \mathbb{E}\big(r(T_1 - 1)\big) \\ &= c(1) - bc'(1) + o(b).\end{aligned}$$

For a proof for the convergence of the remainder term $\mathbb{E}\big(r(T_1 - 1)\big)$ we refer to Chen (2000). Because, following Lemma 1, in the limit the first two moments of both T_1 and T_2 are equivalent, it follows that $\mathbb{E}\big(c(T_1)\big) = \mathbb{E}\big(c(T_2)\big)$, which concludes the proof of this lemma.

□

We can now continue with the proof of Proposition 1. We start with the expectation. Note that $\tilde{c}_{n,b}\big(t^{\pm}\big)$ is a sum of random variables which are non-identically distributed. This fact, combined with the second-order condition (3.11), gives, as $n \to \infty$,

$$\begin{aligned}\mathbb{E}\Big(\tilde{c}_{n,b}\big(t^{\pm}\big)\Big) &= \frac{1}{k}\sum_{i=1}^{n} K_b\left(\frac{i}{n}\right)\mathbb{E}\left(\mathbb{1}_{\left\{X_i > U\left(n/(kt^{\pm})\right)\right\}}\right) \\ &= t^{\pm}\frac{1}{n}\sum_{i=1}^{n} K_b\left(\frac{i}{n}\right)\left[c\left(\frac{i}{n}\right) + O\left(\tilde{A}\left(\frac{n}{kt^{\pm}}\right)\right)\right],\end{aligned} \tag{5.11}$$

because condition (3.11) implies that for all $n \in \mathbb{N}$ and all $i = 1, \ldots, n$, as $n \to \infty$,

$$\begin{aligned}\mathbb{E}\left(\mathbb{1}_{\left\{X_i > U\left(n/(kt^{\pm})\right)\right\}}\right) &= \mathbb{P}\left(X_i > U\left(\frac{n}{kt^{\pm}}\right)\right) \\ &= 1 - F_{n,i}\left(U\left(\frac{n}{kt^{\pm}}\right)\right) \\ &= \frac{kt^{\pm}}{n}\left[c\left(\frac{i}{n}\right) + O\left(\tilde{A}\left(\frac{n}{kt^{\pm}}\right)\right)\right].\end{aligned} \tag{5.12}$$

Using the definition and the corresponding error of the right Riemann sum approximation

to an integral together with Lemma 2, we get, as $n \to \infty$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} c\left(\frac{i}{n}\right) K_b\left(\frac{i}{n}\right) &= \int_0^1 c(u)K_b(u)du + O\left(\frac{1}{n}\right) \\ &= \mathbb{E}\Big(c\big(T_{1/b+1,1}\big)\Big) + O\left(\frac{1}{n}\right) \\ &= c(1) - bc'(1) + o(b) + O\left(\frac{1}{n}\right),\end{aligned} \tag{5.13}$$

where $T_{1/b+1,1}$ is a Beta$(1/b+1, 1)$ variable. Similarly, as $n \to \infty$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} K_b\left(\frac{i}{n}\right) &= \int_0^1 K_b(u)du + O\left(\frac{1}{n}\right) \\ &= 1 + O\left(\frac{1}{n}\right).\end{aligned} \tag{5.14}$$

Combining (5.11), (5.13), and (5.14), then gives, as $n \to \infty$,

$$\begin{aligned}\mathbb{E}\Big(\hat{c}_{n,b}\big(t^{\pm}\big)\Big) = t^{\pm}\Bigg[&c(1) - b'c(1) + o(b) + O\left(\frac{1}{n}\right) \\ &+ \; O\left(\tilde{A}\left(\frac{n}{kt^{\pm}}\right)\right)\left(1 + O\left(\frac{1}{n}\right)\right)\Bigg].\end{aligned} \tag{5.15}$$

Now note that, as $n \to \infty$, $kb \to \infty$ and $n/k \to \infty$ together imply $o(b) + O(1/n) = o(b)$. Further, as we assume that $k^{1/3}b \to \lambda_b \in [0, \infty) = O(1)$, we have $\sqrt{kb}o(b) = o(\lambda_b^{3/2}) = o(1)$, or, equivalently, $o(b) = o(1/\sqrt{kb})$. Condition (3.12) implies $\tilde{A}(n/(kt^{\pm})) = o(1/\sqrt{kb})$, because, as $n \to \infty$,

$$\sqrt{kb}\tilde{A}\left(\frac{n}{kt^{\pm}}\right) \leq \sqrt{kb}\tilde{A}\left(\frac{n}{2k}\right) \leq \sqrt{k}\tilde{A}\left(\frac{n}{2k}\right) \to 0.$$

Lastly, $t^{\pm} = 1 \pm 1/(\delta\sqrt{k}) = 1 \pm o(1/\sqrt{kb})$ if we choose δ such that $\sqrt{b}/\delta \to 0$. Therefore, (5.15) reduces to the expression in the proposition, which completes this part of the proof.

We now continue with the variance. Similar to (5.11), using (5.12) together with $\mathbb{V}(\mathbb{1}_A) = \mathbb{P}(A)\,(1 - \mathbb{P}(A))$ and the independence of the random variables, we get, as

$n \to \infty$,

$$\begin{aligned}
\mathbb{V}\Big(\tilde{c}_{n,b}\big(t^{\pm}\big)\Big) &= \frac{1}{k^2}\sum_{i=1}^{n} K_b^2\left(\frac{i}{n}\right)\mathbb{V}\left(\mathbb{1}_{\left\{X_i > U\left(n/(kt^{\pm})\right)\right\}}\right) \\
&= \frac{t^{\pm}}{k}\frac{1}{n}\sum_{i=1}^{n} K_b^2\left(\frac{i}{n}\right)\left[c\left(\frac{i}{n}\right) + O\left(\tilde{A}\left(\frac{n}{kt^{\pm}}\right)\right)\right] \\
&\quad - \frac{(t^{\pm})^2}{n}\frac{1}{n}\sum_{i=1}^{n} K_b^2\left(\frac{i}{n}\right)\left[c\left(\frac{i}{n}\right) + O\left(\tilde{A}\left(\frac{n}{kt^{\pm}}\right)\right)\right]^2 .
\end{aligned} \tag{5.16}$$

Again, using the definition and the corresponding error of the right Riemann sum approximation to an integral, together with Lemma 2 and the properties of the beta density function in (4.5), we get, as $n \to \infty$,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} c\left(\frac{i}{n}\right) K_b^2\left(\frac{i}{n}\right) &= \int_0^1 c(u)K_b^2(u)du + O\left(\frac{1}{n}\right) \\
&= \frac{\mathcal{B}(2/b+1,1)}{\mathcal{B}^2(1/b+1,1)}\mathbb{E}\Big(c\big(T_{2/b+1,1}\big)\Big) + O\left(\frac{1}{n}\right) \\
&= \frac{1}{b}\left(\frac{1}{2} + O\big(b^2\big)\right)\Big(c(1) - bc'(1) + o(b)\Big) + O\left(\frac{1}{n}\right) \\
&= \frac{1}{2b}c(1) + O(1)\,,
\end{aligned} \tag{5.17}$$

where $\mathcal{B}$ denotes the beta function as defined in (4.6) and where $T_{2/b+1,1}$ is a Beta$(2/b+1,1)$ variable. This follows from the fact that, for a Beta(α,β) variable with density function $K_{\alpha,\beta}$ as defined in (4.5),

$$\begin{aligned}
K_{\alpha,\beta}^2(x) &= \frac{1}{\mathcal{B}^2(\alpha,\beta)}x^{(2\alpha-1)-1}(1-x)^{(2\beta-1)-1} \\
&= \frac{\mathcal{B}(2\alpha-1,2\beta-1)}{\mathcal{B}^2(\alpha,\beta)}K_{2\alpha-1,2\beta-1}(x),
\end{aligned}$$

and, by the definition of the beta function in (4.6) and a Taylor expansion around $b=0$,

$$\frac{\mathcal{B}(2/b+1,1)}{\mathcal{B}^2(1/b+1,1)} = \frac{b/(b+2)}{b^2/(b+1)^2} = \frac{1}{b}\left(\frac{1}{2} + O\big(b^2\big)\right).$$

Additionally, this implies, as $n \to \infty$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} K_b^2\left(\frac{i}{n}\right) &= \int_0^1 K_b^2(u)du + O\left(\frac{1}{n}\right) \\ &= \frac{1}{b}\left(\frac{1}{2} + O\big(b^2\big)\right) + O\left(\frac{1}{n}\right) \\ &= \frac{1}{2b} + O(b)\,.\end{aligned} \tag{5.18}$$

Combining (5.17) and (5.18) gives for the first term of (5.16), as $n \to \infty$,

$$\mathbb{V}\Big(\tilde{c}_{n,b}\big(t^{\pm}\big)\Big) = \frac{t^{\pm}}{k}\left[\frac{1}{2b}c(1) + O(1) + O\left(\tilde{A}\left(\frac{n}{kt^{\pm}}\right)\right)\left(\frac{1}{2b} + O(b)\right)\right]. \tag{5.19}$$

By the same arguments as for the mean, $t^{\pm}/k = 1/k \pm o(1/(k\sqrt{kb}))$ and $o(1/(k\sqrt{kb})) = o(1/k) = o(1/(kb))$, such that (5.19) reduces to the expression in the proposition. For the second term in (5.16), Lemma 2 can be extended to c^2 in a straightforward manner, which gives an expression of the same order: $c^2(1) = O(1)$. Combined with (5.17) and (5.18) and because $(t^{\pm})^2/n = O(1/n)$, it can be shown that the second term in (5.16) is $O(1/(nb)) = o(1/(kb))$, which completes the proof for Proposition 1.

□

Now, given Proposition 1, we use the Lindeberg-Feller theorem, which allows for independent but non-identically distributed random variables, to prove asymptotic normality of $\tilde{c}_{n,b}\big(t^{\pm}\big)$. First, let

$$\begin{aligned}Y_i &\coloneqq \frac{1}{k}\left[\mathbb{1}_{\left\{X_i > U\left(n/(kt^{\pm})\right)\right\}} K_b\left(\frac{i}{n}\right) - \mathbb{E}\left(\mathbb{1}_{\left\{X_i > U\left(n/(kt^{\pm})\right)\right\}} K_b\left(\frac{i}{n}\right)\right)\right] \\ &= \frac{1}{k}K_b\left(\frac{i}{n}\right)\left[\mathbb{1}_{\left\{X_i > U\left(n/(kt^{\pm})\right)\right\}} - \mathbb{P}\left(X_i > U\left(\frac{n}{kt^{\pm}}\right)\right)\right] \\ &\coloneqq \frac{1}{k}K_b\left(\frac{i}{n}\right) Z_i\end{aligned} \tag{5.20}$$

and

$$\begin{aligned}s_n^2 &\coloneqq \sum_{i=1}^{n} \mathbb{V}(Y_i) \\ &= \frac{1}{k^2}\sum_{i=1}^{n} K_b^2\left(\frac{i}{n}\right)\mathbb{V}\left(\mathbb{1}_{\left\{X_i > U\left(n/(kt^{\pm})\right)\right\}}\right) \\ &= \mathbb{V}\Big(\tilde{c}_{n,b}\big(t^{\pm}\big)\Big)\,,\end{aligned} \tag{5.21}$$

of which the latter follows directly from (5.16). Then, following van der Vaart (2000),

among others, the Lindeberg-Feller condition for sequences assures that $\sum_{i=1}^{n} Y_i/s_n \xrightarrow{d} \mathcal{N}(0,1)$ if for every $\omega > 0$, as $n \to \infty$,

$$L_n \coloneqq \frac{1}{s_n^2} \sum_{i=1}^{n} \mathbb{E}\left(Y_i^2 \mathbb{1}_{\{|Y_i| \geq \omega s_n\}} \right) \to 0. \tag{5.22}$$

This condition essentially imposes that each of the individual variances of the independent, but possibly non-identically distributed variables, however large, is sufficiently small compared to the total variance. Combining (5.20) and (5.22) gives

$$\begin{aligned} L_n &= \sum_{i=1}^{n} \left(\frac{K_b\big(i/n\big)}{k s_n} \right)^2 \mathbb{E}\left(Z_i^2 \mathbb{1}_{\left\{|Z_i| \geq \omega k s_n / K_b\left(i/n\right)\right\}} \right) \\ &\coloneqq \sum_{i=1}^{n} I_{n,i}^{-2} \mathbb{E}\left(Z_i^2 \mathbb{1}_{\{|Z_i| \geq \omega I_{n,i}\}} \right) \\ &\coloneqq \sum_{i=1}^{n} J_{1,n,i} J_{2,n,i}. \end{aligned}$$

Clearly, $|Z_i| \leq 1$, implying $J_{2,n,i} \to 0$ whenever $\omega I_{n,i} > 1$, which is always satisfied if $I_{n,i} \to \infty$. Similarly, $J_{1,n.i} \to 0$ when $I_{n,i} \to \infty$. Therefore, we need $I_{n,i} \to \infty$ as a sufficient condition for $L_n \to 0$ as $n \to \infty$. Using Proposition 1 for s_n, we get

$$I_{n,i} \sim \frac{\sqrt{c(1)/2}}{\sqrt{b/k} K_b(i/n)},$$

where, following (5.2), the denominator is

$$\begin{aligned} \sqrt{\frac{b}{k}} K_b\left(\frac{i}{n}\right) &= \sqrt{\frac{b}{k}} \left(\frac{1}{b} + 1 \right) \left(\frac{i}{n} \right)^{1/b} \\ &= \left(\frac{1}{\sqrt{kb}} + \sqrt{\frac{b}{k}} \right) \left(\frac{i}{n} \right)^{1/b} \\ &\coloneqq M_1 M_{2,n,i}. \end{aligned}$$

It holds that $M_{2,n,i} \in (0,1]$ and, as $n \to \infty$, $M_1 \to 0$ because $kb \to \infty$ by assumption. Therefore, $M_1 M_{2,n,i} \to 0$. Together with the assumption that c is positive and bounded on $[0,1]$, $I_{n,i} \to \infty$ for every n, i, and the Lindeberg-Feller condition is satisfied. Using

(5.20) and (5.21), this implies that, as $n \to \infty$, for some λ_b such that $k^{1/3}b \to \lambda_b \in [0, \infty)$,

$$\begin{aligned}\frac{\sum_{i=1}^{n} Y_i}{s_n} &= \frac{\tilde{c}_{n,b}\big(t^{\pm}\big) - \mathbb{E}\Big(\tilde{c}_{n,b}\big(t^{\pm}\big)\Big)}{\sqrt{\mathbb{V}\big(\tilde{c}_{n,b}(t^{\pm})\big)}} \\ &\sim \frac{\tilde{c}_{n,b}\big(t^{\pm}\big) - c(1) + bc'(1)}{\sqrt{c(1)/2kb}} \\ &= \frac{\sqrt{kb}\big(\tilde{c}_{n,b}\big(t^{\pm}\big) - c(1)\big) - \big(-\lambda_b^{3/2}c'(1)\big)}{\sqrt{c(1)/2}},\end{aligned}$$

convergences in distribution to a standard normal variable, which is equivalent to (5.10).

Expressions (5.7), (5.9), and (5.10) together prove Theorem 1.

□

Proof of Theorem 2

For this proof, we assume asymptotic independence of $\hat{c}_{n,b}(1)$ and $\hat{\gamma}_n$. For some intuitive arguments why the dependence between these estimators may become negligible, we refer to Appendix A.2.

Now, let

$$\theta \coloneqq (c(1), \gamma)^{\intercal}, \quad T_n \coloneqq (\hat{c}_{n,b}(1)\,, \hat{\gamma}_n)^{\intercal}, \quad \text{and} \quad \mu \coloneqq \big(-\lambda^{3/2}c'(1), 0\big)^{\intercal} \tag{5.23}$$

be 2×1 row vectors, with $^{\intercal}$ denoting their transpose, and

$$\Sigma \coloneqq \operatorname{diag}\left(\frac{1}{2}c(1), b\gamma^2\right) \tag{5.24}$$

be a 2×2 diagonal matrix. Then, under the assumption of asymptotic independence of $\hat{c}_{n,b}(1)$ and $\hat{\gamma}_n$ and using result (3.15) and Theorem 1, it holds that, as $n \to \infty$,

$$\sqrt{kb}(T_n - \theta) \xrightarrow{d} \mathcal{N}(\mu, \Sigma)\,, \tag{5.25}$$

where in this case $\mathcal{N}$ denotes a bivariate normal distribution. Then, following van der Vaart (2000), among others, the delta method assures that, for a function $\phi : \mathbb{R}^2 \mapsto \mathbb{R}$ defined on a subset of $\mathbb{R}^2$ and differentiable at θ, with $\nabla\phi$ denoting its gradient, and the random vector T_n taking values in the domain of ϕ,

$$\sqrt{kb}(\phi(T_n) - \phi(\theta)) \xrightarrow{d} \mathcal{N}\big(\nabla\phi(\theta)^{\intercal}\mu, \nabla\phi(\theta)^{\intercal}\Sigma\nabla\phi(\theta)\big)\,. \tag{5.26}$$

Here $\mathcal{N}$ is a univariate normal distribution. Now let

$$\phi(\theta) := \left(\frac{kc(1)}{np}\right)^{\gamma},$$

such that

$$\frac{\partial\phi}{\partial c(1)} = \phi(\theta)\frac{\gamma}{c(1)}, \quad \frac{\partial\phi}{\partial\gamma} = \phi(\theta)\log\left(\frac{kc(1)}{np}\right),$$

and $\nabla\phi = (\partial\phi/(\partial c(1)), \partial\phi/(\partial\gamma))^{\intercal}$. Then it follows that

$$\nabla\phi(\theta)^{\intercal}\mu = \phi(\theta)\left(-\lambda^{3/2}\frac{\gamma c'(1)}{c(1)}\right) \tag{5.27}$$

and

$$\nabla\phi(\theta)^{\intercal}\Sigma\nabla\phi(\theta) = \phi^2(\theta)\gamma^2\left(\frac{1}{2c(1)} + b\log\left(\frac{kc(1)}{np}\right)^2\right). \tag{5.28}$$

Further,

$$b\log\left(\frac{kc(1)}{np}\right)^2 = b\left(\log\left(\frac{k}{np}\right)^2 + 2\log\left(\frac{k}{np}\right)\log c(1) + \log\big(c(1)\big)^2\right) \to \beta_b^2,$$

as $n \to \infty$. Note that, following the extreme quantile prediction (5.3),

$$\frac{\phi(T_n)}{\phi(\theta)} = \frac{\hat{U}_{n,n+1}(1/p)}{X_{n,n-k}\phi(\theta)}. \tag{5.29}$$

Thus, we have to prove that $X_{n,n-k}\phi(\theta)$ converges to $U_{n,n+1}(1/p)$. Letting $t_n = (n/k)(1 - F(X_{n,n-k}))$, as (5.7) implies $t_n = 1 + o_p(1/\sqrt{kb})$, and using the continuous mapping theorem and condition (3.4), we get, as $n \to \infty$,

$$U_{n,n+1}\left(\frac{1}{p}\right) \sim U\left(\frac{n}{kt_n}\right)\left(\frac{kt_nc(1)}{np}\right)^{\gamma} = X_{n,n-k}\phi(\theta) + o_p\left(\frac{1}{\sqrt{kb}}\right).$$

Note that this is also a consequence of result A.8 in Einmahl et al. (2016), which implies that $X_{n,n-k}/U(n/k) = U(n/(kt_n))/U(n/k) = 1 + o_p(1/\sqrt{kb})$. In combination with (5.26), (5.27), (5.28), (5.29), this proves Theorem 2.

□

6 Monte Carlo simulation

We continue with a Monte Carlo simulation study on extreme quantile prediction in the heteroscedastic extremes framework for both our beta kernel scedasis estimator and the convolution-type biweight boundary kernel estimator considered in Einmahl et al. (2016). The goal of this simulation study is to verify our derived asymptotic results and compare the two estimators for a variety of sample sizes and data generating processes.

For an overview of the programs and functions implemented for our simulation, we refer to Appendix A.6.

6.1 Data generating processes

We consider six data-generating processes (DGPs) in the heavy-tailed case of the heteroscedastic extremes framework. Each independent observation $i = 1, \ldots, n$ follows a scaled Fréchet distribution with cdf

$$F_{n,i}^{(d)} = \exp\left(-\left(\frac{x}{c_d(i/n)}\right)^{-1}\right), \quad \text{for } x > 0, \tag{6.1}$$

where c_d, for $d = 1, \ldots, 6$, denotes the scedasis function corresponding to each of the following six DGPs.

- DGP 1: $c_1(s) = 1$, is the iid or homoscedastic extremes case.
- DGP 2: $c_2(s) = 0.5 + s$, is a linear trend in the relative frequency of extremes.
- DGP 3: $c_3(s) = 0.5 + 2s$, for $s \in [0, 0.5]$, and $c_3(s) = 2.5 - 2s$, for $s \in (0.5, 1]$, is a symmetric pike.
- DGP 4: $c_4(s) = 0.8$, for $s \in [0, 0.4] \cup [0.6, 1]$, $c_4(s) = -7.2 + 20s$, for $s \in (0.4, 0.5]$, and $c_4(s) = 12.8 - 20s$, for $s \in (0.5, 0.6]$, is a symmetric pike with a flat start and end.
- DGP 5: $c_5(s) = 0.5 + 0.5\exp(s)/(e - 1)$, such that $c'(s)/c''(s) = 1$, is gradually exponentially increasing.
- DGP6: $c_6(s) = 0.5 + 5\exp(10s)/(\exp(10) - 1)$, such that $c'(s)/c''(s) = 10$, is abruptly exponentially increasing.

The first four DGPs are identical to those in Einmahl et al. (2016). Recall that proper estimation of $c(1)$ is important for extreme quantile prediction, which clearly follows from (3.9) and (5.3). Thus, we consider the additional DGPs, 5 and 6, which are mixtures of the log-linear model in Mefleh et al. (2020), for two reasons. First, DGPs 1 through

Table 6.1: Relevant characteristics of the scedasis functions at their right boundary point for each DGP.

DGP	$c(1)$	$c'(1)$	$c''(1)$
1	1.000	0.000	0.000
2	1.500	0.500	0.000
3	0.500	−2.000	0.000
4	0.800	0.000	0.000
5	1.291	0.791	0.791
6	5.500	50.002	500.023

4 have scedasis functions with zero second order derivatives. This is unrealistic in an application and makes it unfair to compare the beta kernel with the convolution-type boundary kernel estimator, which have biases depending on the first and second order derivatives, respectively. The exponential form of the scedasis functions of DGP 5 and 6 makes it easy to compare this part of the bias of both estimators. Second, in risk management, especially from a perspective of regulation and prudence, it is important that the estimators perform well in periods with relatively higher occurrences of extremes. The additional scedasis functions reflect this by their exponential increase: gradually in DGP 5 and more abruptly in DGP 6. An overview of relevant characteristics of each of the scedasis functions at their right boundary point is shown in Table 6.1 and a plot of each can be found in Figure A.1 in Appendix A.3.

Using the inverse cdf method and (6.1), the observations for each simulation replication r can be generated by

$$X_i^{(d,r)} = c_d(i/n) \log\Big(-W_i^{(r)}\Big)^{-1},$$

where $W_i^{(r)}$ are iid standard uniform (pseudo)random variables. The theoretical extreme quantile at time $n+1$ is equal to

$$U_{n,n+1}^{(d)}\left(\frac{1}{p}\right) = U_{n,n}^{(d)}\left(\frac{1}{p}\right) = c_d(1) \log\left(\frac{1}{1-p}\right)^{-1},$$

which we use to evaluate the outcomes of the predictions.

6.2 Methodology

We simulate 1000 samples of sizes $n = 5000, 1250, 625$ and take $k = 400, 100, 50$ for each n, respectively, for each of the six DGPs. This follows the set-up in Einmahl et al. (2016), who consider $n = 5000$ and take $k = 400$, but with additional smaller sample sizes. In each replication, we make extreme quantile predictions $\hat{U}_{n,n+1}$ following (5.3) and (3.9) for

$p = 0.02$, using both the beta kernel estimator (5.2) and the convolution-type boundary kernel estimator (4.4) with biweight kernel (4.2), respectively. We use the Hill estimator (2.13) to estimate the extreme value index. Note that, theoretically, we need $k/n \to 0$ for the asymptotic properties of the estimators to hold, but we keep this ratio constant for simplicity. Moreover, since the data are very heavy-tailed, as $\gamma = 1$, the regularly varying behaviour of the tail of the distribution is attained for relatively low values of x and thus higher values of k suffice (Embrechts et al., 2013).

To verify our asymptotic results, we follow Einmahl et al. (2016) with $n = 5000$ and $k = 400$ and by using the fixed, non-optimised bandwidth $h = 0.1$ for the biweight boundary kernel estimator. In correspondence with our discussion in Section 5.2, we set $b = h^{5/3} \approx 0.022$ for the beta kernel estimator. Moreover, to examine the asymptotic normality, we consider the additional sample size $n = 50000$, with $k = 4000$, and scale the fixed bandwidths according to their local optimal counterparts, i.e. $h = O(1/k^{1/5})$ and $b = O(1/k^{1/3})$.

As suggested in Section 4, comparison of the beta and convolution-type boundary kernel estimators is meaningless without proper bandwidth selection due to the corresponding bias-variance trade-off. Preliminary research suggests that explicitly following asymptotic optimal bandwidth expressions and our cross-validation criterion both lead to inferior results. The former may be due to the complexity of the boundary bias of the symmetric convolution-type kernel. As finding data-driven (local) optimal bandwidths is not the goal of our research, we select approximately optimal bandwidths by minimising the simulation sample MSE. That is, we perform our simulation study over a grid of 200 logarithmically spaced bandwidths such that $b \in [0.01k^{-1/3}, 1000k^{-1/3}]$ for the beta and $h \in [0.01k^{-1/5}, 1000k^{-1/5}]$ for the biweight boundary kernel estimator, respectively, and consider the results corresponding to the bandwidths which lead to minimal simulation sample MSE.

We calculate and report the bias, standard deviation (SD), and root mean squared error (RMSE) of $\hat{U}_{n,n+1}(1/p)/U_{n,n+1}(1/p) - 1$, both as obtained by the simulation sample and by following the theoretical asymptotic values in correspondence with Theorem 2 and result (3.17). For the results corresponding to the fixed, non-optimised bandwidths, we set $\sqrt{b}\log(k/(np)) \to \beta_b = 0$ and $\sqrt{h}\log(k/(np)) \to \beta_h = 0$, following the set-up in Einmahl et al. (2016). However, in our optimised bandwidth case, because the selected bandwidths may become large in some instances, we choose not to neglect this term.

6.3 Results

Table 6.2 shows the results for the fixed, non-optimised bandwidths with $n = 5000$ and $k = 400$, i.e. the set-up in Einmahl et al. (2016). The simulation statistics generally agree with their theoretical asymptotic counterparts, which are reported in parentheses, and

Table 6.2: Simulation results for the one-step ahead predictions with $n = 5000$, $k = 400$, $p = 0.02$, and the fixed, non-optimised bandwidths $h = 0.1$, and $b \approx 0.022$. The theoretical asymptotic results are shown in parentheses. Note that these results are mainly for verification of the asymptotic results and not for comparison between the estimators.

	Beta			Biweight boundary		
DGP	Bias	SD	RMSE	Bias	SD	RMSE
1	-0.005 (0.000)	0.246 (0.241)	0.246 (0.241)	-0.007 (0.000)	0.347 (0.358)	0.347 (0.358)
2	-0.025 (-0.007)	0.207 (0.197)	0.208 (0.197)	-0.013 (0.000)	0.287 (0.292)	0.287 (0.292)
3	0.088 (0.086)	0.342 (0.341)	0.353 (0.351)	0.007 (0.000)	0.491 (0.506)	0.491 (0.506)
4	0.005 (0.000)	0.273 (0.269)	0.273 (0.269)	0.003 (0.000)	0.392 (0.400)	0.392 (0.400)
5	-0.020 (-0.013)	0.219 (0.212)	0.220 (0.212)	-0.009 (-0.000)	0.307 (0.315)	0.307 (0.315)
6	-0.208 (-0.196)	0.140 (0.103)	0.251 (0.221)	-0.101 (-0.040)	0.181 (0.152)	0.207 (0.158)

the results for the biweight boundary kernel estimator are in correspondence with those in Einmahl et al. (2016). As expected, the beta kernel estimator has a nonzero bias for the DGPs with nonzero $c'(1)$ and the convolution-type boundary kernel estimator shows a clear nonzero bias for DGP 6, in which $c''(1) \approx 500$. The quantile-quantile plots in Figure A.2 in Appendix A.4 verify the asymptotic normality of our beta kernel estimator, though the convergence seems rather slow. Following the Jargque-Bera test for normality, at a 1% significance level, the estimator attains normality for $n = 50000$ for all DGPs except DGP 3. However, for $n = 5000$, the corresponding p-values are virtually zero and are therefore not shown, indicating significant deviations from a normal distribution.

Table 6.3 shows the results when optimising over the grid of bandwidths to compare the two estimators. Our beta kernel scedasis estimator performs best in terms of the RMSE, though by small margins, for $n = 5000$ in DGP 1 and for $n = 625$ in DGPs 1, 2, and 5. The larger effective sample size does not seem to lead to substantial performance gains relative to the biweight boundary kernel estimator. While the variance is slightly lower in some instances for the smaller samples, which is conform the findings of Chen (1999), this is generally outweighed in the RMSE by a larger bias. The beta kernel estimator underperforms for the symmetric scedasis functions in DGPs 3 and 4, where its simulation and theoretical statistics are often multiples of that of the biweight boundary kernel estimator. There is also no clear advantage for the beta kernel estimator during a period of an exponentially increasing frequency of extremes: while $c''(1)/c'(1) = 10$ in DGP 6, the bias of the biweight boundary kernel estimator is always relatively smaller.

Table 6.3: Simulation results of the one-step ahead predictions for the optimised bandwidths and $p = 0.02$. The theoretical asymptotic values are shown in parentheses. These results allow for comparison between the estimators.

		Beta				Biweight boundary			
DGP	n	Bias	SD	RMSE	b	Bias	SD	RMSE	h
1	5000	0.000 (0.000)	0.084 (0.069)	0.084* (0.069)	135.721	0.016 (0.000)	0.087 (0.089)	0.088 (0.089)	4.171
	1250	0.002 (0.000)	0.178 (0.139)	0.178 (0.139)	215.443	-0.012 (0.000)	0.175 (0.176)	0.176* (0.176)	4.367
	625	0.014 (0.000)	0.249 (0.196)	0.249* (0.196)	271.442	0.023 (0.000)	0.254 (0.250)	0.255 (0.250)	4.217
2	5000	-0.067 (-0.037)	0.117 (0.111)	0.135 (0.117)	0.110	0.001 (0.000)	0.103 (0.093)	0.103* (0.093)	2.208
	1250	-0.096 (-0.065)	0.214 (0.190)	0.235 (0.201)	0.196	0.006 (0.000)	0.221 (0.187)	0.221* (0.187)	2.181
	625	-0.099 (-0.093)	0.282 (0.250)	0.299* (0.266)	0.278	0.007 (0.000)	0.301 (0.259)	0.300 (0.259)	2.365
3	5000	0.166 (0.175)	0.255 (0.249)	0.305 (0.304)	0.044	-0.014 (0.000)	0.060 (0.085)	0.062* (0.085)	10.527
	1250	0.254 (0.294)	0.432 (0.394)	0.501 (0.492)	0.073	-0.007 (0.000)	0.124 (0.170)	0.124* (0.170)	10.401
	625	0.361 (0.416)	0.562 (0.481)	0.668 (0.635)	0.104	-0.027 (0.000)	0.165 (0.240)	0.167* (0.240)	10.642
4	5000	0.037 (0.000)	0.121 (0.119)	0.126 (0.119)	0.165	-0.003 (0.000)	0.075 (0.087)	0.075* (0.087)	5.903
	1250	0.080 (0.000)	0.235 (0.211)	0.248 (0.211)	0.248	0.005 (0.000)	0.159 (0.174)	0.159* (0.174)	5.832
	625	0.133 (0.000)	0.326 (0.280)	0.352 (0.280)	0.312	-0.005 (0.000)	0.216 (0.245)	0.216* (0.245)	5.967
5	5000	-0.067 (-0.090)	0.109 (0.107)	0.128 (0.140)	0.147	0.007 (-0.210)	0.096 (0.091)	0.096* (0.229)	2.782
	1250	-0.095 (-0.170)	0.200 (0.182)	0.222 (0.249)	0.278	-0.009 (-0.230)	0.200 (0.181)	0.200* (0.293)	2.913
	625	-0.104 (-0.270)	0.255 (0.237)	0.276* (0.359)	0.441	0.024 (-0.215)	0.287 (0.258)	0.288 (0.336)	2.813
6	5000	-0.120 (-0.062)	0.193 (0.195)	0.227 (0.205)	0.007	-0.102 (-0.043)	0.180 (0.166)	0.206* (0.171)	0.103
	1250	-0.119 (-0.132)	0.380 (0.286)	0.398 (0.315)	0.015	-0.101 (-0.132)	0.352 (0.266)	0.366* (0.297)	0.181
	625	-0.070 (-0.187)	0.554 (0.356)	0.558 (0.402)	0.021	-0.052 (-0.196)	0.543 (0.350)	0.545* (0.401)	0.221

* Performing best in terms of RMSE for specific DGP and sample size.

The results in Table 6.3 for the standard deviation and RMSE also clearly agree with the theoretical asymptotic order of convergence, $\sqrt{k}$. Ignoring the influence of the bandwidth, the standard deviation increases approximately by a factor of $\sqrt{4} = 2$ and $\sqrt{2} \approx 1.414$ as k decreases with the sample size n from 400 to 100 and 100 to 50, respectively. Further, as expected, the difference between the theoretical and simulation statistics becomes larger as the sample size decreases. Note that the asymptotic statistics are based on $b, h \to 0$. However, the selected bandwidths are often not small, which is an explanation for deviations between the theory and the simulation results, especially for the largest sample size.

The selected approximately optimal bandwidths for the beta kernel estimator are generally also in close accordance with the theory, increasing as the sample size decreases with a factor corresponding to result (5.5). However, this is not true for the instances where $c'(1) = 0$, and similarly for the biweight boundary kernel estimator when $c''(1) = 0$. In these cases, the theory implies asymptotically optimal bandwidths $b^*, h^* \to \infty$, which results in unexpected bandwidth choices.

7 Empirical application

We now demonstrate the heteroscedastic extremes statistics, in particular the beta kernel scedasis estimator, on a financial data set. Note that in the field of finance, an extreme quantile of the loss distribution is generally called the Value-at-Risk (VaR), such that $U(1/p) = \text{VaR}_{1-p}$. Therefore, we essentially make VaR predictions for the first day of 2008 with $\hat{U}_{n,n+1}$. The goal of this empirical application is not to showcase a complete methodological framework, but to guide one through the steps to use the heteroscedastic extremes model with the knowledge we have now.

For an overview of the programs and functions implemented for our application, we refer to Appendix A.6.

7.1 Data

We consider the 20-years daily log-losses from the S&P 500 index from 1988 until 2007, which is identical to the data set and period consider by Einmahl et al. (2016). Excluding non-trading days, this results in 5043 observations, of which 2350 are losses. This period does not cover observations related to the black Monday crash on October 19th, 1987, and the Global Financial Crisis in 2008, but includes the crash of the dot-com-bubble between 2000 and 2002 and other sharp market fluctuations. Additionally, we consider the 4-years subsample of 2004 until 2007 and 1-year subsample of 2007, which have 1006 and 251 observations, respectively. Some relevant summary statistics can be found in Table 7.1(a). The statistics are given for the log-losses in percentages.

The heteroscedastic extremes model is valid for this data, specifically for the full sample period, for the following reasons. Jansen and de Vries (1991) and Kearns and Pagan (1997), among many others, find that losses of financial equity have heavy-tailed distributions. With this empirical finding, reflecting assumption (2.11), together with condition (3.1), Einmahl et al. (2016) find no significant evidence that the extreme value index varies over time. These constitute two important assumptions with respect to the heteroscedastic extremes framework. Additionally, Einmahl et al. (2016) find strong evidence that the scedasis function varies over time, meaning that the heteroscedastic extremes model is relevant. Therefore, following their conclusion, the frequency of extremes changes over time while their size remains the same. Einmahl et al. (2016) further show robustness by replicating these results with weekly observations, suggesting the absence of significant serial dependence. To shorten our discussion, because testing is no important part of our research as a whole, and as the other two sample periods are subsamples of what is discussed above, we assume that the conditions for the heteroscedastic extremes model are met for the subperiods as well.

7.2 Methodology

For each of the three sample periods, we estimate the scedasis and extreme quantile predictions functions. We consider both the beta kernel estimator (5.1) and convolution-type biweight kernel estimator (3.7) with (4.2), for which in the latter case we use the boundary kernel (4.4) in the boundary regions. For the extreme value index, we use the Hill estimator (2.13). For further comparison, we also consider the homoscedastic extreme quantile estimator (2.17) and the empirical quantile function.

As suggested in Section 5.2, we choose the bandwidth and number of upper order statistics sequentially, selecting k in the Hill plot as approximately the middle of the first stable part and the bandwidth in relation to k. As we leave data-driven bandwidth selection for scedasis estimation to future research, we set $h = 0.1$ as the global and local bandwidth of the biweight kernel estimator for the full sample period, directly following Einmahl et al. (2016). As discussed in Section 5.2, we adjust this bandwidth for the subsamples and the beta kernel estimator in correspondence to their local and global optimal orders. That is, for the convolution-type boundary kernel estimator, $h = O(1/k^{1/5})$, both globally and locally. Then, for the beta kernel estimator, we set $b = h^{5/3}$ and $b = h^2$ for the local and global scedasis estimates, respectively. We emphasise that these bandwidth choices are not optimal and so neither are our results.

7.3 Results

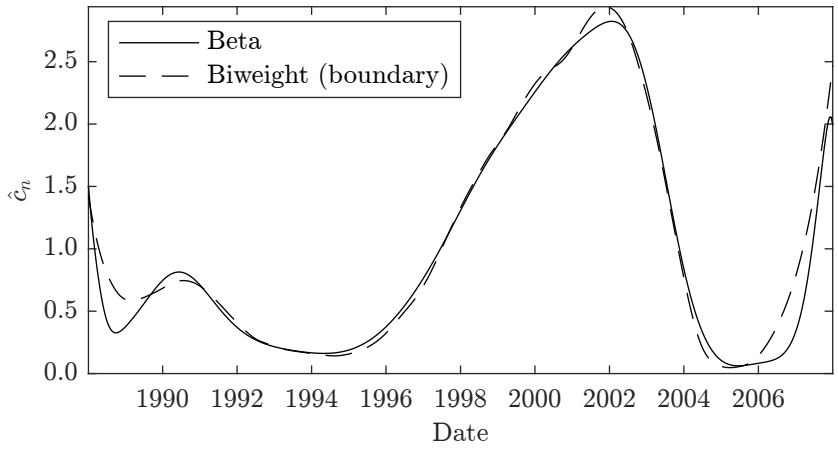
Table 7.1(b) gives an overview of the relevant extreme value statistics for each sample period. The Hill plots corresponding to the choice of k can be found in Appendix A.5,

Table 7.1: Summary statistics, extreme value statistics, and extreme quantile predictions for the S&P 500 log-losses, in percentages, for the three sample periods.

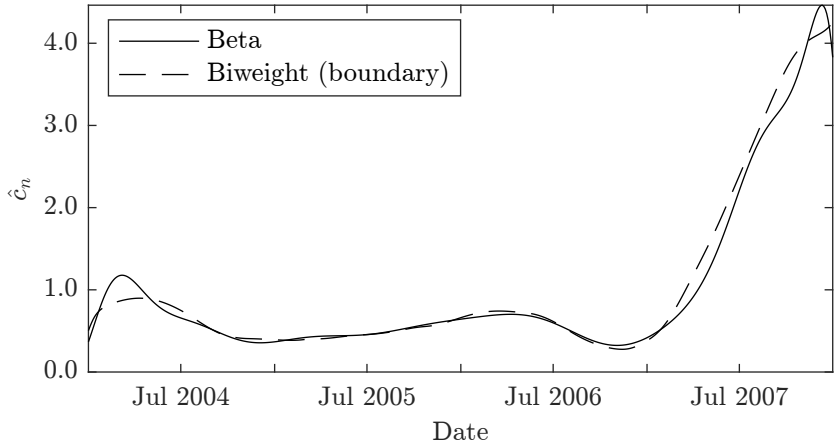
		1988-2007	2004-2007	2007
(a) *Summary statistics*				
n		5043	1006	251
Losses		2350	448	114
Mean		-0.035	-0.028	-0.014
Median		-0.052	-0.073	-0.081
Maximum		7.113	3.534	3.534
(b) *Extreme value statistics*				
k		180	32	25
$\hat{\gamma}_n$		0.298	0.291	0.367
$X_{n,n-k}$		1.806	1.445	1.369
Beta	$\hat{c}_{n,b}(1)$	1.682	3.905	2.194
	b	0.022	0.038	0.042
Boundary	$\hat{c}_{n,b}(1)$	2.471	4.252	2.236
	h	0.100	0.141	0.148
(c) *Extreme quantile predictions* $\hat{U}_{n,n+1}(1/p)$				
Beta	$p = 0.05$	1.907	1.883	2.355
	$p = 0.01$	3.078	3.007	4.253
Boundary	$p = 0.05$	2.138	1.930	2.371
	$p = 0.01$	3.451	3.083	4.283
Homoscedastic	$p = 0.05$	1.633	1.267	1.764
	$p = 0.01$	2.637	2.024	3.187
Empirical	$p = 0.05$	1.592	1.304	1.829
	$p = 0.01$	2.627	2.187	2.978

giving the interested reader an opportunity to judge for him- or herself. The extreme value index estimates $\hat{\gamma}_n$ for the 20- and 4-years periods lie around 0.3, which is in line with some examples in the literature, for instance, Nolde and Zhou (2021). For the 1-year 2007 period, the estimate is substantially larger, suggesting infinite variance data as $\alpha = 1/\gamma < 3$ as explained in Section 2.1. However, repeating Section 7.1, we assume γ does not vary significantly between the subsamples. Further, for the 2007 subperiod, we choose $k = 25$, which is a substantially larger share of the upper-order statistics than for the other samples and close to the 32 chosen for the 2004-2007 subperiod. However, heavier-tailed data, as implied by the larger $\hat{\gamma}_n$, attains tail behaviour for lower upper-order statistics.

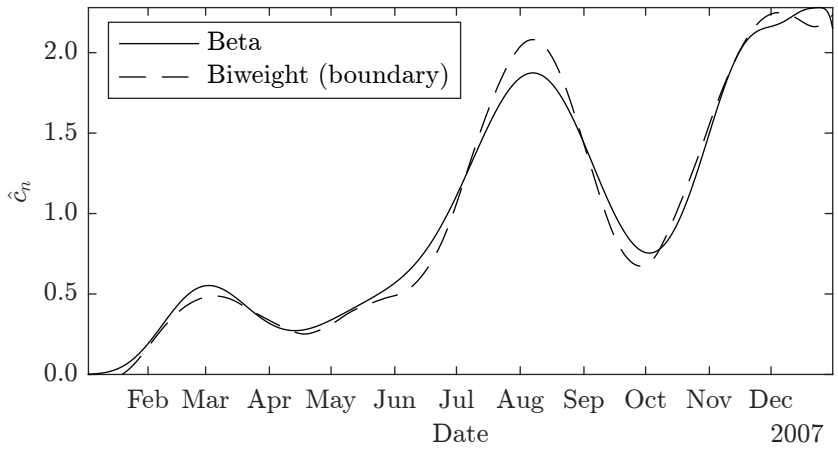
Figure 7.1 shows the global estimates of the scedasis function. For the full sample period 1988-2007 (a), these are similar to those in Einmahl et al. (2016). There clearly is an increased frequency of extremes during the dot-com-bubble around the year 2000,



(a) 20-Years period 1988-2007.



(b) 4-Years period 2004-2007.



(c) 1-Year period 2007.

Figure 7.1: Estimated scedasis functions for both the beta kernel and convolution-type biweight (boundary) kernel estimator for the S&P 500 log-losses over the 20-years period 1988-2007 (top), 4-years period 2004-2007 (middle), and 1-year period 2007 (bottom).

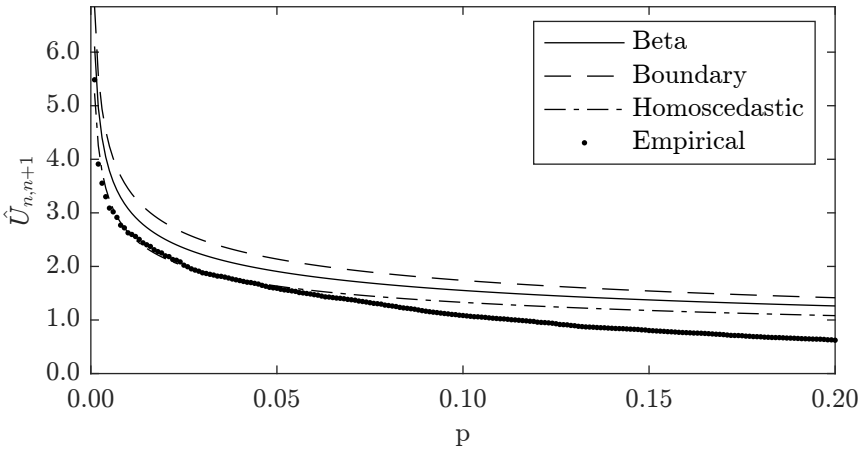
which starts to decrease after the burst around 2002, and an increase in the scedasis in 2007, which corresponds to the start of the financial crisis of 2008. Visually, the boundary

region appears similar to the sixth DGP in our simulation study. The scedasis function estimates for the subsample 2004-2007 (b) show a similar large increase in 2007, which indicate that most of the extreme losses occur in 2007. This explains the proximity of the choices of k between the 2004-2007 and 2007 subsamples. For the 1-year subperiod 2007 (c) there is a rough positive trend with local fluctuations. The beta and convolution-type kernel estimates generally agree globally. However, the boundary kernel results in negative scedasis estimates in the left boundary region in Figure 7.1(c), which highlights a potential drawback of this estimator.

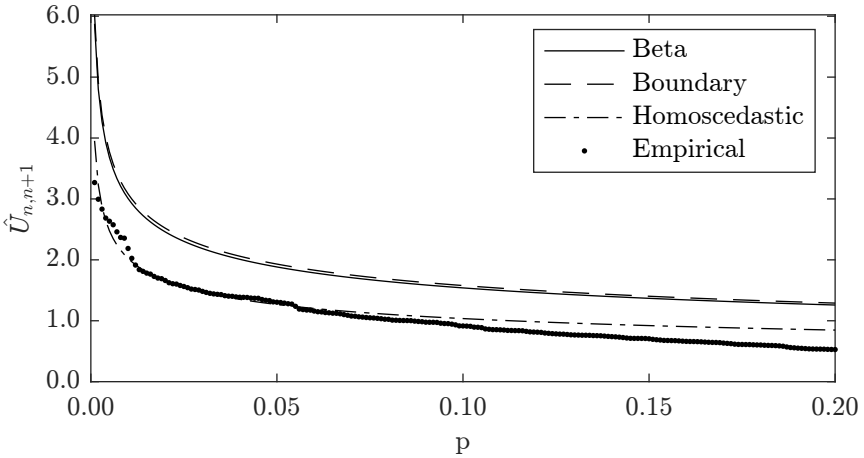
Table 7.1(b) further presents the $c(1)$ estimates. For the 1988-2007 sample period, the beta kernel estimate of approximately 1.7 is substantially lower than that of the biweight boundary kernel of about 2.5. This may be a result of the downward bias of the beta kernel estimator, as the scedasis slope is large in this region. The $c(1)$ estimates of both kernel estimators are close together for the other two subsample periods. The relative frequency of extremes is about four times larger than the baseline homoscedastic case at the end of the 2004-2007 subsample period. For the 1-year 2007 subsample, in which the financial crisis of 2008 has already started, the estimates suggest a scedasis of approximately 2 at the end of 2007.

Condition (3.4) and the estimators (3.9) and (5.3) imply that higher scedasis results in higher extreme quantiles, which is reflected by the estimated extreme quantile predictions in Figure 7.2 and Table 7.1(c). The estimates corresponding to the heteroscedastic extremes model, for both the beta and boundary kernel estimators, lie above those of the static homoscedastic and empirical estimators for all sample periods, suggesting higher VaR predictions than the homoscedastic model. The richness of our considered data set also allows us to highlight the accuracy of the tail estimates of the iid EVT model in the static case, as the homoscedastic extreme quantile estimates closely follow their empirical counterparts in the upper 5% and 5-10% region for the two largest samples and the smallest sample, respectively. However, whereas the heteroscedastic extremes model estimators capture the emerging crisis in the extreme quantile predictions in every sample, the other estimators do this only marginally for the 1-year 2007 subperiod.

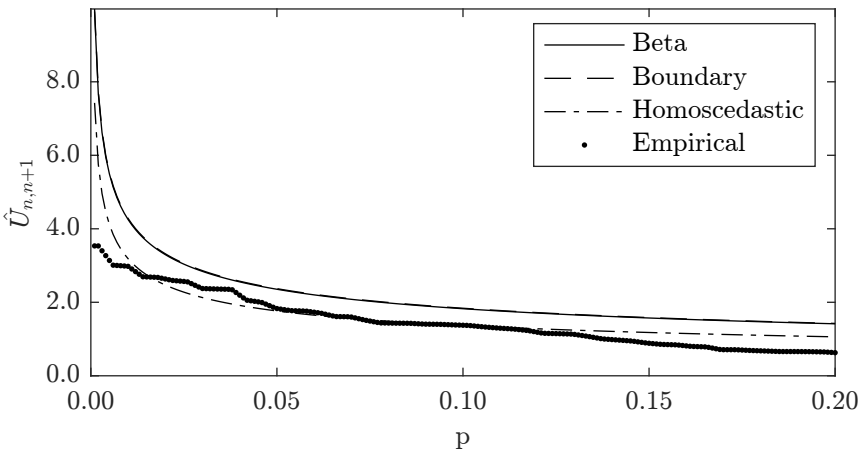
Due to the smaller sample size, only the 25 upper-order statistics are considered for the extreme value statistics corresponding to the 1-year 2007 subsample. A recurring subject in our research, this presents a typical bias-variance trade-off. Though the smaller sample size may lead to relatively more uncertainty (i.e. higher variance) in the statistical results, zooming into 2007 may better reflect the characteristics of the distribution underlying the losses during the emerging financial crisis (i.e. lower bias). But, based on our results, we find that the heteroscedastic extremes model is also better at capturing changes in economic environment than the two traditional approaches. We note, however, that this comes at the cost of other uncertainties, such as the selection of proper bandwidth parameters.



(a) 20-Years period 1988-2007.



(b) 4-Years period 2004-2007.



(c) 1-Year period 2007.

Figure 7.2: Estimated one-step ahead extreme quantile prediction functions for the beta kernel and convolution-type biweight boundary kernel heteroscedastic extremes estimators and the homoscedastic (iid EVT) and empirical estimators for the first trading day of 2008 using the S&P 500 log-losses over the 20-years period 1988-2007 (top), 4-years period 2004-2007 (middle), and 1-year period 2007 (bottom).

8 Conclusion

In this research, we propose a non-parametric beta kernel scedasis estimator for the heavy-tailed heteroscedastic extremes model. In contrast to traditional convolution-type kernel estimators, its compact support matches that of the scedasis function, leading to a larger effective sample size, and it has no boundary bias problem. Therefore, we are interested in its properties for extreme quantile prediction, individually and in comparison to a convolution-type boundary kernel scedasis estimator.

We establish the asymptotic normality of our beta kernel scedasis estimator at the right boundary point and, under the assumption of asymptotic independence of the scedasis and extreme value estimator, of the corresponding extreme quantile prediction. We further find that its local optimal MSE convergence is of larger asymptotic order than that of the convolution-type estimator. Our asymptotic results are verified by a simulation study, but the beta kernel estimator generally does not outperform a convolution-type boundary kernel counterpart in terms of MSE in a variety of sample sizes and data generating processes. This suggest that the larger effective sample size does not lead to meaningful finite-sample performance increases. In our empirical application to S&P 500 loss data, zooming into the emerging financial crisis of 2008, both heteroscedastic extremes estimators are generally in close correspondence and suggest higher extreme quantile predictions than traditional extreme value and empirical statistics.

We conclude that our beta kernel scedasis estimator has clear asymptotic statistical properties for extreme quantile prediction, which are in agreement with existing literature. Its performance is generally not better than that of the convolution-type boundary kernel estimator, though it is often similar in finite sample applications. Therefore, we think it is potentially useful in practice.

A few comments are in order. First, we pragmatically consider (approximately) optimal bandwidths in the simulation study and use fixed, non-optimal bandwidths in the empirical application. However, a solid methodology for data-driven bandwidth selection for kernel scedasis estimation, deserving a study on its own, may facilitate a more realistic comparison of the estimators in practice. Second, as the problems of selecting the bandwidth and the number of upper order statics are interconnected, insight in simultaneous selection may provide better strategies than our sequential approach. Moreover, data-driven selection methods may allow the heteroscedastic extremes model to be formally compared to existing approaches for extreme quantile prediction in practical applications. Finally, while our asymptotic results are verified by a simulation study, a formal proof for the asymptotic independence of the beta kernel scedasis and extreme value estimators, if possible, may strengthen the basis of our theory.

Acknowledgements

I would like to thank my thesis supervisor, Alex Koning, for his pleasant and fruitful discussions throughout the process of writing this thesis and for the time and opportunity he gave to explore and work on some of the more theoretical aspects of the subject. Moreover, I would like to thank for the support and patience of my family and friends and the useful comments of my father and girlfriend.

References

Balkema, A. A. and De Haan, L. (1974). Residual life time at great age. *The Annals of probability*, 2(5):792–804.

Chen, S. X. (1999). Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, 31(2):131–145.

Chen, S. X. (2000). Beta kernel smoothers for regression curves. *Statistica Sinica*, 10:73–91.

Danielsson, J., de Haan, L., Peng, L., and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate analysis*, 76(2):226–248.

Danielsson, J. and de Vries, C. G. (1997). Tail index and quantile estimation with very high frequency data. *Journal of empirical Finance*, 4(2-3):241–257.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425.

de Haan, L. (1984). Slow variation and characterization of domains of attraction. *Statistical Extremes and Applications*, 131:31.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, volume 21. Springer Science & Business Media.

de Haan, L. and Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, 52(1):60–70.

de Haan, L., Tank, A. K., and Neves, C. (2015). On tail trend detection: modeling relative risk. *Extremes*, 18(2):141–178.

de Haan, L. and Zhou, C. (2021). Trends in extreme value indices. *Journal of the American Statistical Association*, 116(535):1265–1279.

Dekkers, A. L. and de Haan, L. (1993). Optimal choice of sample fraction in extreme-value estimation. *Journal of Multivariate Analysis*, 47(2):173–195.

Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *The Annals of Applied Probability*, 10(4):1274–1301.

Einmahl, J. H., de Haan, L., and Zhou, C. (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):31–51.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media.

Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes*, 13(2):177–204.

Goegebeur, Y., Guillou, A., and Schorgen, A. (2014). Nonparametric regression estimation of conditional tails: the random covariate case. *Statistics*, 48(4):732–755.

Hall, P. and Marron, J. S. (1991). Local minima in cross-validation functions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(1):245–252.

Heidenreich, N.-B., Schindler, A., and Sperlich, S. (2013). Bandwidth selection for kernel density estimation: a review of fully automatic selectors. *AStA Advances in Statistical Analysis*, 97(4):403–433.

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

Hsing, T. (1991). On tail index estimation using dependent data. *The Annals of Statistics*, 19(3):1547–1569.

Jansen, D. W. and de Vries, C. G. (1991). On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics*, 73(1):18–24.

Jones, M. and Foster, P. (1996). A simple nonnegative boundary correction method for kernel density estimation. *Statistica Sinica*, 6(4):1005–1013.

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

Jones, M. C., Marron, J. S., and Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American statistical association*, 91(433):401–407.

Kearns, P. and Pagan, A. (1997). Estimating the density tail index for financial time series. *Review of Economics and Statistics*, 79(2):171–175.

Koedijk, K. G., Schafgans, M. M., and de Vries, C. G. (1990). The tail index of exchange rate returns. *Journal of international economics*, 29(1-2):93–108.

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

Mandelbrot, B. (1967). The variation of some other speculative prices. *The Journal of Business*, 40(4):393–413.

McElroy, T. (2016). On the measurement and treatment of extremes in time series. *Extremes*, 19(3):467–490.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management: concepts, techniques and tools - revised edition*. Princeton university press.

Mefleh, A., Biard, R., Dombry, C., and Khraibani, Z. (2020). Trend detection for heteroscedastic extremes. *Extremes*, 23(1):85–115.

Nolde, N. and Zhou, C. (2021). Extreme value analysis for financial risk management. *Annual Review of Statistics and Its Application*, 8:217–240.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, volume 26. CRC Press.

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, 4(4):367–377.

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable*, volume 2. Random house.

Thombs, L. A. and Sheather, S. J. (1992). Local bandwidth selection for density estimation. In *Computing Science and Statistics*, pages 111–116. Springer.

van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Wand, M. P. and Jones, M. C. (1994). *Kernel smoothing*. CRC press.

Wang, H. and Tsai, C. L. (2009). Tail index regression. *Journal of the American Statistical Association*, 104(487):1233–1240.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364):812–815.

Zambom, A. Z. and Ronaldo, D. (2013). A review of kernel density estimation with applications to econometrics. *International Econometric Review*, 5(1):20–42.

Zhang, S. and Karunamuni, R. J. (2010). Boundary performance of the beta kernel estimators. *Journal of Nonparametric Statistics*, 22(1):81–104.

A Appendix

A.1 Cross-validation criterion

The following discussion about the cross-validation criterion in kernel estimation is quite general. We refer to Jones et al. (1996) for a review on more bandwidth selection methods.

The integrated squared error for a scedasis estimator $\hat{c}$ is

$$\begin{aligned} \mathrm{ISE}(\hat{c}) &\coloneqq \int_0^1 \big(\hat{c}(s) - c(s)\big)^2 ds \\ &= \int_0^1 \hat{c}^2(s) ds - 2\int_0^1 \hat{c}(s)c(s) ds + \int_0^1 c(s) ds. \end{aligned} \tag{A.1}$$

For a kernel scedasis estimator, minimisation of this expression with respect to the bandwidth parameter gives the well-known cross-validation (CV) criterion. Because, traditionally, this requires the convolution of the kernel function with itself, we derive the CV criterion for the symmetric convolution-type kernel estimator (3.7). A similar criterion for a beta kernel estimator would require numerical integration to estimate the first term of (A.1).

The last term of (A.1) does not involve the estimator and can therefore be ignored. For the first term, using the variable transformation $u = (s - i/n)/h$, we can write

$$\begin{aligned} \int_0^1 \hat{c}^2_{n,h}(s) ds &= \frac{1}{k^2 h} \sum_{i=1}^n \sum_{j=1}^n \mathbb{1}_{\{X_i, X_j > X_{n,n-k}\}} \frac{1}{h} \int_0^1 K\left(\frac{s - i/n}{h}\right) K\left(\frac{s - j/n}{h}\right) ds \\ &\coloneqq \frac{1}{k^2 h} \sum_{i=1}^n \sum_{j=1}^n \mathbb{1}_{\{X_i, X_j > X_{n,n-k}\}} K \star K\left(\frac{i/n - j/n}{h}\right), \end{aligned} \tag{A.2}$$

where $K \star K$ denotes the convolution of the kernel function with itself. For the second

term in (A.1),

$$\begin{aligned}\int_0^1 \hat{c}_{n,h}(s)c(s)ds &= \mathbb{E}\big(\hat{c}_{n,h}(s)\big) \\ &\approx \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\{X_i > X_{n,n-k}\}} \hat{c}_{n,h,-i}(i/n) \\ &\coloneqq \frac{1}{k^2 h}\sum_{i=1}^{n}\sum_{j\neq i}^{n} \mathbb{1}_{\{X_i, X_j > X_{n,n-k}\}} K\left(\frac{i/n - j/n}{h}\right),\end{aligned} \tag{A.3}$$

where $\hat{c}_{n,h,-i}$ is the leave-one-out estimator, linking this criterion to cross-validation. Note that $\hat{c}_{n,h,-i}$ has a varying number of nonzero terms, which is either k or $(k-1)$, but we have chosen k in (A.3) as its divider for simplicity. The CV criterion is then the sum of expressions (A.3) and (A.2), of which minimisation with respect to h gives some approximation to the optimal bandwidth.

Hall and Marron (1991) show that spurious local minima of the CV criterion tend to occur at too small values of the bandwidth rather than at larger ones, which is why Jones et al. (1996) suggest choosing the largest local minimiser of the CV criterion as the bandwidth h. As mentioned in Section 5, 6, and 7, this criterion does not result in satisfactory scedasis estimates, which is why no further results are shown.

A.2 Asymptotic independence assumption for Theorem 2

We briefly discuss the (in)dependence of the scedasis and extreme value estimators. In Theorem 2, we assume that $\hat{c}_{n,b}(1)$ and $\hat{\gamma}_n$ are asymptotically independent. We have no formal proof for this, but provide an intuitive argument.

In contrast to our beta kernel scedasis estimator, the estimator (3.10) by Einmahl et al. (2016) has a (boundary) kernel such that $K(u) = 0$ for $|u| > 1$. This means that, given its bandwidth h, all terms of the estimator equal zero in case $i < (1-h)n$, such that all the corresponding observations receive no weight. Now consider estimator $\hat{\gamma}^*_{(0,1-h]}$, which is $\hat{\gamma}_n$ based on only the first $n(1-h)$ observations. This estimator is clearly asymptotically (note that $\hat{c}_{n,h}(1)$ contains $X_{n,n-k}$) independent of $\hat{c}_{n,h}(1)$. Einmahl et al. (2016) prove that that $\hat{\gamma}^*_{(0,1-h]} - \hat{\gamma}_n = o_p(1/\sqrt{k})$, such that $\hat{c}_{n,h}(1)$ and $\hat{\gamma}_n$ are asymptotically independent.

Our estimator gives nonzero weight to all observations in the sample, as the beta kernel matches the compact support of the scedasis function. Therefore, the proof by Einmahl et al. (2016) cannot be directly extended to our case. However, an intuitive argument why the dependence between $\hat{c}_{n,b}(1)$ and $\hat{\gamma}_n$ becomes asymptotically negligible is as follows. The kernel $K_b(u) = (1/b+1)u^{1/b}$ increasingly concentrates on the right endpoint $u = 1$ as $b \to 0$, which can be seen in Figure 5.1. This implicates that the observations further away from the last observation, X_n, receive increasingly less (though nonzero) weight in $\hat{c}_{n,b}(1)$. More formally, consider some constant $h \in (0,1)$, omitting

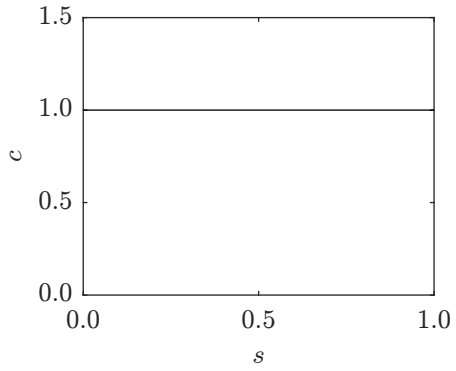
the trivial cases $h = 0, 1$, such that we can write, in correspondence with (5.8),

$$\begin{aligned}\tilde{c}_{n,b}\big(t^{\pm}\big) = \frac{1}{k}\sum_{i=1}^{[(1-h)n]} \mathbb{1}_{\left\{X_i > U\left(n/(kt^{\pm})\right)\right\}} K_b\left(\frac{i}{n}\right) \\ + \frac{1}{k}\sum_{i=[(1-h)n]+1}^{n} \mathbb{1}_{\left\{X_i > U\left(n/(kt^{\pm})\right)\right\}} K_b\left(\frac{i}{n}\right).\end{aligned} \tag{A.4}$$

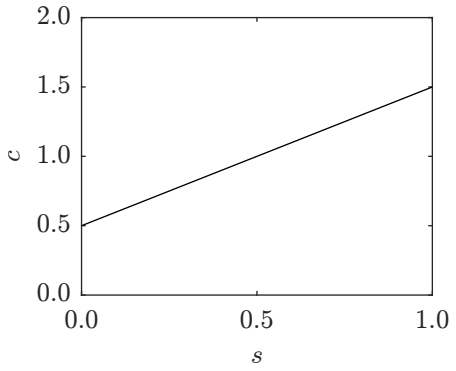
Clearly, the second term is independent of $\hat{\gamma}^*_{(0,1-h]}$. For asymptotic independence, we would need the first term to go to zero as $n \to \infty$. We argue, heuristically, that it becomes asymptotically negligible.

Lastly, as mentioned in Section 3.1, Mefleh et al. (2020) prove asymptotic independence of the exceedance times and corresponding exceedance values, where the former is associated with estimation of c, and the latter of γ. This provides another argument for the asymptotic independence.

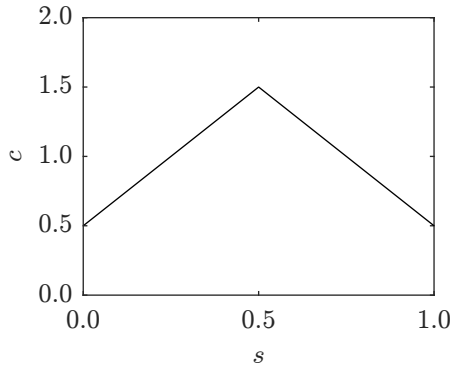
A.3 Simulation scedasis function plots



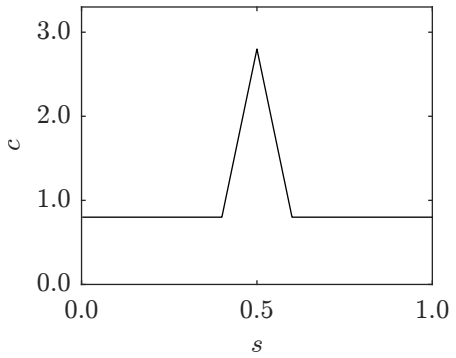
DGP 1.



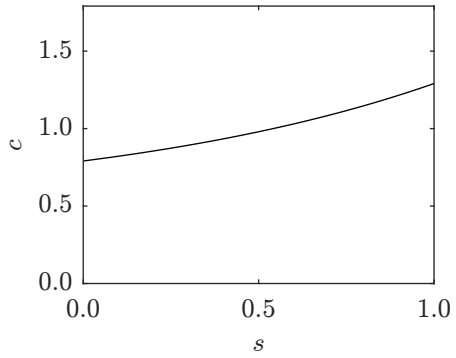
DGP 2.



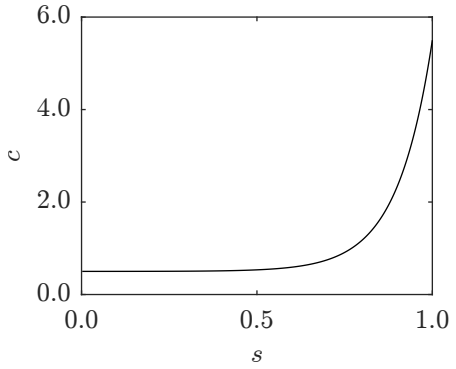
DGP 3.



DGP 4.



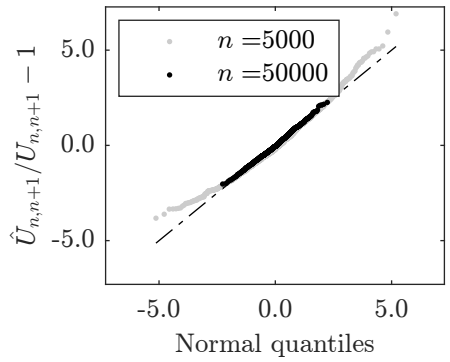
DGP 5.



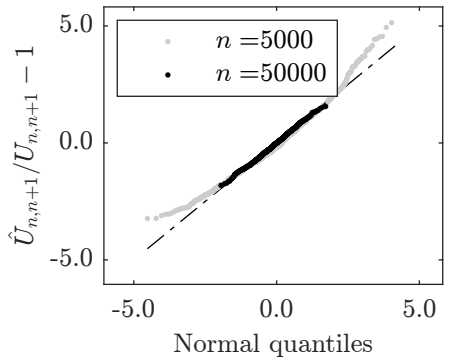
DGP 6.

Figure A.1: Scedasis functions corresponding to each of the six DGPs in the simulation study in Section 6 .

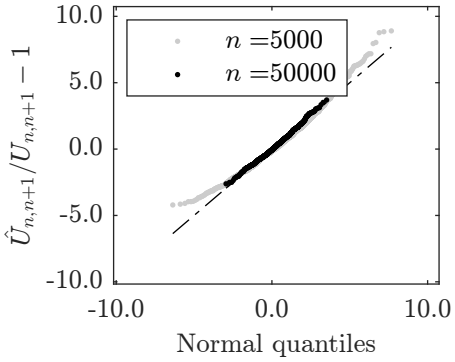
A.4 Simulation quantile-quantile plots



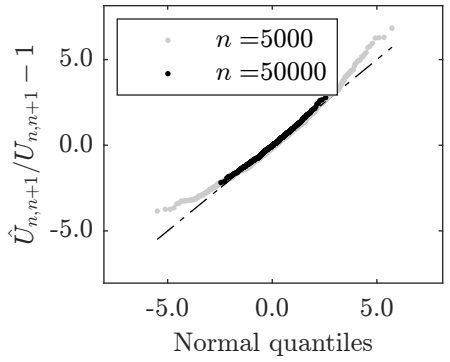
DGP 1, p-value = 0.054.



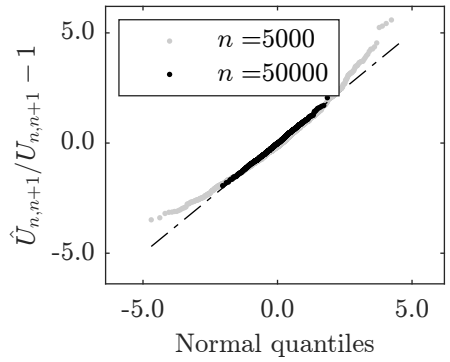
DGP 2, p-value = 0.397.



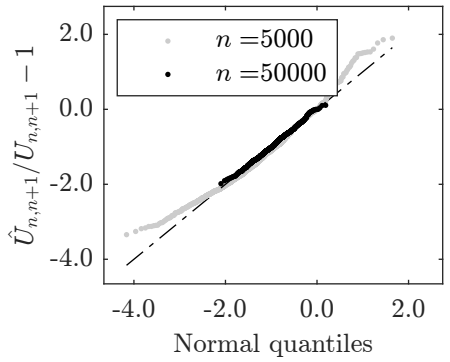
DGP 3, p-value = 0.007.



DGP 4, p-value = 0.014.



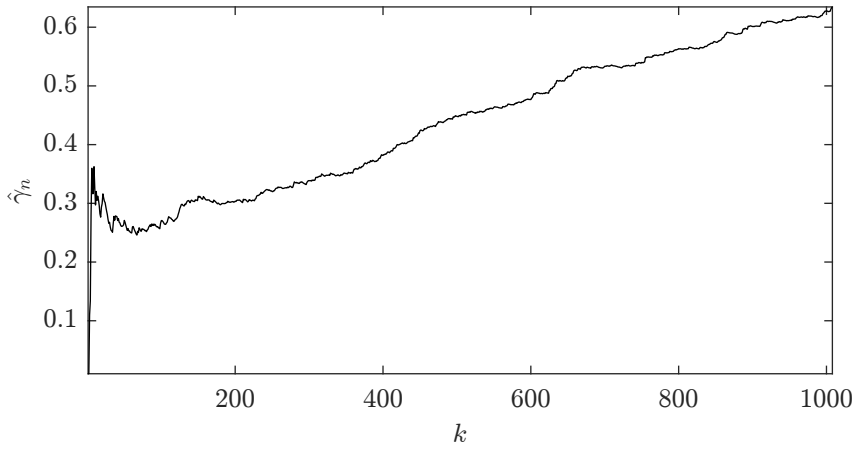
DGP 5, p-value = 0.359.



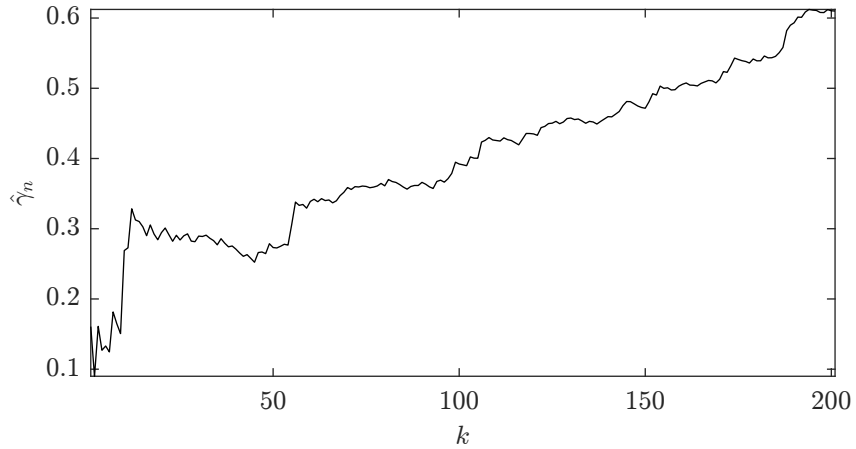
DGP 6, p-value = 0.358.

Figure A.2: Quantile-quantile plots for the one-step ahead predictions against normal quantiles for the fixed, non-optimised bandwidths corresponding to each of the six DGPs in the simulation study in Section 6. The p-values of the Jarque-Bera test for normality for $n = 50000$ are given below each figure. Those for $n = 5000$ are all virtually zero.

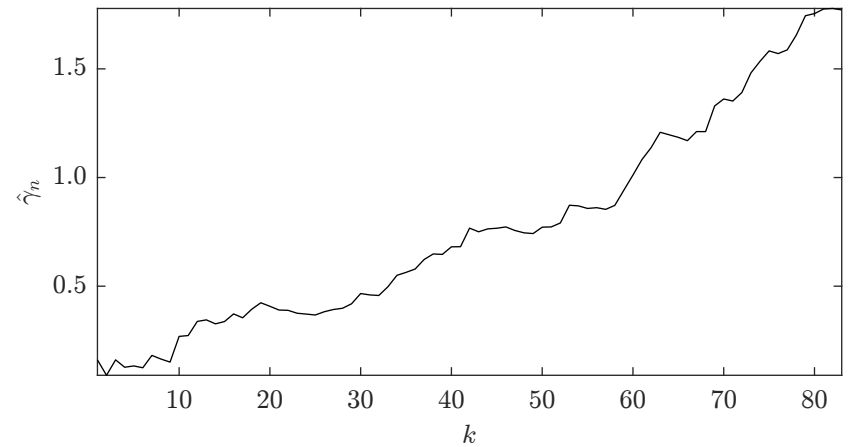
A.5 Empirical application Hill plots



(a) 20-Years period 1988-2007.



(b) 4-Years period 2004-2007.



(c) 1-Year period 2007.

Figure A.3: Hill plots for the S&P 500 log-losses over the 20-years period 1988-2007 (top), 4-years period 2004-2007 (middle), and 1-year period 2007 (bottom), for the empirical application in Section 7.

A.6 MATLAB programs and functions

Here we give an overview of the programs and functions implemented in MATLAB, version R2020b. The (main) programs are

- *simulation_replication.m*: results for the Monte Carlo simulation in Section 6, Table 6.2 and Figure A.2 for the fixed, non-optimised bandwidth parameters, and for Table 6.1 and Figure A.1 for the theoretical scedasis functions. This program calls the functions *c_dgp.m*, *BetaKSE.m*, and *BiweightKSE.m*.

- *simulation_optimisation.m*: results for the Monte Carlo simulation in Section 6, Table 6.3 for the optimised bandwidth parameters. This program calls the functions *c_dgp.m*, *BetaKSE.m*, and *BiweightKSE.m*.

- *application.m*: results for the empirical application in Section 7, Table 7.1 and Figure 7.1, 7.2, and A.5. This program calls the functions *BetaKSE.m* and *BiweightKSE.m*.

- *Kb_plots.m*: plots the beta kernel at the boundary for a small and a large bandwidth for Figure 5.1.

The functions are

- *c_dgp.m*: theoretical scedasis for each DGP.

- *BetaKSE.m*: beta kernel scedasis estimator.

- *BiweightKSE.m*: convolution-type biweight (boundary) kernel scedasis estimator.

- *CV_BiweightKSE.m*: cross-validation criterion corresponding to (A.2) and (A.3) for the convolution-type biweight kernel scedasis estimator, used in preliminary research by means of simulation and application.