Fitting tails affected by truncation

Jan Beirlant[*]

Department of Mathematics and LStat, KU Leuven

Department of Mathematical Statistics and Actuarial Science, University of the Free State

Isabel Fraga Alves

Department of Statistics and Operations Research, University of Lisbon

and

Tom Reynkens

Department of Mathematics and LStat, KU Leuven

Abstract: In several applications, ultimately at the largest data, truncation effects can be observed when analysing tail characteristics of statistical distributions. In some cases truncation effects are forecasted through physical models such as the Gutenberg-Richter relation in geophysics, while at other instances the nature of the measurement process itself may cause under recovery of large values, for instance due to flooding in river discharge readings. Recently, Beirlant, Fraga Alves and Gomes (2016) discussed tail fitting for truncated Pareto-type distributions. Using examples from earthquake analysis, hydrology and diamond valuation we demonstrate the need for a unified treatment of extreme value analysis for truncated heavy and light tails. We generalise the classical Peaks over Threshold approach for the different max-domains of attraction with shape parameter $\xi > -1/2$ to allow for truncation effects. We use a pseudo maximum likelihood approach to estimate the model parameters and consider extreme quantile estimation and reconstruction of quantile levels before truncation whenever appropriate. We report on some simulation experiments and provide some basic asymptotic results.

Keywords and phrases: Tail estimation, truncation, maximum likelihood estimation.
MSC 2010 subject classifications: 62G32.

1. Introduction

Modelling extreme events has recently received a lot of interest. Assessing the risk of rare events through estimation of extreme quantiles or corresponding return periods has been developed extensively and was applied to a wide variety of fields such as meteorology, finance, insurance and geology, among others.

---
[*]Corresponding author. Address: Celestijnenlaan 200B, 3001 Leuven, Belgium; e-mail: jan.beirlant@kuleuven.be

The methodology on modelling the univariate upper tail of the distribution of such quantities Y relies on the fact that the maximum of independent measurements Y_i, $i = 1, \ldots, n$, can be approximated by the generalised extreme value distribution: as $n \to \infty$

$$\mathbb{P}\left(\frac{\max\limits_{i=1,\ldots,n} Y_i - b_n}{a_n} \leq y\right) \to G_\xi(y) = \exp\left(-(1 + \xi y)^{-1/\xi}\right), \quad 1 + \xi y > 0, \quad (1)$$

where $b_n \in \mathbb{R}$, $a_n > 0$ and $\xi \in \mathbb{R}$ are the location, scale and shape parameters, respectively. For $\xi = 0$, $G_0(y)$ has to be read as $\exp\{-\exp(-y)\}$. In fact, (1) represents the only possible non-degenerate limits for maxima of independent and identically distributed sequences Y_i. Condition (1) is equivalent to the convergence of the distribution of excesses (or peaks) over high thresholds t to the generalised Pareto distribution (GPD): as t tends to the endpoint of the distribution of Y, then, with $\bar{F}$ the right tail function (RTF) of a given distribution,

$$\mathbb{P}\left(\frac{Y - t}{\sigma_Y(t)} > y \,\bigg|\, Y > t\right) = \frac{\bar{F}_Y(t + y\sigma_Y(t))}{\bar{F}_Y(t)} \to H_\xi(y) = -\log G_\xi(y) = (1 + \xi y)^{-1/\xi}, \quad (2)$$

where $\sigma_Y(t) > 0$. Below we set $\sigma_Y(t) = \sigma_t$. Setting t at the $(k+1)$th largest observation $y_{n-k,n}$ for some $k \in \{1 \ldots, n-1\}$ so that k data points are larger than the threshold t, (2) leads to the estimator

$$\hat{p}_c = \frac{k}{n} H_{\hat{\xi}}\left(\frac{c - y_{n-k,n}}{\hat{\sigma}}\right) \quad (3)$$

of the tail probability $\mathbb{P}(Y > c)$ for $c > 0$ large, where $(\hat{\xi}, \hat{\sigma})$ denote estimators for (ξ, σ_t). The modelling of extreme values and the estimation of tail parameters through the peaks over threshold (POT) methodology has been discussed for instance in Coles (2001), Embrechts, Klüppelberg and Mikosch (1997), Beirlant et al. (2004), and de Haan and Ferreira (2006).

Recently, Aban, Meerschaert and Panorska (2006); Chakrabarty and Samorodnitsky (2012); Beirlant, Fraga Alves and Gomes (2016) have addressed the problem of using unbounded probability mass leading to levels that are unreasonably large or physically impossible. All of these papers consider cases with shape parameter $\xi > 0$. In Beirlant, Fraga Alves and Gomes (2016) it was observed that the above mentioned extreme value methods, even when using a negative extreme value index, are not able to capture truncation at high levels. However, in several other fields, such as hydrology and earthquake magnitude modelling, the underlying distributions appear to be lighter tailed than Pareto. In this paper we will propose an adaptation of the classical approach to truncated tails over the whole range of max-convergence (1) with $\xi > -0.5$ as in the original POT approach.

First, we consider recent **magnitude data** (expressed on the Richter scale) of the 200 largest earthquakes in the Groningen area (the Netherlands), in the period 2003–2015, which are caused by gas extraction. In Figure 1, we present the time plot and the exponential QQ-plot $(x_{n-j+1,n}, \log(j/n))$ $(j = 1, 2, \ldots, n)$
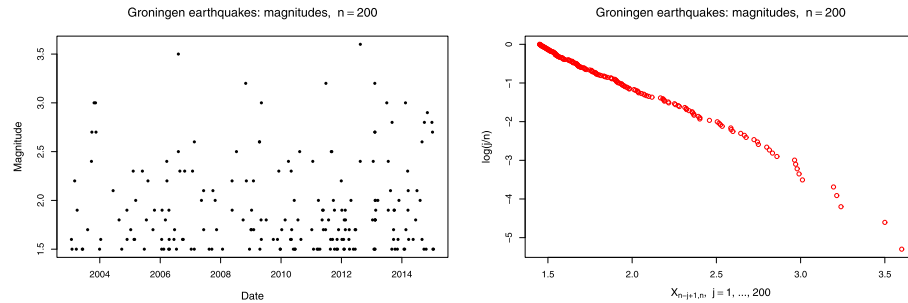
FIG 1. *Time plot and exponential QQ-plot of earthquake magnitude data from the Groningen area.*

where $x_{1,n} \leq \ldots \leq x_{n-j+1,n} \leq \ldots \leq x_{n,n}$ denote the ordered data. Along the Gutenberg-Richter (1956) law the magnitudes of independent earthquakes are drawn from a doubly truncated exponential distribution

$$\mathbb{P}(M > m) = \frac{e^{-\lambda m} - e^{-\lambda T_M}}{e^{-\lambda m_0} - e^{-\lambda T_M}}, \ m_0 < m < T_M.$$

Kijko and Singh (2011) provide a review of the vast literature on estimating the maximum possible magnitude T_M. The energy E released by earthquakes, expressed in megajoule (MJ), relates to the magnitude M, expressed on the Richter scale, by

$$M = \log_{10}(E/2)/1.5 + 1.$$

In Figure 1, a linear pattern is visible for a large section of the magnitudes data, while some curvature appears at the largest values. The data set was tested for serial correlation and no significance could be detected.

Secondly, we revisit the **diamond size data** considered in Verster et al. (2012). The nature of metallurgical recovery processes in diamond mining may cause under recovery of large diamonds between 30 and 60 cts per stone. If stones are not recovered during this process they are discarded onto tailing dumps from which they can be recovered during future re-mining programs. Because even a small number of large diamonds can have a large value, the question arises whether re-mining a mine dump can be made profitable by recovering these large diamonds. Therefore, the expected number of large diamonds above certain carat values c is of interest and the original non-truncated values are to be reconstructed from the data, which exhibit truncation. In Figure 2, the Pareto QQ-plot or log-log plot $(\log x_{n-j+1,n}, \log(j/n))$ $(j = 1, 2, \ldots, n)$ of the available carat data is presented. Again, a curvature near the top data is visible.

Thirdly, we study the river flows of the Molenbeek river at Erpe-Mere in Belgium (in $m^3/s, n = 426$) obtained between 1986 and 1996. The data are peaks over threshold values taken from a complete series of hourly flow measurements which was filtered in order to satisfy hydrological independence as discussed in Willems (2009). This river is prone to flooding at high flow levels and hence the
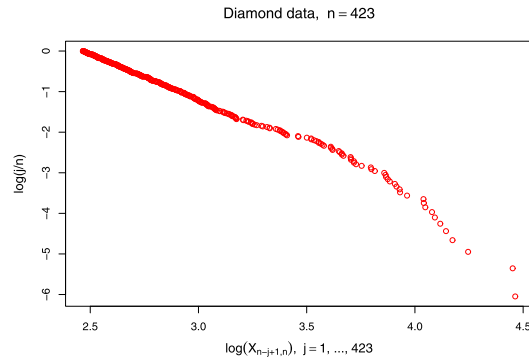
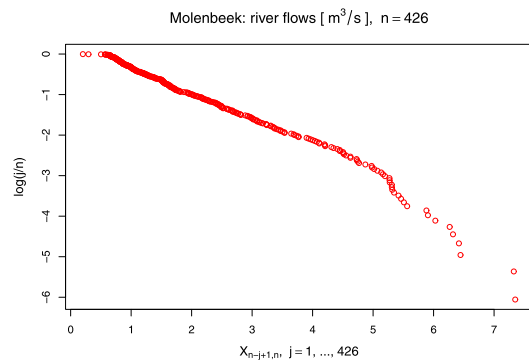Fig 2. *Log-log plot of diamond size data from Verster et al. (2012).*

Fig 3. *Exponential QQ-plot of the Molenbeek flow data.*

measurements can be truncated. In Figure 3 the exponential QQ-plot is given, which exhibits a linear (i.e. exponential) pattern with a downward curvature near the largest floods.

In this paper, we aim to provide a statistical model being able to approximate tail characteristics of distributions truncated at high levels. Moreover, the statistical estimation methods should also include the case of no-truncation in order for these methods to be useful and competitive both in cases with and without truncation. In the case of Pareto-type tails with $\xi > 0$ the proposed methods should also be compared with the methods which have been developed specifically for that sub-case. To this purpose we extend the classical POT technique with maximum likelihood estimation of the GPD parameters ξ and σ. Of course estimators for tail probabilities and extreme quantiles of a truncated distribution are to be discussed. Estimation of the endpoint T of a truncated distribution is of particular importance as discussed above in earthquake applications. Motivated by the river flow and diamond valuation examples, we finally consider the problem of reconstructing quantiles of the underlying unobserved variable Y before truncation.

2. Model

Let Y denote a parent random variable with distribution function $F_Y(y) = \mathbb{P}(Y \le y)$, RTF $\bar{F}_Y(y) = 1 - F_Y(y)$, quantile function $Q_Y(p) = \inf\{y : F_Y(y) \ge p\}$ $(0 < p < 1)$, and tail quantile function $U_Y(v) = Q_Y(1 - \frac{1}{v})$ $(v > 1)$. We consider the right truncated distribution from which independent and identically distributed data $X_1, X_2, \ldots, X_n$ are observed with, for some $T > 0$,

$$X =_d Y \,|\, Y < T. \tag{4}$$

The corresponding RTF is denoted by $\bar{F}_T(x) = \mathbb{P}(X > x)$ and the tail quantile function is given by $U_T(u) = Q_T(1 - \frac{1}{u})$ $(u > 1)$. Then,

$$\bar{F}_T(x) = \frac{\bar{F}_Y(x) - \bar{F}_Y(T)}{1 - \bar{F}_Y(T)} = (1 + D_T)\bar{F}_Y(x) - D_T, \tag{5}$$

$$U_T(u) = U_Y\left(\frac{u}{F_Y(T)}[1 + uD_T]^{-1}\right) \tag{6}$$

$$= U_Y\left(\frac{1}{\bar{F}_Y(T)}\left[1 + \frac{1}{uD_T}\right]^{-1}\right), \tag{7}$$

where $D_T = \bar{F}_Y(T)/F_Y(T)$ equals the odds of the truncated probability mass under the untruncated distribution Y.

The goal of this paper is to provide a test for truncation and to estimate

- the model parameters ξ and $\sigma = \sigma_t$,
- the odds D_T,
- quantiles $Q_T(1 - p)$ (p small) of the truncated distribution and the truncation point $T = Q_T(1)$,
- tail probabilities $\mathbb{P}(X > c)$ (c large) of the truncated distribution,
- and reconstruct quantile levels $Q_Y(1 - p)$ of the parent variable Y before truncation,

all on the basis of a pure random sample from X (possibly) truncated at some large T.

We assume that the distribution of Y satisfies (1) or, equivalently, (2). Condition (2) is also known to be equivalent to the following condition relating extreme quantile levels at $1 - \frac{1}{vy}$ and $1 - \frac{1}{y}$ close to the endpoint of the distribution: there exists a positive measurable function a such that

$$\frac{U_Y(vy) - U_Y(y)}{a(y)} \to \frac{v^\xi - 1}{\xi} \quad \text{when } y \to \infty, \tag{8}$$

with $a(1/\bar{F}_Y(t_{k,n})) = \sigma_t$ where $t = t_{k,n} = U_T(n/k)$. The right hand side of (8) is to be read as $\log v$ for $\xi = 0$.

The specific case $\xi > 0$ of Pareto-type distributions satisfies

$$\frac{U_Y(vy)}{U_Y(y)} \to_{y \to \infty} v^\xi \quad \text{and} \quad \mathbb{P}(Y/t > y \,|\, Y > t) = \frac{\bar{F}_Y(ty)}{\bar{F}_Y(t)} \to_{t \to \infty} y^{-1/\xi}. \tag{9}$$

Also when $\xi > 0$, $\sigma_t \sim \xi t$ as $t \to \infty$. Furthermore, it is known that $\sigma_t/t \to 0$ when $\xi \leq 0$.

Note that for a given T fixed, the tail of a truncated model X defined through (4) has an extreme value index $\xi_X = -1$, see for instance Figure 2.8 in Beirlant et al. (2004).

Truncation of a distribution Y satisfying (2) at a value T necessarily requires $t < T \to \infty$. The threshold t is mostly taken at the theoretical quantile $Q_T(1-\frac{k}{n})$ $= U_T(n/k)$, which in practice is estimated by the empirical quantile $X_{n-k,n}$. Given the fact that our model is only defined choosing $t = t_n, T = T_n \to \infty$ as the sample size $n \to \infty$, the underlying model depends on n and a triangular array formulation $X_{n1}, \ldots, X_{nn}$ of the observations should be used in order to emphasise the nature of the model. However, in statistical procedures as presented here, when a single sample is given, the notation $X_1, \ldots, X_n$ is more natural and will be used throughout.

The model considered in this paper is then given by

($\mathcal{M}$) For a sequence $T_n \to \infty$, $\{X_{n1}, \ldots, X_{nn}\} = \{X_1, \ldots, X_n\}$ are independent copies of a random variable $X = X_{T_n}$ where $X = X_{T_n}$ is distributed as $Y|Y < T_n$, with Y satisfying (2) or equivalently (8).

Now we consider the distribution of the POT values for the data of the truncated distribution under ($\mathcal{M}$):

$$
\begin{aligned}
\mathbb{P}\left(\frac{X-t}{\sigma_t} > x \,\Big|\, X > t\right) &= \mathbb{P}\left(\frac{Y-t}{\sigma_t} > x \,\Big|\, t < Y < T\right) \\
&= \frac{\mathbb{P}(Y > t + x\sigma_t) - \mathbb{P}(Y > T)}{\mathbb{P}(Y > t) - \mathbb{P}(Y > T)} \\
&= \frac{\frac{\mathbb{P}(Y > t + x\sigma_t)}{\mathbb{P}(Y > t)} - \frac{\mathbb{P}(Y > T)}{\mathbb{P}(Y > t)}}{1 - \frac{\mathbb{P}(Y > T)}{\mathbb{P}(Y > t)}}.
\end{aligned}
\tag{10}
$$

One can now consider two cases as $t, T \to \infty$:

- ($\mathcal{T}_t$) *Rough truncation with the threshold $t = t_n$:*

$$
(T - t)/\sigma_t \to \kappa > 0,
\tag{11}
$$

and hence from (2) and with local uniform convergence in (2)

$$
\frac{\mathbb{P}(Y > T)}{\mathbb{P}(Y > t)} \to (1 + \xi\kappa)^{-1/\xi}.
\tag{12}
$$

This entails that for $x \in (0, \kappa)$

$$
\mathbb{P}\left(\frac{X-t}{\sigma_t} > x \,\Big|\, X > t\right) \to \frac{(1 + \xi x)^{-1/\xi} - (1 + \xi\kappa)^{-1/\xi}}{1 - (1 + \xi\kappa)^{-1/\xi}} =: \bar{F}_{\xi,\kappa}(x).
\tag{13}
$$

This corresponds to situations where the deviation from the Pareto behaviour due to truncation at a high value T will be visible in the data from t on, and the approximation of the POT distribution using the limit distribution in (13) appears more appropriate than with a simple GPD.

- $(\bar{\mathcal{T}}_t)$ *Light truncation with the threshold* $t = t_n$: $\frac{\mathbb{P}(Y>T)}{\mathbb{P}(Y>t)} \to 0$.
  This entails

$$\mathbb{P}\left(\frac{X-t}{\sigma_t} > x \,\middle|\, X > t\right) \to (1+\xi x)^{-1/\xi}, \quad 1+\xi x > 0. \qquad (14)$$

Light truncation is introduced for mathematical completeness. But $(\bar{\mathcal{T}}_t)$ means that the truncation is not really visible in the data above t, and the classical extreme value modelling without truncation is appropriate. Hence, it will be practically impossible to discriminate light truncation from no truncation (i.e. $T = \infty$).

Under $(\mathcal{T}_t)$ with $t = t_{k,n} = U_T(n/k)$ we find from applying F_Y to both sides of (6) with $u = n/k$ that

$$\bar{F}_Y(t) = F_Y(T)\frac{1+(n/k)D_T}{n/k} = F_Y(T)\left(\frac{k}{n} + D_T\right),$$

from which, dividing by $\bar{F}_Y(T)$, we obtain

$$\frac{\bar{F}_Y(t)}{\bar{F}_Y(T)} = \frac{1}{D_T}\left(\frac{k}{n} + D_T\right),$$

while, using (2) and $(\mathcal{T}_t)$,

$$\frac{\bar{F}_Y(T)}{\bar{F}_Y(t)} \to (1+\xi\kappa)^{-1/\xi},$$

and hence under $(\mathcal{T}_t)$

$$\frac{k}{nD_T} \to (1+\xi\kappa)^{1/\xi} - 1. \qquad (15)$$

Now in order to be able to construct extreme quantile estimators under $(\mathcal{T}_t)$, remark that from (8) with $vy = 1/p$, $y = 1/\bar{F}_Y(t)$ and $k_\xi(u) = (u^\xi - 1)/\xi$, we have as $t \to \infty$ and $\bar{F}_Y(t)/p \to C$ for some constant $C > 0$ that

$$\frac{Q_Y(1-p) - t}{\sigma_t} - k_\xi\left(\frac{\bar{F}_Y(t)}{p}\right) \to 0.$$

Hence, with (7) and $p = \bar{F}_Y(T)(1 + \frac{1}{uD_T})$ we obtain

$$
\begin{aligned}
\frac{U_T(u) - t}{\sigma_t} &= \frac{U_Y\left(\frac{1}{\bar{F}_Y(t)}[1 + \frac{1}{uD_T}]^{-1}\right) - t}{\sigma_t} \\
&= k_\xi\left(\frac{\bar{F}_Y(t)}{\bar{F}_Y(T)[1 + \frac{1}{uD_T}]}\right) + o(1).
\end{aligned}
$$

Using (15) and (2) with $y = \kappa$ we obtain under $(\mathcal{T}_t)$ that

$$\frac{\bar{F}_Y(t)}{\bar{F}_Y(T)} \sim (1+\xi\kappa)^{1/\xi} \sim 1 + \frac{k}{nD_T}.$$

Hence, we conclude that under $(\mathcal{T}_t)$ for $1/(uD_T) \to 0$

$$\frac{U_T(u) - t}{\sigma_t} - k_\xi \left(\frac{1 + \frac{k}{nD_T}}{1 + \frac{1}{uD_T}} \right) \to 0. \tag{16}$$

These derivations will motivate the proposed estimators of D_T and extreme quantiles $Q_T(1 - p)$.

3. Inference

3.1. Estimators and goodness-of-fit

Estimation of the parameters (ξ, σ) in the classical POT without truncation is well-developed (Coles, 2001; Beirlant et al., 2004). Fitting the scaled GPD with RTF $(1 + \frac{\xi}{\sigma}x)^{-1/\xi}$ to the excesses $X - t$ given $X > t$ (based on (13)) using maximum likelihood is by far the most popular method in this respect. Here we rely on the generalisation (13) under $(\mathcal{T}_t)$, with t replaced by a random threshold $X_{n-k,n}$ and using the exceedances $E_{j,k} = X_{n-j+1,n} - X_{n-k,n}$ $(j = 1, 2, \ldots, k)$ for some $k \geq 2$. Substituting $E_{1,k}/\sigma$ for κ following (11), the log-likelihood is given by

$$\begin{aligned}
\log L_{k,n}(\xi, \sigma) &= \log \left(\prod_{j=2}^{k} \frac{\sigma^{-1} \left(1 + \frac{\xi}{\sigma} E_{j,k}\right)^{-(1/\xi)-1}}{1 - \left(1 + \frac{\xi}{\sigma} E_{1,k}\right)^{-1/\xi}} \right) \\
&= -(k-1) \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{j=2}^{k} \log \left(1 + \frac{\xi}{\sigma} E_{j,k}\right) \\
&\quad -(k-1) \log \left(1 - \left(1 + \frac{\xi}{\sigma} E_{1,k}\right)^{-1/\xi}\right),
\end{aligned}$$

or, by reparametrising (ξ, σ) to (ξ, τ) with $\tau = \xi/\sigma$,

$$\begin{aligned}
\log L_{k,n}(\xi, \tau) &= (k-1) \log \tau - (k-1) \log \xi - \left(1 + \frac{1}{\xi}\right) \sum_{j=2}^{k} \log(1 + \tau E_{j,k}) \\
&\quad -(k-1) \log \left(1 - (1 + \tau E_{1,k})^{-1/\xi}\right).
\end{aligned}$$

The partial derivatives are given by

$$\begin{aligned}
\frac{1}{k-1} \frac{\partial \log L_{k,n}(\xi, \tau)}{\partial \xi} &= -\frac{1}{\xi} + \frac{1}{\xi^2} \frac{1}{k-1} \sum_{j=2}^{k} \log(1 + \tau E_{j,k}) \\
&\quad + \frac{1}{\xi^2} \frac{(1 + \tau E_{1,k})^{-1/\xi} \log(1 + \tau E_{1,k})}{1 - (1 + \tau E_{1,k})^{-1/\xi}},
\end{aligned}$$

$$\frac{1}{k-1}\frac{\partial \log L_{k,n}(\xi,\tau)}{\partial \tau} = \frac{1}{\tau} - \left(1+\frac{1}{\xi}\right)\frac{1}{k-1}\sum_{j=2}^{k}\frac{E_{j,k}}{1+\tau E_{j,k}}$$
$$- \frac{1}{\xi}E_{1,k}\frac{(1+\tau E_{1,k})^{-1-1/\xi}}{1-(1+\tau E_{1,k})^{-1/\xi}},$$

from which the likelihood equations defining the pseudo maximum likelihood estimators $(\hat{\xi}_k,\hat{\tau}_k)$ are obtained:

$$\frac{1}{k-1}\sum_{j=2}^{k}\log(1+\hat{\tau}_k E_{j,k}) + \frac{(1+\hat{\tau}_k E_{1,k})^{-1/\hat{\xi}_k}\log(1+\hat{\tau}_k E_{1,k})}{1-(1+\hat{\tau}_k E_{1,k})^{-1/\hat{\xi}_k}} = \hat{\xi}_k \qquad (17)$$

$$\frac{1}{k-1}\sum_{j=2}^{k}\frac{1}{1+\hat{\tau}_k E_{j,k}} = \frac{1}{1+\hat{\xi}_k}\frac{1-(1+\hat{\tau}_k E_{1,k})^{-1-1/\hat{\xi}_k}}{1-(1+\hat{\tau}_k E_{1,k})^{-1/\hat{\xi}_k}}. \qquad (18)$$

When computing $(\hat{\xi}_k,\hat{\tau}_k)$, one has to impose the model restrictions. In order to meet the restrictions $\sigma = \xi/\tau > 0$ and $1+\tau E_{j,k} > 0$ for $j = 1,\ldots,k$, in our implementation we require the estimates of these quantities to be larger than the numerical tolerance value 10^{-10}.

An estimator of D_T now follows from taking $u = n$ in (16):

$$U_T(n) - U_T(n/k) \approx \sigma k_\xi \left(\frac{1+\frac{k}{nD_T}}{1+\frac{1}{nD_T}}\right).$$

Estimating $U_T(n) - U_T(n/k)$ by $E_{1,k}$ we obtain

$$\hat{D}_{T,k} := \max\left\{0, \frac{k}{n}\frac{(1+\hat{\tau}_k E_{1,k})^{-1/\hat{\xi}_k} - \frac{1}{k}}{1-(1+\hat{\tau}_k E_{1,k})^{-1/\hat{\xi}_k}}\right\}. \qquad (19)$$

Similarly taking $u = 1/p$ in (16) with $np/k \to 0$, we obtain estimators for $Q_T(1-p)$:

$$\hat{Q}_{T,k}(1-p) = X_{n-k,n} + \frac{1}{\hat{\tau}_k}\left[\left\{\frac{\hat{D}_{T,k}+\frac{k}{n}}{\hat{D}_{T,k}+p}\right\}^{\hat{\xi}_k} - 1\right]. \qquad (20)$$

Setting $p = 0$ in (20) one obtains an estimator for the truncation point T:

$$\hat{T}_k = X_{n-k,n} + \frac{1}{\hat{\tau}_k}\left[\left\{\frac{1-k^{-1}}{(1+\hat{\tau}_k E_{1,k})^{-1/\hat{\xi}_k} - k^{-1}}\right\}^{\hat{\xi}_k} - 1\right]. \qquad (21)$$

Based on (3) and (5) an estimator for tail probabilities $\mathbb{P}(X > c)$ can be derived:

$$\hat{p}_{T,k}(c) = (1+\hat{D}_{T,k})\frac{k}{n}(1+\hat{\tau}_k(c - X_{n-k,n}))^{-1/\hat{\xi}_k} - \hat{D}_{T,k}. \qquad (22)$$

Note that all proposed estimators from (17), (18), (20) and (22) are direct generalisations of the classical POT estimators under no-truncation which are obtained by setting $\hat{D}_{T,k}$ equal to 0.

From (6) it follows that when $p - (1-p)D_T > 0$, or $p > D_T/(1+D_T) = \bar{F}_Y(T)$

$$Q_Y(1-p) = Q_T((1-p)(1+D_T)) = Q_T(1 - \{p - (1-p)D_T\}),$$

from which the following estimator reconstructing $Q_Y(1-p)$ of the parent distribution Y follows:

$$
\begin{aligned}
\hat{Q}_{Y,k}(1-p) &= \hat{Q}_{T,k}\left(1 - [p - (1-p)\hat{D}_{T,k}]\right) \\
&= X_{n-k,n} + \frac{1}{\hat{\tau}_k}\left[\left\{\frac{\hat{D}_{T,k} + \frac{k}{n}}{p(\hat{D}_{T,k}+1)}\right\}^{\hat{\xi}_k} - 1\right].
\end{aligned}
\tag{23}
$$

In the specific case $\xi > 0$ the estimators developed above can be compared with those developed in Beirlant, Fraga Alves and Gomes (2016) for this special Pareto-type case:

$$
\begin{aligned}
H_{k,n} &= \hat{\xi}_k^+ + \frac{R_{k,n}^{1/\hat{\xi}_k^+}\log R_{k,n}}{1 - R_{k,n}^{1/\hat{\xi}_k^+}}, \\
\hat{D}_{T,k}^+ &= \max\left\{0, \frac{k}{n}\frac{R_{k,n}^{1/\hat{\xi}_k^+} - \frac{1}{k}}{1 - R_{k,n}^{1/\hat{\xi}_k^+}}\right\}, \\
\log\hat{Q}_{T,k}^+(1-p) &= \log X_{n-k,n} + \hat{\xi}_k^+\log\left(\frac{\hat{D}_{T,k}^+ + \frac{k}{n}}{\hat{D}_{T,k}^+ + p}\right),
\end{aligned}
$$

with $H_{k,n} = \frac{1}{k}\sum_{j=1}^{k}\log X_{n-j+1,n} - \log X_{n-k,n}$ the Hill (1975) statistic, and $R_{k,n} = X_{n-k,n}/X_{n,n}$.

Of course, in practice there is a clear need for detecting rough truncation. Let $(\bar{\mathcal{T}}_k)$ and $(\mathcal{T}_k)$ denote light and rough truncation with the thresholds $X_{n-k,n}$. A test for

$$H_{0,k} : (\bar{\mathcal{T}}_k) \text{ versus } H_{1,k} : (\mathcal{T}_k)$$

can be constructed generalising the goodness-of-fit test which was proposed by Aban, Meerschaert and Panorska (2006) within a Pareto context, rejecting $H_{0,k}$ at asymptotic level $q \in (0,1)$ when

$$T_{k,n} := k\left(1 + \hat{\tau}_k E_{1,k}\right)^{-1/\hat{\xi}_k} > \log(1/q), \tag{24}$$

while the P-value is given by $e^{-T_{k,n}}$, as under $H_{0,k}$, $T_{k,n}$ approximately follows a standard exponential distribution as will be shown in Theorem 3 below.

3.2. Simulation study

The authors have performed an extensive simulation study concerning all the proposed estimators for different distributions of Y. We compare the results with the results from a Pareto analysis $\hat{\xi}_k^+$ and $\hat{Q}_{T,k}^+(1-p)$ (Aban, Meerschaert and Panorska, 2006; Beirlant, Fraga Alves and Gomes, 2016), with the classical POT maximum likelihood results denoted by $\hat{\xi}_k^\infty$, $\hat{Q}_k^\infty(1-p)$, and with the classical moment estimators (Dekkers, Einmahl and de Haan, 1989)

$$
\hat{\xi}_k^M = M_k^{(1)} + 1 - \frac{1}{2}\left[1 - \frac{\left(M_k^{(1)}\right)^2}{M_k^{(2)}}\right]^{-1},
$$

$$
\hat{Q}_k^M(1-p) = X_{n-k,n} + X_{n-k,n}M_k^{(1)}\left(1 - \hat{\xi}_k^M\right)\frac{\left(\frac{k}{np}\right)^{\hat{\xi}_k^M} - 1}{\hat{\xi}_k^M},
$$

with $M_k^{(j)} := \frac{1}{k}\sum_{l=1}^k \log^j\left(X_{n-l+1,n}/X_{n-k,n}\right)$, $j = 1, 2$. In the Appendix we give a selection from these simulation results for Y following the standard Pareto distribution, the standard lognormal distribution, the standard exponential distribution, and the GPD with RTF $H_{-0.2}$. For each setting, 1000 samples for X of size 500 were generated where we consider different levels of truncation: $T = Q_Y(0.975)$, $T = Q_Y(0.99)$ and $T = Q_Y(1)$. Note that the last case corresponds to no truncation, or $X =_d Y$. The samples were generated using inverse transform sampling with the quantile function $Q_T(p) = Q_Y(pF_Y(T))$ (which can easily be deduced from (5)).

To show the performance of the test for truncation, we plot the average P-values over the 1000 simulations as a function of k in the first columns of Figures 7–10 (full line). Additionally, the median (dashed line), first quartile (dotted line) and third quartile (dotted line) of the P-values over the 1000 simulations are also plotted as a function of k. This corresponds to the box of the boxplot of P-values as a function of k. Finally, we add blue horizontal lines (dash-dotted line) indicating the standard significance levels of 1% and 5%. When truncation is present ($T = Q_Y(0.975)$ or $T = Q_Y(0.99)$), the average P-values show that the test rejects the null hypothesis of no truncation when k is large enough. For the standard exponential, standard lognormal and GPD(-0.2,1) truncated at $T = Q_Y(0.99)$, the average P-value is higher than, or just below, the 5% significance level, even for high values of k. However, when looking at the median values and the third quartile, we see that the majority, and sometimes more than 75%, of the P-values are below the 5% significance level. When the data are not truncated, i.e. $X =_d Y$, the P-values are on average always well above the considered significance levels, hence correctly not rejecting the null hypothesis. The first quartile of the P-values is also above the 5% significance level, except for smaller values of k. Note that when we look at $Y \sim \text{GPD}(-0.2, 1)$, Y itself is bounded by $-\sigma/\xi = 5$, but still $X =_d Y$ when we set $T = Q_Y(1)$. The simulation results

show that the test performs as expected: rejecting the null hypothesis when $T = Q_Y(0.975)$ or $T = Q_Y(0.99)$, and not rejecting the null hypothesis when $T = Q_Y(1)$.

Concerning the estimation of ξ, see the second and third columns in Figures 7–10, the behaviour of $\hat{\xi}_k$ in the standard Pareto case exhibits a slightly smaller bias but quite a larger variance compared to $\hat{\xi}^+_{T,k}$ from Aban, Meerschaert and Panorska (2006); Beirlant, Fraga Alves and Gomes (2016) which was constructed exclusively for the case $\xi > 0$. The classical POT and moment estimators exhibit large bias under truncation, as they tend to -1 when the threshold tends to $x_{n,n}$. The mean squared error of $\hat{\xi}_k$ is comparable to the mean squared error (MSE) of these estimators for $k \geq 200$. In case of no truncation the bias of $\hat{\xi}_k$ is the smallest for $k \geq 100$ while the mean squared error is the worst of the four estimators.

When $\xi \leq 0$, the estimator $\hat{\xi}^+_{T,k}$ from the Pareto analysis is breaking down as can be expected whereas the difference between the classical estimators and the newly proposed POT estimator is small for $k \geq 200$ in case $\xi = 0$ and $k \geq 300$ in the case $\xi < 0$. In all presented cases $\hat{\xi}_k$ compares well for k sufficiently large with the classical estimators when there is no truncation. Note that all estimators have a large bias for the (truncated) lognormal distribution. As can clearly be seen, the bias of all estimators decreases as truncation becomes lighter, or when there is no truncation, as expected. Moreover, the stable area of the $\hat{\xi}_k$ estimates starts for smaller values of k when the truncation point gets larger.

Concerning the estimation of $Q_T(1-p)$, see Figures 11–18 with $p = 0.01$ and 0.005 and $T = Q_Y(0.975), Q_Y(0.99)$, the estimator $\hat{Q}_{T,k}(1-p)$ has the smallest bias, uniformly over all distributions and values of p considered, while the MSE values are always comparable with the best performing estimators. Even in case of no truncation $\hat{Q}_{T,k}(1-p)$ does not lose too much accuracy in comparison with the classical MLE estimator.

3.3. Asymptotic results

Here we present the asymptotic normality of $(\hat{\xi}, \hat{\tau})$ and $\hat{Q}_{T,k}(1-p)$ under rough truncation, and the asymptotic null distribution of the goodness-of-fit test statistic $T_{k,n}$. The proofs are provided in the Appendix.

We assume a second-order remainder relation in (8) as in Theorem 3.4.2 in de Haan and Ferreira (2006): with $\xi > -\frac{1}{2}$,

$$\lim_{t\to\infty} \frac{\frac{U_Y(tx)-U_Y(t)}{a_Y(t)} - \frac{x^\xi-1}{\xi}}{A(t)} = \Psi_{\xi,\rho}(x) \text{ for all } x > 0, \qquad (25)$$

where

$$\Psi_{\xi,\rho}(x) = \int_1^x s^{\xi-1} \int_1^s u^{\rho-1} \, du \, ds,$$

with $\rho \leq 0$. Furthermore, we introduce the notations $b_{T,k,n} := \frac{k+1}{(n+1)D_T}$, $a_{T,k,n} := a_Y\left(1/(\bar{F}_Y(T)(1 + b_{T,k,n}))\right)$, and we denote the limit of $k/(nD_T)$ under rough truncation as derived in (15) by $\beta := (1 + \xi\kappa)^{1/\xi} - 1$.

Theorem 1. *Let $X_1, X_2, \ldots,$ be i.i.d. random variables with distribution function F_T following* (5) *where U_Y satisfies* (25). *Let $n, k = k_n \to \infty$, $\frac{k}{n} \to 0$, $T \to \infty$. Then, under $(\mathcal{T}_t)$ we have that as $\sqrt{k}A(1/[\bar{F}_Y(T)(1 + b_{T,k,n})]) \to \lambda \in \mathbb{R}$*

$$\sqrt{k}\left(\hat{\xi}_k - \xi, \hat{\tau}_k a_{T,k,n} - \xi\right)' = \mathcal{I}_\beta^{-1}\mathbf{N}_{\xi,\beta} + \lambda\mathcal{I}_\beta^{-1}\mathbf{f}_{\xi,\beta,\rho} + o_p(1)\mathbf{1},$$

where

$$\mathcal{I}_\beta = \begin{pmatrix} 1 - \frac{1+\beta}{\beta^2}\log^2(1+\beta) & \frac{1}{\xi}\left[-\frac{\xi}{1+\xi}\frac{1+\beta}{\beta}(1-(1+\beta)^{-1-\xi}) \right. \\ & \left. +\frac{1+\beta}{\beta^2}\log(1+\beta)(1-(1+\beta)^{-\xi})\right] \\ \\ -\frac{1}{\xi}\left[-\frac{\xi}{1+\xi}\frac{1+\beta}{\beta}(1-(1+\beta)^{-1-\xi}) \right. & -\frac{1}{\xi\beta}\left[\frac{\xi}{1+2\xi}(1+\beta)(1-(1+\beta)^{-1-2\xi}) \right. \\ \left. +\frac{1+\beta}{\beta^2}\log(1+\beta)(1-(1+\beta)^{-\xi})\right] & \left. -\frac{1+\beta}{\beta}\frac{1}{\xi}(1-(1+\beta)^{-\xi})^2\right] \end{pmatrix},$$

$$\mathbf{N}_{\xi,\beta} = \frac{\beta}{1+\beta}\begin{pmatrix} \xi\int_0^1 W_n(u)\left(\frac{1+u\beta}{1+\beta}\right)^{-1}du \\ -\xi W_n(1)\left(-\frac{(1+\beta)^{1-\xi}\log(1+\beta)}{\beta^2} + \frac{\xi(1+\beta)^{-\xi}+(1+\beta)}{(1+\xi)\beta}\right) \\ \\ \xi(1+\xi)\int_0^1 W_n(u)\left(\frac{1+u\beta}{1+\beta}\right)^{-1+\xi}du \\ -W_n(1)\left(\frac{\xi(1+\xi)(1+\beta)}{(1+2\xi)\beta}(1-(1+\beta)^{-1-2\xi})\right. \\ \left. -\frac{(1+\beta)^{1-\xi}}{\beta^2}(1-(1+\beta)^{-\xi})\right) \end{pmatrix},$$

and

$$\mathbf{f}_{\xi,\beta,\rho} = \begin{pmatrix} \xi\int_0^1 \Psi_{\xi,\rho}(\frac{1+\beta}{1+u\beta})\left(\frac{1+u\beta}{1+\beta}\right)^\xi du \\ -\xi\Psi_{\xi,\rho}(1+\beta)(1+\beta)^{-\xi}\left(\frac{(1+\beta)\log(1+\beta)}{\beta^2} - \frac{1}{\beta}\right) \\ \\ \xi(1+\xi)\int_0^1 \Psi_{\xi,\rho}(\frac{1+\beta}{1+u\beta})\left(\frac{1+u\beta}{1+\beta}\right)^{2\xi}du \\ -\Psi_{\xi,\rho}(1+\beta)\frac{(1+\beta)^{1-\xi}}{\beta^2}(1-(1+\beta)^{-\xi}) \end{pmatrix},$$

for a sequence of Brownian motions $\{W_n(s); s \geq 0\}$.

Under $(\bar{\mathcal{T}}_t)$ the asymptotic result for $(\hat{\xi}_k, \hat{\tau}_k)$ can be checked to be identical to that of the classical MLE estimators under no truncation as given in Theorem 3.4.2 in de Haan and Ferreira (2006).

Note that the information matrix $\mathcal{I}_\beta$ equals 0 when $\kappa = 0$, or equivalently $\beta = 0$, so that the asymptotic variances are unbounded in such case. In practice this induces large variances for smaller values of k. This also appears in Figures 7–10. Fortunately, the bias stays reasonably small for larger values of k, as can be deduced for instance in case of the lognormal distribution.

In order to state the asymptotic result for the quantile estimator $\hat{Q}_{T,k}(1-p)$ with $p = p_n \to 0$, we use the notation $d_n = k/(np_n)$. Furthermore, we will use the result that when U_Y satisfies (25), we have that

$$\lim_{t\to\infty}\frac{\frac{a_Y(tx)}{a_Y(t)} - x^\xi}{A(t)} = Cx^\xi\frac{x^\rho - 1}{\rho} \tag{26}$$

for some constant C (see B.3.4 in de Haan and Ferreira (2006)).

Theorem 2. *Let $X_1, X_2, \ldots,$ be i.i.d. random variables with distribution function F_T following (5) where U_Y satisfies (25). Let $n, k = k_n \to \infty$, $\frac{k}{n} \to 0$, $T \to \infty$, $p = p_n \to 0$ and $n p_n / \sqrt{k} \to 0$. Then, under $(\mathcal{T}_t)$ we have that*

$$
\frac{\left(\hat{Q}_{T,k}(1-p) - Q_T(1-p) \right)}{a_Y \left(\frac{1}{\overline{F}_Y(T)} \right)}
$$

$$
= -\frac{\beta}{k}(E - 1) + O_p \left(\frac{1}{k^2} \vee \frac{1}{d_n^2} \right)
$$

$$
- \beta \left(\frac{1}{d_n} - \frac{1}{k} \right) \left[A \left(\frac{1}{\overline{F}_Y(T)} \right) C \frac{(1+\beta)^{-\rho} - 1}{\rho} \right.
$$

$$
+ \left(\frac{\hat{\xi}_k}{\hat{\tau}_k} \frac{1}{a_{T,k,n}} - 1 \right)
$$

$$
- \left(\hat{\xi}_k - \xi \right) \frac{1}{\xi} \frac{(1+\beta) \log(1+\beta)}{\beta}
$$

$$
+ (\hat{\tau}_k a_{T,k,n} - \xi) \frac{1 - (1+\beta)^{-\xi}}{\xi} \left(1 + \frac{1+\beta}{\xi\beta} \right)
$$

$$
+ (1+\beta)^{-\xi} \left(\frac{1+\beta}{\beta} + \xi \right)
$$

$$
\left. \times \left(-\frac{W_n(1)}{\sqrt{k}} + A \left(\frac{1}{\overline{F}_Y(T)} \right) (1+\beta)^{-\xi} \Psi_{\xi,\rho}(1+\beta) \right) \right],
$$

where E is a standard exponential random variable and $\{W_n(s); s \geq 0\}$ a sequence of Brownian motions.

This result should be compared to Theorem 4.3.1 in de Haan and Ferreira (2006) stating the basic asymptotic result for the quantile estimator based on the classical ML estimators under no truncation. Note that under $(\mathcal{T}_t)$ the rate of the stochastic part in the asymptotic representation is $O_p(1/k)$ rather than the classical $O_p(1/\sqrt{k})$.

Theorem 3. *Let $X_1, X_2, \ldots,$ be i.i.d. random variables with distribution function F_T following (5) where U_Y satisfies (25). Let $n, k = k_n \to \infty$, $\frac{k}{n} \to 0$, $T \to \infty$. Then, under $(\overline{\mathcal{T}}_k)$ with $n D_T \to 0$ we have that*

$$
T_{k,n} =_d E(1 + o_p(1))
$$

where E is a standard exponential random variable.

4. Case studies

Analysing the magnitude data from the Groningen area, it appears that the given 200 top data confirm the Gutenberg-Richter law with $\hat{\xi}_k$ clearly indicating that $Y = M$ belongs to the Gumbel $\xi = 0$ domain. The goodness-of-fit test
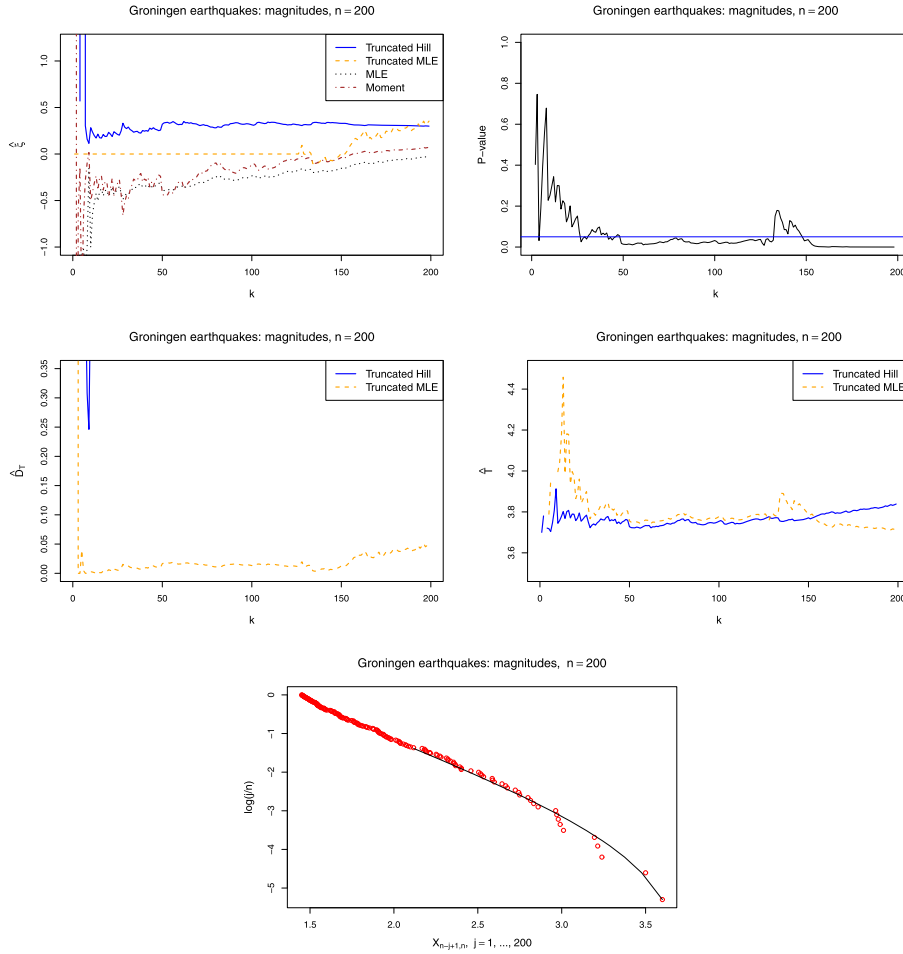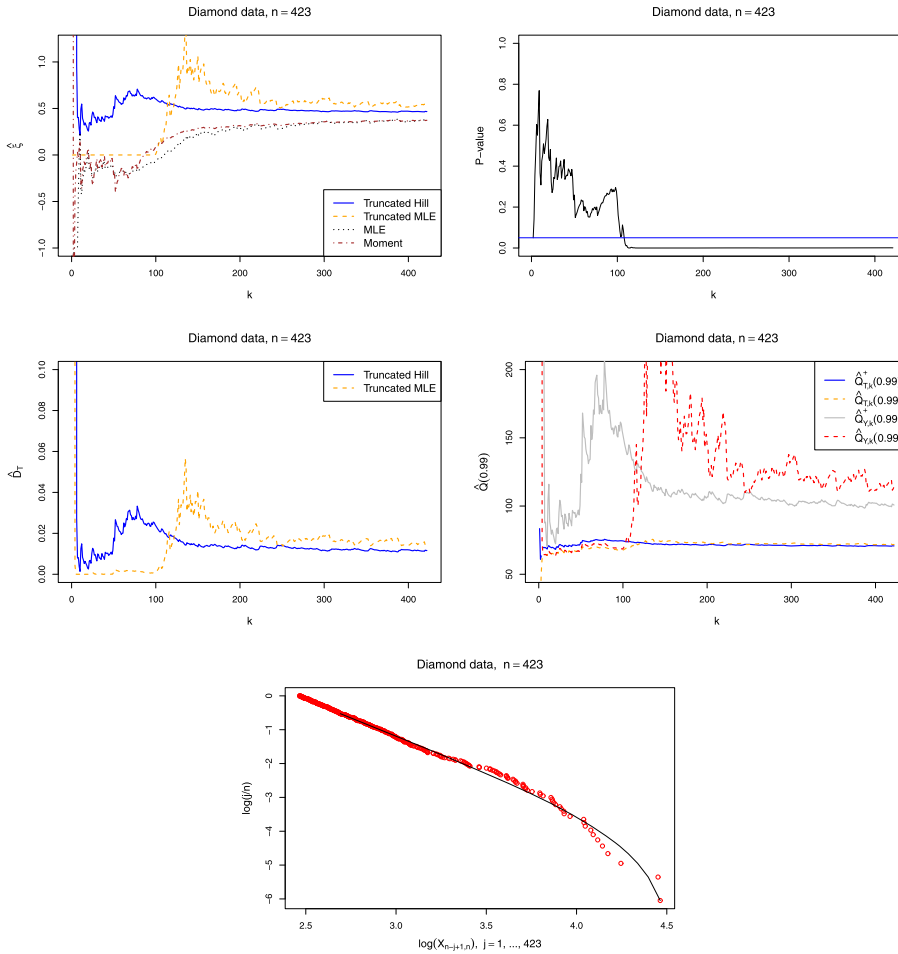
FIG 4. *Groningen earthquake magnitude data:* $\hat{\xi}_k^+$, $\hat{\xi}_k$, $\hat{\xi}_k^\infty$ *and* $\hat{\xi}_k^M$ *(top left); P-values for test for truncation (top right);* $\hat{D}_{T,k}^+$ *and* $\hat{D}_{T,k}$ *(middle left); the logarithmic transform of* $\hat{T}_{E,k}^+$ *and* $\hat{T}_{M,k}$ *(middle right); exponential QQ-plot with fit based on* $k = 50$ *largest magnitudes (bottom).*

rejects light truncation for $k \geq 40$ and the proposed truncation model fits well to the top 50 data as indicated on the exponential QQ-plot in Figure 4. Furthermore, $\hat{D}_{T,k}$ indicates a truncation volume D_T between 0.01 and 0.02. Finally, the endpoint can be estimated in two ways: directly on the magnitude data using $\hat{T}_{M,k} = \hat{Q}_{T,k}(1)$, or using a Pareto analysis $\hat{T}_{E,k}^+ = \hat{Q}_{T,k}^+(1)$ based on the energy data and transforming back to the magnitude scale with a logarithmic transformation. Both approaches lead to a value around 3.75.

Concerning the diamond data introduced in Figure 2, $\hat{\xi}_k$ and $\hat{\xi}_k^+$, respectively $\hat{D}_{T,k}$ and $\hat{D}_{T,k}^+$, correspond well for $k \geq 250$ and lead to a Pareto fit with extreme
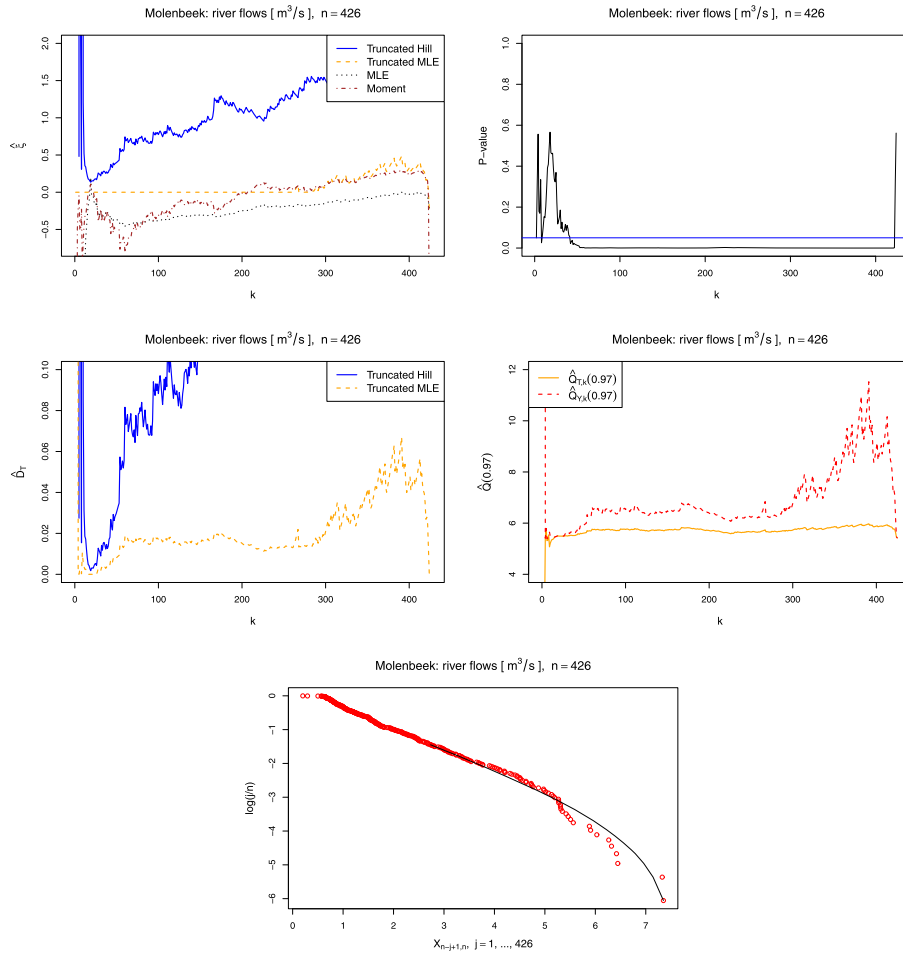
FIG 5. *Diamond data:* $\hat{\xi}_k^+$, $\hat{\xi}_k$, $\hat{\xi}_k^\infty$ *and* $\hat{\xi}_k^M$ *(top left); P-values for test for truncation (top right);* $\hat{D}_{T,k}^+$ *and* $\hat{D}_{T,k}$ *(middle left);* $\hat{Q}_{T,k}^+(0.99)$, $\hat{Q}_{T,k}(0.99)$, $\hat{Q}_{Y,k}^+(0.99)$ *and* $\hat{Q}_{Y,k}(0.99)$ *(middle right); log-log plot with fit based on $k = 250$ largest sizes (bottom).*

value index around 0.5 and a truncation odds D_T around 0.02. The goodness-of-fit test now rejects light truncation for $k \geq 110$. Reconstructing $Q_Y(0.99)$ with $\hat{Q}_{Y,k}(0.99)$ and $\hat{Q}_{Y,k}^+(0.99)$ leads to a value of 120 cts at $k = 250$.

Finally with the Molenbeek data, the goodness-of-fit test and the fit of the proposed truncation model on the exponential QQ-plot on the top 100 data, again indicate that this Y belongs to the Gumbel domain with an odds D_T around 0.02. Here, the Pareto domain estimators $\hat{\xi}_k^+$ and $\hat{D}_{T,k}^+$ clearly do not show a stable pattern as a function of k. Reconstructing $Q_Y(0.97)$ leads to a value $\hat{Q}_{Y,100}(0.97) = 6.5 \ m^3/s$.

FIG 6. *Molenbeek flow data: $\hat{\xi}_k^+$, $\hat{\xi}_k$, $\hat{\xi}_k^\infty$ and $\hat{\xi}_k^M$ (top left); P-values for test for truncation (top right); $\hat{D}_{T,k}^+$ and $\hat{D}_{T,k}$ (middle left); $\hat{Q}_{T,k}(0.97)$ and $\hat{Q}_{Y,k}(0.97)$ (middle right); exponential QQ-plot with fit based on $k = 100$ largest flows (bottom).*

5. Discussion

We proposed a general tail estimation approach for cases where truncation affects the ultimate right tail of the distribution. Using applications from geophysics, hydrology and geology we motivated the importance of this problem. The proposed estimators of the extreme value index, and quantiles of the truncated and underlying non-truncated distribution, in most cases compare well with the best performing alternatives, even in case there is no truncation. The proposed estimator of extreme quantiles of a truncated distribution is performing uniformly best. While the alternative procedures sometimes break down in

at least one situation, our proposals remain always useful for large enough k. Hence, in addition to the existing methods, this method can be an interesting extra tool when analysing tails.

Appendix: Proofs of Theorems

Proposition 1. *Under the condition of Theorem 1, one can define a sequence of Brownian motions $\{W_n(s) \,|\, s > 0\}$, such that for $\epsilon > 0$*

(a)

$$
\max_{j=1,\ldots,k} \left(\frac{j}{k+1} \right)^{0.5+\varepsilon} \left| \sqrt{k} \left[\frac{X_{n-j+1,n} - U_T\left(\frac{n+1}{k+1}\right)}{a_{T,k,n}} - \frac{1}{\xi}\left(\left(\frac{1 + \frac{j}{k+1}b_{T,k,n}}{1 + b_{T,k,n}} \right)^{-\xi} - 1 \right) \right] \right.
$$
$$
+ \frac{b_{T,k,n}}{1 + b_{T,k,n}} \left(\frac{1 + \frac{j}{k+1}b_{T,k,n}}{1 + b_{T,k,n}} \right)^{-1-\xi} W_n\left(\frac{j}{k+1} \right)
$$
$$
\left. + \sqrt{k} A\left(\frac{1}{\bar{F}_Y(T)(1 + b_{T,k,n})} \right) \Psi_{\xi,\rho}\left(\frac{1 + b_{T,k,n}}{1 + \frac{j}{k+1}b_{T,k,n}} \right) \right| \to_p 0
$$

(b)

$$
\max_{j=1,\ldots,k} \left(\frac{j}{k+1} \right)^{0.5+\varepsilon} \left| \sqrt{k} \left[\frac{X_{n-j+1,n} - X_{n-k,n}}{a_{T,k,n}} - \frac{1}{\xi}\left(\left(\frac{1 + \frac{j}{k+1}b_{T,k,n}}{1 + b_{T,k,n}} \right)^{-\xi} - 1 \right) \right] \right.
$$
$$
+ \frac{b_{T,k,n}}{1 + b_{T,k,n}} \left[\left(\frac{1 + \frac{j}{k+1}b_{T,k,n}}{1 + b_{T,k,n}} \right)^{-1-\xi} W_n\left(\frac{j}{k+1} \right) - W_n(1) \right]
$$
$$
\left. + \sqrt{k} A\left(\frac{1}{\bar{F}_Y(T)(1 + b_{T,k,n})} \right) \Psi_{\xi,\rho}\left(\frac{1 + b_{T,k,n}}{1 + \frac{j}{k+1}b_{T,k,n}} \right) \right| \to_p 0.
$$

Proof. In order to derive (a), note that for $j = 1, \ldots, k$,

$$
X_{n-j+1,n} - U_T\left(\frac{n+1}{k+1} \right) =_d U_T(Y_{n-j+1,n}) - U_T\left(\frac{n+1}{k+1} \right)
$$
$$
= U_Y\left(\frac{1 + b_{T,k,n}}{1 + \frac{1}{Y_{n-j+1,n}D_T}} \frac{1}{\bar{F}_Y(T)(1 + b_{T,k,n})} \right) - U_Y\left(\frac{1}{\bar{F}_Y(T)(1 + b_{T,k,n})} \right)
$$

where we used (7), and where $Y_{1,n} \leq Y_{2,n} \leq \ldots \leq Y_{n,n}$ denote the order statistics of an i.i.d. sample from a standard Pareto distribution with distribution function $1 - 1/x$ for $x \geq 1$. Hence, using (25) with

$$
t = \frac{1}{\bar{F}_Y(T)(1 + b_{T,k,n})} \quad \text{and} \quad x = \frac{1 + b_{T,k,n}}{1 + \frac{n+1}{jY_{n-j+1,n}}\frac{j}{k+1}b_{T,k,n}},
$$

we obtain

$$\frac{X_{n-j+1,n} - U_T\left(\frac{n+1}{k+1}\right)}{a_{T,k,n}} = \frac{1}{\xi}\left(\left(\frac{1 + \frac{n+1}{jY_{n-j+1,n}}\frac{j}{k+1}b_{T,k,n}}{1 + b_{T,k,n}}\right)^{-\xi} - 1\right)$$

$$+ A\left(\frac{1}{\overline{F}_Y(T)(1+b_{T,k,n})}\right)\Psi_{\xi,\rho}\left(\frac{1+b_{T,k,n}}{1+\frac{j}{k+1}b_{T,k,n}}\right) + o_p(1). \quad (27)$$

Using Lemma 2.4.10 in de Haan and Ferreira (2006) applied to the standard Pareto distribution one gets

$$\max_{j=1,\ldots,k}\left(\frac{j}{k+1}\right)^{0.5+\varepsilon}\left|\sqrt{k}\left(Y_{n-j+1,n}\frac{j}{n} - 1\right) - \left(\frac{j}{k+1}\right)^{-1}W_n\left(\frac{j}{k+1}\right)\right| \to_p 0.$$

Using the mean value theorem we now obtain

$$\frac{1}{\xi}\left(\left(\frac{1 + \frac{n+1}{jY_{n-j+1,n}}\frac{j}{k+1}b_{T,k,n}}{1 + b_{T,k,n}}\right)^{-\xi} - \left(\frac{1+\frac{j}{k+1}b_{T,k,n}}{1+b_{T,k,n}}\right)^{-\xi}\right)$$

$$= \frac{b_{T,k,n}}{1+b_{T,k,n}}\frac{j}{k+1}\left(\frac{1+\frac{j}{k+1}b_{T,k,n}}{1+b_{T,k,n}}\right)^{-1-\xi}\left(\frac{jY_{n-j+1,n}}{n} - 1\right)(1+o_p(1)).$$

Hence, combining this with (27) and the result from Lemma 2.4.10 in de Haan and Ferreira (2006), we arrive at (a). Combining (a) with the analogous result for $j = k + 1$, one arrives at (b). To this end note that $\Psi_{\xi,\rho}(1) = 0$. □

Proof of Theorem 1. This proof follows the approach of the proof of Theorem 3.4.2 in de Haan and Ferreira (2006). Let $\hat{\tau}_k a_{T,k,n} = \hat{\tau}_k^s$, and

$$Z_{T,k,n}\left(\frac{j}{k+1}\right) = \frac{b_{T,k,n}}{1+b_{T,k,n}}\left(\left(\frac{1+\frac{j}{k+1}b_{T,k,n}}{1+b_{T,k,n}}\right)^{-1-\xi}W_n\left(\frac{j}{k+1}\right) - W_n(1)\right)$$

$$+ \sqrt{k}A\left(\frac{1}{\overline{F}_Y(T)(1+b_{T,k,n})}\right)\Psi_{\xi,\rho}\left(\frac{1+b_{T,k,n}}{1+\frac{j}{k+1}b_{T,k,n}}\right).$$

Then, uniformly in $j \in \{1,\ldots,k\}$,

$$1 + \hat{\tau}_k^s\frac{E_{j,k}}{a_{T,k,n}} = \left(\frac{1+\frac{j}{k+1}b_{T,k,n}}{1+b_{T,k,n}}\right)^{-\xi} + \frac{1}{\xi}(\hat{\tau}_k^s - \xi)\left(\left(\frac{1+\frac{j}{k+1}b_{T,k,n}}{1+b_{T,k,n}}\right)^{-\xi} - 1\right)$$

$$+ \hat{\tau}_k^s\frac{1}{\sqrt{k}}Z_{T,k,n}\left(\frac{j}{k+1}\right) + o_p(1).$$

Using $\log(1 + u) = u(1 + o(1))$ if $u \downarrow 0$, we get

$$\log\left(\left(\frac{1+\frac{j}{k+1}b_{T,k,n}}{1+b_{T,k,n}}\right)^{\xi}\left(1+\hat{\tau}_k^s\frac{E_{j,k}}{a_{T,k,n}}\right)\right)=\frac{1}{\xi}(\hat{\tau}_k^s-\xi)\left(1-\left(\frac{1+\frac{j}{k+1}b_{T,k,n}}{1+b_{T,k,n}}\right)^{\xi}\right)$$

$$+\hat{\tau}_k^s\frac{1}{\sqrt{k}}Z_{T,k,n}\left(\frac{j}{k+1}\right)\left(\frac{1+\frac{j}{k+1}b_{T,k,n}}{1+b_{T,k,n}}\right)^{\xi}+o_p(1).$$

Hence, the first term on the left hand side of (17) is given by

$$\frac{1}{k-1}\sum_{j=2}^{k}\log(1+\hat{\tau}_kE_{j,k})=\left[-\xi\int_0^1\log\left(\frac{1+ub_{T,k,n}}{1+b_{T,k,n}}\right)\,du\right.$$

$$+\frac{1}{\xi}(\hat{\tau}_k^s-\xi)\int_0^1\left(1-\left(\frac{1+ub_{T,k,n}}{1+b_{T,k,n}}\right)^{\xi}\right)\,du$$

$$+\hat{\tau}_k^s\frac{1}{\sqrt{k}}\int_0^1 Z_{T,k,n}\left(u\right)\left(\frac{1+ub_{T,k,n}}{1+b_{T,k,n}}\right)^{\xi}\,du\Bigg]$$

$$\sim\left[\xi\left(1-\frac{\log(1+b_{T,k,n})}{b_{T,k,n}}\right)\right.$$

$$+\frac{1}{\xi}(\hat{\tau}_k^s-\xi)\left(1-\frac{1+b_{T,k,n}}{b_{T,k,n}(1+\xi)}\left(1-(1+b_{T,k,n})^{-1-\xi}\right)\right)$$

$$+\hat{\tau}_k^s\frac{1}{\sqrt{k}}\int_0^1 Z_{T,k,n}\left(u\right)\left(\frac{1+ub_{T,k,n}}{1+b_{T,k,n}}\right)^{\xi}\,du\Bigg].\qquad(28)$$

Moreover, using Proposition 1(b) with $j=1$, we obtain

$$\left(1+\hat{\tau}_k^s\frac{E_{1,k}}{a_{T,k,n}}\right)^{-1/\hat{\xi}_k}=\left(1+\hat{\tau}_k^s\frac{1}{\xi}\left((1+b_{T,k,n})^{\xi}-1\right)+\hat{\tau}_k^s\frac{1}{\sqrt{k}}Z_{T,k,n}\left(\frac{1}{k+1}\right)\right)^{-1/\hat{\xi}_k}$$

$$=(1+b_{T,k,n})^{-\frac{\xi}{\hat{\xi}_k}}\left(1+(\hat{\tau}_k^s-\xi)\frac{1}{\xi}\left(1-(1+b_{T,k,n})^{-\xi}\right)\right.$$

$$\left.+\hat{\tau}_k^s\frac{1}{\sqrt{k}}Z_{T,k,n}\left(\frac{1}{k+1}\right)(1+b_{T,k,n})^{-\xi}\right)^{-1/\hat{\xi}_k}$$

$$=(1+b_{T,k,n})^{-1}\left(1+(\hat{\xi}_k-\xi)\frac{1}{\xi}\log(1+b_{T,k,n})\right.$$

$$-(\hat{\tau}_k^s-\xi)\frac{1}{\xi^2}\left(1-(1+b_{T,k,n})^{-\xi}\right)$$

$$\left.-\frac{\hat{\tau}_k^s}{\hat{\xi}_k}\frac{1}{\sqrt{k}}Z_{T,k,n}\left(\frac{1}{k+1}\right)(1+b_{T,k,n})^{-\xi}\right)(1+o_p(1))$$

where we used the series expansions

$$e^{-\left(\frac{\xi}{\hat{\xi}_k}-1\right)\log(1+b_{T,n,k})}=1-\left(\frac{\xi}{\hat{\xi}_k}-1\right)\log(1+b_{T,n,k})(1+o_p(1))$$

and $(1+u)^{-1/\xi}=1-\frac{u}{\xi}(1+o(1))$. Hence, the second term on the left hand side

of (17) equals

$$
-\hat{\xi}_k \frac{\left(1 + \hat{\tau}_k^s \frac{E_{1,k}}{a_{T,k,n}}\right)^{-1/\hat{\xi}_k} \log\left(1 + \hat{\tau}_k^s \frac{E_{1,k}}{a_{T,k,n}}\right)^{-1/\hat{\xi}_k}}{1 - \left(1 + \hat{\tau}_k^s \frac{E_{1,k}}{a_{T,k,n}}\right)^{-1/\hat{\xi}_k}}
$$

$$
= -\hat{\xi}_k (1 + b_{T,k,n})^{-1} \left(1 + \frac{(\hat{\xi}_k - \xi)}{\xi} \log(1 + b_{T,k,n}) - \frac{(\hat{\tau}_k^s - \xi)}{\xi^2}\left(1 - (1 + b_{T,k,n})^{-\xi}\right)\right.
$$

$$
\left. - \frac{\hat{\tau}_k^s}{\hat{\xi}_k} \frac{1}{\sqrt{k}} Z_{T,k,n}\left(\frac{1}{k+1}\right)(1 + b_{T,k,n})^{-\xi}\right)
$$

$$
\times \log(1 + b_{T,k,n})^{-1} \times \left(1 - \frac{(\hat{\xi}_k - \xi)}{\xi} + \frac{(\hat{\tau}_k^s - \xi)}{\xi^2}\frac{\left(1 - (1 + b_{T,k,n})^{-\xi}\right)}{\log(1 + b_{T,k,n})}\right.
$$

$$
\left. + \frac{\hat{\tau}_k^s}{\hat{\xi}_k} \frac{1}{\sqrt{k}} Z_{T,k,n}\left(\frac{1}{k+1}\right)\frac{(1 + b_{T,k,n})^{-\xi}}{\log(1 + b_{T,k,n})}\right)
$$

$$
/ \left[\frac{b_{T,k,n}}{1 + b_{T,k,n}}\left(1 - \frac{(\hat{\xi}_k - \xi)}{\xi}\frac{\log(1 + b_{T,k,n})}{b_{T,k,n}} + \frac{(\hat{\tau}_k^s - \xi)}{\xi^2}\frac{\left(1 - (1 + b_{T,k,n})^{-\xi}\right)}{b_{T,k,n}}\right.\right.
$$

$$
\left.\left. + \frac{\hat{\tau}_k^s}{\hat{\xi}_k}\frac{1}{\sqrt{k}} Z_{T,k,n}\left(\frac{1}{k+1}\right)\frac{(1 + b_{T,k,n})^{-\xi}}{b_{T,k,n}}\right)\right](1 + o_p(1))
$$

$$
\sim \left[\hat{\xi}_k \frac{\log(1 + b_{T,k,n})}{b_{T,k,n}} + (\hat{\xi}_k - \xi)\frac{\log(1 + b_{T,k,n})}{b_{T,k,n}}\left(-1 + \frac{1 + b_{T,k,n}}{b_{T,k,n}}\log(1 + b_{T,k,n})\right)\right.
$$

$$
- \frac{(\hat{\tau}_k^s - \xi)}{\xi}\left(1 - (1 + b_{T,k,n})^{-\xi}\right)\left(\frac{1 + b_{T,k,n}}{b_{T,k,n}} - \frac{1}{\log(1 + b_{T,k,n})}\right)\frac{\log(1 + b_{T,k,n})}{b_{T,k,n}}
$$

$$
- \hat{\tau}_k^s \frac{1}{\sqrt{k}} Z_{T,k,n}\left(\frac{1}{k+1}\right)(1 + b_{T,k,n})^{-\xi}
$$

$$
\left. \times \left(\frac{1 + b_{T,k,n}}{b_{T,k,n}} - \frac{1}{\log(1 + b_{T,k,n})}\right)\frac{\log(1 + b_{T,k,n})}{b_{T,k,n}}\right]. \tag{29}
$$

Combining (17), (28) and (29) gives

$$
\left[\xi\left(1 - \frac{\log(1 + b_{T,k,n})}{b_{T,k,n}}\right) + \frac{1}{\xi}(\hat{\tau}_k^s - \xi)\left(1 - \frac{1 + b_{T,k,n}}{b_{T,k,n}(1 + \xi)}\left(1 - (1 + b_{T,k,n})^{-1-\xi}\right)\right)\right.
$$

$$
\left. + \hat{\tau}_k^s \frac{1}{\sqrt{k}} \int_0^1 Z_{T,k,n}(u)\left(\frac{1 + u b_{T,k,n}}{1 + b_{T,k,n}}\right)^{\xi} du\right](1 + o_p(1))
$$

$$
+ \left[\hat{\xi}_k \frac{\log(1 + b_{T,k,n})}{b_{T,k,n}} + (\hat{\xi}_k - \xi)\frac{\log(1 + b_{T,k,n})}{b_{T,k,n}}\left(-1 + \frac{1 + b_{T,k,n}}{b_{T,k,n}}\log(1 + b_{T,k,n})\right)\right.
$$

$$
- \frac{(\hat{\tau}_k^s - \xi)}{\xi}\left(1 - (1 + b_{T,k,n})^{-\xi}\right)\left(\frac{1 + b_{T,k,n}}{b_{T,k,n}} - \frac{1}{\log(1 + b_{T,k,n})}\right)\frac{\log(1 + b_{T,k,n})}{b_{T,k,n}}
$$

$$
- \hat{\tau}_k^s \frac{1}{\sqrt{k}} Z_{T,k,n}\left(\frac{1}{k+1}\right)(1 + b_{T,k,n})^{-\xi}
$$

$$
\left. \times \left(\frac{1 + b_{T,k,n}}{b_{T,k,n}} - \frac{1}{\log(1 + b_{T,k,n})}\right)\frac{\log(1 + b_{T,k,n})}{b_{T,k,n}}\right] = \hat{\xi}_k(1 + o_p(1)).
$$

This equation can be written as

$$
\begin{aligned}
0 = \Bigg[&(\hat{\xi}_k - \xi)\left(-1 + \frac{(1 + b_{T,k,n})\log^2(1 + b_{T,k,n})}{b_{T,k,n}^2}\right) \\
&+ \frac{1}{\xi}(\hat{\tau}_k^s - \xi)\left(\frac{\xi}{1+\xi}\frac{1 + b_{T,k,n}}{b_{T,k,n}}\left(1 - (1 + b_{T,k,n})^{-1-\xi}\right)\right. \\
&\left. \qquad - \frac{(1 + b_{T,k,n})}{b_{T,k,n}^2}\log(1 + b_{T,k,n})\left(1 - (1 + b_{T,k,n})^{-\xi}\right)\right) \\
&+ \frac{\hat{\tau}_k^s}{\sqrt{k}}\int_0^1 Z_{T,k,n}(u)\left(\frac{1 + ub_{T,k,n}}{1 + b_{T,k,n}}\right)^{\xi} du \\
&- \frac{\hat{\tau}_k^s}{\sqrt{k}}Z_{T,k,n}\left(\frac{1}{k+1}\right)(1 + b_{T,k,n})^{-\xi} \\
&\qquad \times \left(\frac{(1 + b_{T,k,n})\log(1 + b_{T,k,n})}{b_{T,k,n}^2} - \frac{1}{b_{T,k,n}}\right)\Bigg](1 + o_p(1)). \quad (30)
\end{aligned}
$$

The left hand side of (18) yields, using similar asymptotic methods as above,

$$
\begin{aligned}
&\frac{1}{k-1}\sum_{j=2}^k \left(\frac{1 + \frac{j}{k+1}b_{T,k,n}}{1 + b_{T,k,n}}\right)^{\xi}\left[1 - \frac{(\hat{\tau}_k^s - \xi)}{\xi}\left(1 - \left(\frac{1 + \frac{j}{k+1}b_{T,k,n}}{1 + b_{T,k,n}}\right)^{\xi}\right)\right. \\
&\left. \qquad - \frac{\hat{\tau}_k^s}{\sqrt{k}}Z_{T,k,n}\left(\frac{j}{k+1}\right)\left(1 - \left(\frac{1 + \frac{j}{k+1}b_{T,k,n}}{1 + b_{T,k,n}}\right)^{\xi}\right)\right](1 + o_p(1)) \\
&= \Bigg[\frac{1}{1+\xi}\frac{1 + b_{T,k,n}}{b_{T,k,n}}\left(1 - (1 + b_{T,k,n})^{-1-\xi}\right) \\
&\qquad - \frac{(\hat{\tau}_k^s - \xi)}{\xi}\left(\frac{\xi(1 + b_{T,k,n})}{(1+\xi)(1+2\xi)} - \frac{(1 + b_{T,k,n})^{-\xi}}{1+\xi} + \frac{(1 + b_{T,k,n})^{-2\xi}}{1+2\xi}\right) \\
&\qquad - \frac{\hat{\tau}_k^s}{\sqrt{k}}\int_0^1 Z_{T,k,n}(u)\left(\frac{1 + ub_{T,k,n}}{1 + b_{T,k,n}}\right)^{2\xi} du\Bigg](1 + o_p(1)). \quad (31)
\end{aligned}
$$

The right hand side of (18) is asymptotically equivalent to (where we used again Proposition 1(b) with $j = 1$)

$$
\begin{aligned}
&\frac{1}{1 + \hat{\xi}_k}\Bigg[1 - \frac{1}{1 + b_{T,k,n}}\left(1 + \frac{\hat{\xi}_k - \xi}{\xi}\log(1 + b_{T,k,n}) - \frac{\hat{\tau}_k^s - \xi}{\xi^2}\left(1 - (1 + b_{T,k,n})^{-\xi}\right)\right. \\
&\left. \qquad - \hat{\tau}_k^s \frac{1}{\hat{\xi}_k\sqrt{k}}Z_{T,k,n}\left(\frac{1}{k+1}\right)(1 + b_{T,k,n})^{-\xi}\right) \\
&\qquad \times \left((1 + b_{T,k,n})^{\xi} + \frac{\hat{\tau}_k^s - \xi}{\xi}\left((1 + b_{T,k,n})^{\xi} - 1\right) + \hat{\tau}_k^s \frac{1}{\sqrt{k}}Z_{T,k,n}\left(\frac{1}{k+1}\right)\right)^{-1}\Bigg] \\
&\times \Bigg[1 - \frac{1}{1 + b_{T,k,n}}\left(1 + \frac{\hat{\xi}_k - \xi}{\xi}\log(1 + b_{T,k,n}) - \frac{\hat{\tau}_k^s - \xi}{\xi^2}\left(1 - (1 + b_{T,k,n})^{-\xi}\right)\right.
\end{aligned}
$$

$$\left. -\hat{\tau}_k^s \frac{1}{\hat{\xi}_k \sqrt{k}} Z_{T,k,n} \left(\frac{1}{k+1} \right) (1 + b_{T,k,n})^{-\xi} \right)\right]^{-1}$$

$$\sim \frac{1}{1+\hat{\xi}_k} \frac{(1 + b_{T,k,n}) \left(1 - (1 + b_{T,k,n})^{-1-\xi} \right)}{b_{T,k,n}}$$

$$\times \left(1 - \frac{\hat{\xi}_k - \xi}{\xi} \log(1 + b_{T,k,n}) \frac{(1 + b_{T,k,n})^{-1-\xi}}{1 - (1 + b_{T,k,n})^{-1-\xi}} \right.$$

$$+ (\hat{\tau}_k^s - \xi) \frac{1+\xi}{\xi^2} \frac{(1 + b_{T,k,n})^{-1-\xi}}{1 - (1 + b_{T,k,n})^{-1-\xi}} \left(1 - (1 + b_{T,k,n})^{-\xi} \right)$$

$$\left. + \frac{\hat{\tau}_k^s}{\hat{\xi}_k} \frac{1}{\sqrt{k}} Z_{T,k,n} \left(\frac{1}{k+1} \right) \frac{1+\hat{\xi}_k}{\hat{\xi}_k} \frac{(1 + b_{T,k,n})^{-1-2\xi}}{1 - (1 + b_{T,k,n})^{-1-\xi}} \right)$$

$$\times \left[1 - \frac{\hat{\xi}_k - \xi}{\xi} \frac{\log(1 + b_{T,k,n})}{b_{T,k,n}} + \frac{\hat{\tau}_k^s - \xi}{\xi^2} \frac{1 - (1 + b_{T,k,n})^{-\xi}}{b_{T,k,n}} \right.$$

$$\left. + \frac{\hat{\tau}_k^s}{\hat{\xi}_k} \frac{1}{\sqrt{k}} Z_{T,k,n} \left(\frac{1}{k+1} \right) \frac{(1 + b_{T,k,n})^{-\xi}}{b_{T,k,n}} \right]^{-1}$$

$$\sim \frac{1}{1+\hat{\xi}_k} \frac{(1 + b_{T,k,n}) \left(1 - (1 + b_{T,k,n})^{-1-\xi} \right)}{b_{T,k,n}}$$

$$+ \frac{(\hat{\xi}_k - \xi)}{\xi(1+\xi)} \frac{1 + b_{T,k,n}}{b_{T,k,n}^2} \log(1 + b_{T,k,n}) \left(1 - (1 + b_{T,k,n})^{-\xi} \right)$$

$$+ \frac{\hat{\tau}_k^s - \xi}{\xi^2(1+\xi)} \frac{1 + b_{T,k,n}}{b_{T,k,n}} \left(1 - (1 + b_{T,k,n})^{-\xi} \right) \left(- \frac{1 - (1 + b_{T,k,n})^{-1-\xi}}{b_{T,k,n}} \right.$$

$$\left. + (1+\xi)(1 + b_{T,k,n})^{-1-\xi} \right)$$

$$- \frac{1}{1+\xi} \frac{\hat{\tau}_k^s}{\hat{\xi}_k} \frac{1}{\sqrt{k}} Z_{T,k,n} \left(\frac{1}{k+1} \right) \frac{(1 + b_{T,k,n})^{1-\xi}}{b_{T,k,n}^2} \left(1 - (1 + b_{T,k,n})^{-\xi} \right). \qquad (32)$$

Combining (18), (31) and (32) leads to (after some lengthy calculations)

$$\frac{1}{\xi} \frac{1 + b_{T,k,n}}{b_{T,k,n}} \left(\frac{\xi}{1+\xi} \left(1 - (1 + b_{T,k,n})^{-1-\xi} \right) \right.$$

$$\left. - \frac{\log(1 + b_{T,k,n})}{b_{T,k,n}} \left(1 - (1 + b_{T,k,n})^{-\xi} \right) \right)$$

$$- (\hat{\tau}_k^s - \xi) \frac{1 + b_{T,k,n}}{b_{T,k,n}} \frac{1}{\xi} \left(\frac{\xi}{1+2\xi} \left(1 - (1 + b_{T,k,n})^{-1-2\xi} \right) \right.$$

$$\left. - \frac{1}{b_{T,k,n}} \frac{1}{\xi} \left(1 - (1 + b_{T,k,n})^{-\xi} \right)^2 \right)$$

$$= \frac{\xi(\xi+1)}{\sqrt{k}} \int_0^1 Z_{T,k,n}(u) \left(\frac{1 + u b_{T,k,n}}{1 + b_{T,k,n}} \right)^{2\xi} du$$

$$- \frac{1}{\sqrt{k}} Z_{T,k,n} \left(\frac{1}{k+1} \right) \frac{(1 + b_{T,k,n})^{1-\xi}}{b_{T,k,n}^2} \left(1 - (1 + b_{T,k,n})^{-\xi} \right). \qquad (33)$$

$\square$

Proof of Theorem 2.

$$\hat{Q}_{T,k}(1-p)$$

$$= X_{n-k,n} + \frac{1}{\hat{\tau}_k}\left(\left(1+\frac{k}{n\hat{D}_T}\right)^{\hat{\xi}_k}\left(1+\frac{1}{d_n}\frac{k}{n\hat{D}_T}\right)^{-\hat{\xi}_k} - 1\right)$$

$$= X_{n-k,n} + \frac{1}{\hat{\tau}_k}\left(\left(\frac{1-\frac{1}{k}}{(1+\hat{\tau}_k E_{1,k})^{-\frac{1}{\hat{\xi}_k}} - \frac{1}{k}}\right)^{\hat{\xi}_k}\left(1+\frac{1}{d_n}\frac{k}{n\hat{D}_T}\right)^{-\hat{\xi}_k} - 1\right)$$

$$= X_{n-k,n} + \frac{1}{\hat{\tau}_k}\left((1+\hat{\tau}_k E_{1,k})\left(\frac{1-\frac{1}{k}}{1-\frac{1}{k}(1+\hat{\tau}_k E_{1,k})^{\frac{1}{\hat{\xi}_k}}}\right)^{\hat{\xi}_k}\left(1+\frac{1}{d_n}\frac{k}{n\hat{D}_T}\right)^{-\hat{\xi}_k} - 1\right)$$

$$= X_{n-k,n} + \frac{1}{\hat{\tau}_k}\left((1+\hat{\tau}_k E_{1,k})\left(1-\frac{\hat{\xi}_k}{k}\left(1-(1+\hat{\tau}_k E_{1,k})^{\frac{1}{\hat{\xi}_k}}\right)(1+o_p(1))\right)\right.$$

$$\times\left.\left(1-\frac{\hat{\xi}_k}{d_n}\frac{k}{n\hat{D}_T}(1+o_p(1))\right) - 1\right)$$

$$= X_{n-k,n} + \left(E_{1,k} + (1+\hat{\tau}_k E_{1,k})\left(-\frac{\hat{\xi}_k}{\hat{\tau}_k k}\left(1-(1+\hat{\tau}_k E_{1,k})^{\frac{1}{\hat{\xi}_k}}\right) - \frac{\hat{\xi}_k}{d_n\hat{\tau}_k}\frac{k}{n\hat{D}_T}\right)\right.$$

$$\times\left.(1+o_p(1))\right)$$

$$= X_{n,n} - \frac{\hat{\xi}_k}{\hat{\tau}_k}(1+\hat{\tau}_k E_{1,k})\left(\frac{1}{k}\left(1-(1+\hat{\tau}_k E_{1,k})^{\frac{1}{\hat{\xi}_k}}\right) + \frac{1}{d_n}\frac{k}{n\hat{D}_T}\right)(1+o_p(1)).$$

Hence,

$$\hat{Q}_{T,k}(1-p) - Q_T(1-p)$$

$$= \left(X_{n,n} - Q_T\left(1-\frac{1}{n}\right)\right) + \left(Q_T\left(1-\frac{1}{n}\right) - Q_T(1-p)\right)$$

$$- \frac{\hat{\xi}_k}{\hat{\tau}_k}(1+\hat{\tau}_k E_{1,k})\left(\frac{1}{k}\left(1-(1+\hat{\tau}_k E_{1,k})^{\frac{1}{\hat{\xi}_k}}\right) + \frac{1}{d_n}\frac{k}{n\hat{D}_T}\right)\left(1+o_p\left(\frac{1}{d_n}\right)\right).$$
(34)

First, using again the notation $Y_{1,n} \le Y_{2,n} \le \ldots \le Y_{n,n}$ for the order statistics of an i.i.d. sample of size n from a standard Pareto distribution, we obtain using (26)

$$X_{n,n} - Q_T\left(1-\frac{1}{n}\right) =_d U_T(Y_{n,n}) - U_T(n)$$

$$= U_Y\left(\frac{1}{\bar{F}_Y(T)\left(1+\frac{n}{Y_{n,n}}\frac{1}{n\hat{D}_T}\right)}\right) - U_Y\left(\frac{1}{\bar{F}_Y(T)\left(1+\frac{1}{n\hat{D}_T}\right)}\right)$$

$$= U_Y\left(\frac{1+\frac{1}{n\hat{D}_T}}{1+\frac{n}{Y_{n,n}}\frac{1}{n\hat{D}_T}}\frac{1}{\bar{F}_Y(T)\left(1+\frac{1}{n\hat{D}_T}\right)}\right) - U_Y\left(\frac{1}{\bar{F}_Y(T)\left(1+\frac{1}{n\hat{D}_T}\right)}\right)$$

$$= a_Y \left(\frac{1}{\bar{F}_Y(T)\left(1 + \frac{1}{k}b_{T,k,n}\right)} \right) \left(\frac{1}{\xi} \left(\left(\frac{1 + \frac{b_{T,k,n}}{k}}{1 + \frac{n}{Y_{n,n}}\frac{b_{T,k,n}}{k}} \right)^{\xi} - 1 \right) \right)$$

$$+ A\left(\frac{1}{\bar{F}_Y(T)\left(1 + \frac{1}{k}b_{T,k,n}\right)} \right) \Psi_{\xi,\rho}\left(\frac{1 + \frac{b_{T,k,n}}{k}}{1 + \frac{n}{Y_{n,n}}\frac{b_{T,k,n}}{k}} \right) \right) (1 + o_p(1))$$

$$= a_Y\left(\frac{1}{\bar{F}_Y(T)} \right)\left(1 + \frac{b_{T,k,n}}{k} \right)^{-\xi}\left(1 + A\left(\frac{1}{\bar{F}_Y(T)} \right)C\left(\frac{\left(1 + \frac{b_{T,k,n}}{k}\right)^{-\rho} - 1}{\rho} \right) \right)$$

$$\times \left(-b_{T,k,n}\frac{1}{k}\left(\frac{n}{Y_{n,n}} - 1 \right)\left(1 + O_p\left(\frac{1}{k} \right) \right) + O_p\left(\frac{1}{k^2} \right) \right)$$

$$= a_Y\left(\frac{1}{\bar{F}_Y(T)} \right)\left(1 - \frac{\xi b_{T,k,n}}{k} - A\left(\frac{1}{\bar{F}_Y(T)} \right)C\frac{b_{T,k,n}}{k} + O_p\left(\frac{1}{k^2} \right) \right)$$

$$\times \left(-\frac{b_{T,k,n}}{k}(E - 1) + O_p\left(\frac{1}{k^2} \right) \right). \tag{35}$$

Here, we used that $\frac{n}{Y_{n,n}} =_d E + O_p\left(\frac{1}{n} \right)$ and that $\Psi_{\xi,\rho}\left(1 + \frac{D}{k} \right) = O\left(\frac{1}{k^2} \right)$ for any constant D. Furthermore,

$$Q_T\left(1 - \frac{1}{n} \right) - Q_T(1 - p)$$

$$= U_Y\left(\frac{1}{\bar{F}_Y(T)\left(1 + \frac{1}{nD_T}\right)} \right) - U_Y\left(\frac{1}{\bar{F}_Y(T)\left(1 + \frac{p}{D_T}\right)} \right)$$

$$= U_Y\left(\frac{1 + \frac{b_{T,k,n}}{d_n}}{1 + \frac{b_{T,k,n}}{k}}\frac{1}{\bar{F}_Y(T)\left(1 + \frac{b_{T,k,n}}{d_n}\right)} \right) - U_Y\left(\frac{1}{\bar{F}_Y(T)\left(1 + \frac{p}{D_T}\right)} \right)$$

$$= a_Y\left(\frac{1}{\bar{F}_Y(T)\left(1 + \frac{b_{T,k,n}}{d_n}\right)} \right)\left(\frac{1}{\xi}\left(\left(\frac{1 + \frac{b_{T,k,n}}{d_n}}{1 + \frac{b_{T,k,n}}{k}} \right)^{\xi} - 1 \right) \right)$$

$$+ A\left(\frac{1}{\bar{F}_Y(T)\left(1 + \frac{b_{T,k,n}}{d_n}\right)} \right)\Psi_{\xi,\rho}\left(\frac{1 + \frac{b_{T,k,n}}{d_n}}{1 + \frac{b_{T,k,n}}{k}} \right)\right)(1 + o_p(1))$$

$$= a_Y\left(\frac{1}{\bar{F}_Y(T)} \right)\left(1 + \frac{b_{T,k,n}}{d_n} \right)^{-\xi}\left(1 + A\left(\frac{1}{\bar{F}_Y(T)} \right)C\left(\frac{\left(1 + \frac{b_{T,k,n}}{d_n}\right)^{-\rho} - 1}{\rho} \right) \right)$$

$$\times \left(b_{T,k,n}\left(\frac{1}{d_n} - \frac{1}{k} \right)\left(1 + O\left(\frac{1}{d_n} \right) \right) + O\left(\frac{1}{d_n^2} \right) \right)$$

$$= a_Y\left(\frac{1}{\bar{F}_Y(T)} \right)\left(b_{T,k,n}\left(\frac{1}{d_n} - \frac{1}{k} \right) + O\left(\frac{1}{d_n^2} \vee \frac{1}{k^2} \right) \right). \tag{36}$$

Finally, using $k/(n\hat{D}_T) = \left((1 + \hat{\tau}_k E_{1,k})^{\frac{1}{\hat{\xi}_k}} - 1 \right)(1 + O_p(1/k))$ and derivations

as in the proof of Theorem 1, the third term in the right hand side of (34) equals

$$
-\left(\frac{\hat{\xi}_k}{\hat{\tau}_k}\frac{1}{a_{T,k,n}}\right)a_{T,k,n}(1+\hat{\tau}_k E_{1,k})\left(\frac{1}{k}\left(1-(1+\hat{\tau}_k E_{1,k})^{\frac{1}{\hat{\xi}_k}}\right)+\frac{1}{d_n}\frac{k}{n\hat{D}_T}\right)
$$

$$
=-a_Y\left(\frac{1}{\bar{F}_Y(T)}\right)(1+b_{T,k,n})^{-\xi}\left(1+A\left(\frac{1}{\bar{F}_Y(T)}\right)C\left(\frac{(1+b_{T,k,n})^{-\rho}-1}{\rho}\right)\right)
$$

$$
\times\left(1+\left(\frac{\hat{\xi}_k}{\hat{\tau}_k}\frac{1}{a_{T,k,n}}-1\right)\right)
$$

$$
\times(1+b_{T,k,n})^{\xi}\left(1+(\hat{\tau}_k^s-\xi)\frac{1}{\xi}\left(1-(1+b_{T,k,n})^{-\xi}\right)\right.
$$

$$
\left.+\frac{\hat{\tau}_k^s}{\sqrt{k}}Z_{T,k,n}\left(\frac{1}{k+1}\right)(1+b_{T,k,n})^{-\xi}\right)
$$

$$
\times\left(\left(1-(1+\hat{\tau}_k E_{1,k})^{\frac{1}{\hat{\xi}_k}}\right)\left(\frac{1}{d_n}-\frac{1}{k}\right)+O_p\left(\frac{1}{d_n k}\right)\right)
$$

$$
=-a_Y\left(\frac{1}{\bar{F}_Y(T)}\right)\left(1+A\left(\frac{1}{\bar{F}_Y(T)}\right)C\left(\frac{(1+b_{T,k,n})^{-\rho}-1}{\rho}\right)\right)\left(1+O_p\left(\frac{1}{k}\right)\right)
$$

$$
\times\left(1+\left(\frac{\hat{\xi}_k}{\hat{\tau}_k}\frac{1}{a_{T,k,n}}-1\right)\right)
$$

$$
\times\left(1+(\hat{\tau}_k^s-\xi)\frac{1}{\xi}\left(1-(1+b_{T,k,n})^{-\xi}\right)+\frac{\hat{\tau}_k^s}{\sqrt{k}}Z_{T,k,n}\left(\frac{1}{k+1}\right)(1+b_{T,k,n})^{-\xi}\right)
$$

$$
\times b_{T,k,n}\left(\frac{1}{d_n}-\frac{1}{k}\right)
$$

$$
\times\left(1-(\hat{\xi}_k-\xi)\frac{1}{\xi}\frac{1+b_{T,k,n}}{b_{T,k,n}}\log(1+b_{T,k,n})\right.
$$

$$
+(\hat{\tau}_k^s-\xi)\frac{1}{\xi^2}\frac{1+b_{T,k,n}}{b_{T,k,n}}\left(1-(1+b_{T,k,n})^{-\xi}\right)
$$

$$
\left.+\frac{1}{\sqrt{k}}Z_{T,k,n}\left(\frac{1}{k+1}\right)\frac{(1+b_{T,k,n})^{1-\xi}}{b_{T,k,n}}\right). \tag{37}
$$

The result follows from joining (34), (35), (36) and (37) and retaining terms of order $O\left(\frac{1}{k}\right)$, $O\left(\left(\frac{1}{d_n}-\frac{1}{k}\right)A\left(\frac{1}{\bar{F}_Y(T)}\right)\right)$ and $O\left(\left(\frac{1}{d_n}-\frac{1}{k}\right)\frac{1}{\sqrt{k}}\right)$. ☐

Proof of Theorem 3. Note that using (6) and $\bar{F}_Y(T)=D_T F_Y(T)$, we obtain

$$
T_{k,n}=k\left(1+\hat{\tau}_k^s\frac{E_{1,k}}{a_{T,k,n}}\right)^{-1/\hat{\xi}_k}
$$

$$
=k\left(1+\hat{\tau}_k^s\frac{U_T(Y_{n,n})-U_T(Y_{n-k,n})}{a_{T,k,n}}\right)^{-1/\hat{\xi}_k}
$$

$$= k \left(1 + \frac{\hat{\tau}_k^s}{a_Y \left(\frac{1}{D_T(1+b_{T,k,n})F_Y(T)} \right)} \right.$$

$$\left. \times \left[U_Y \left(\frac{Y_{n,n}}{F_Y(T)(1+Y_{n,n}D_T)} \right) - U_Y \left(\frac{Y_{n-k,n}}{F_Y(T)(1+Y_{n-k,n}D_T)} \right) \right] \right)^{-1/\hat{\xi}_k}$$

$$= k \left(1 + \frac{\hat{\tau}_k^s}{a_Y \left(\frac{n/k}{(1+nD_T/k)F_Y(T)} \right)} \right.$$

$$\times \left[U_Y \left(\frac{\frac{Y_{n,n}}{Y_{n-k,n}}}{\frac{1+Y_{n,n}D_T}{1+Y_{n-k,n}D_T}} \frac{\frac{kY_{n-k,n}}{n}\frac{n}{k}}{F_Y(T)\left(1 + \frac{kY_{n-k,n}}{n}\frac{nD_T}{k}\right)} \right) \right.$$

$$\left. \left. - U_Y \left(\frac{\frac{kY_{n-k,n}}{n}\frac{n}{k}}{F_Y(T)\left(1 + \frac{kY_{n-k,n}}{n}\frac{nD_T}{k}\right)} \right) \right] \right)^{-1/\hat{\xi}_k}.$$

Now one applies (25) with $t = \frac{\frac{kY_{n-k,n}}{n}\frac{n}{k}}{F_Y(T)\left(1+\frac{kY_{n-k,n}}{n}\frac{nD_T}{k}\right)} = \frac{n}{k}(1 + o_p(1))$ and $x = \frac{Y_{n,n}}{Y_{n-k,n}}\frac{1+Y_{n-k,n}D_T}{1+Y_{n,n}D_T} = U_{1,k}^{-1}(1 + o_p(1))$ since $\frac{kY_{n-k,n}}{n} = 1 + O_p(1/\sqrt{k})$, $Y_{n,n}/n = 1 + o_p(1)$, $nD_T \to 0$ and $Y_{n-k,n}/Y_{n,n} =_d U_{1,k}$, the minimum of an i.i.d. sample of size k from the uniform (0,1) distribution. This, with $\hat{\tau}_k^s/\xi = 1 + o_p(1)$, yields

$$T_{k,n}$$

$$= k \left(1 + \frac{\hat{\tau}_k^s}{\xi} \left[U_{1,k}^{-\xi}(1 + o_p(1)) - 1 + \xi A \left(\frac{n}{k}(1 + o_p(1)) \right) \Psi_{\xi,\rho} \left(U_{1,k}^{-1}(1 + o_p(1)) \right) \right] \right)^{-1/\hat{\xi}_k}$$

$$= k \left(U_{1,k}^{-\xi}(1 + o_p(1)) + \xi A \left(\frac{n}{k}(1 + o_p(1)) \right) \Psi_{\xi,\rho} \left(U_{1,k}^{-1}(1 + o_p(1)) \right) \right)^{-(1/\xi)(1+O_p(1/\sqrt{k}))}$$

$$= k U_{1,k} \left(1 + o_p(1) + \xi U_{1,k}^{\xi} A \left(\frac{n}{k}(1 + o_p(1)) \right) \Psi_{\xi,\rho}(k(1 + o_p(1))) \right)^{-1/\xi}$$

because $U_{1,k}^{-1} = O_p(k)$ and $U_{1,k}^{\xi}\Psi_{\xi,\rho}(k(1 + o_p(1))) = O_p(1)$. The result now follows from $kU_{1,k} =_d E(1 + o_p(1))$. $\qquad\square$
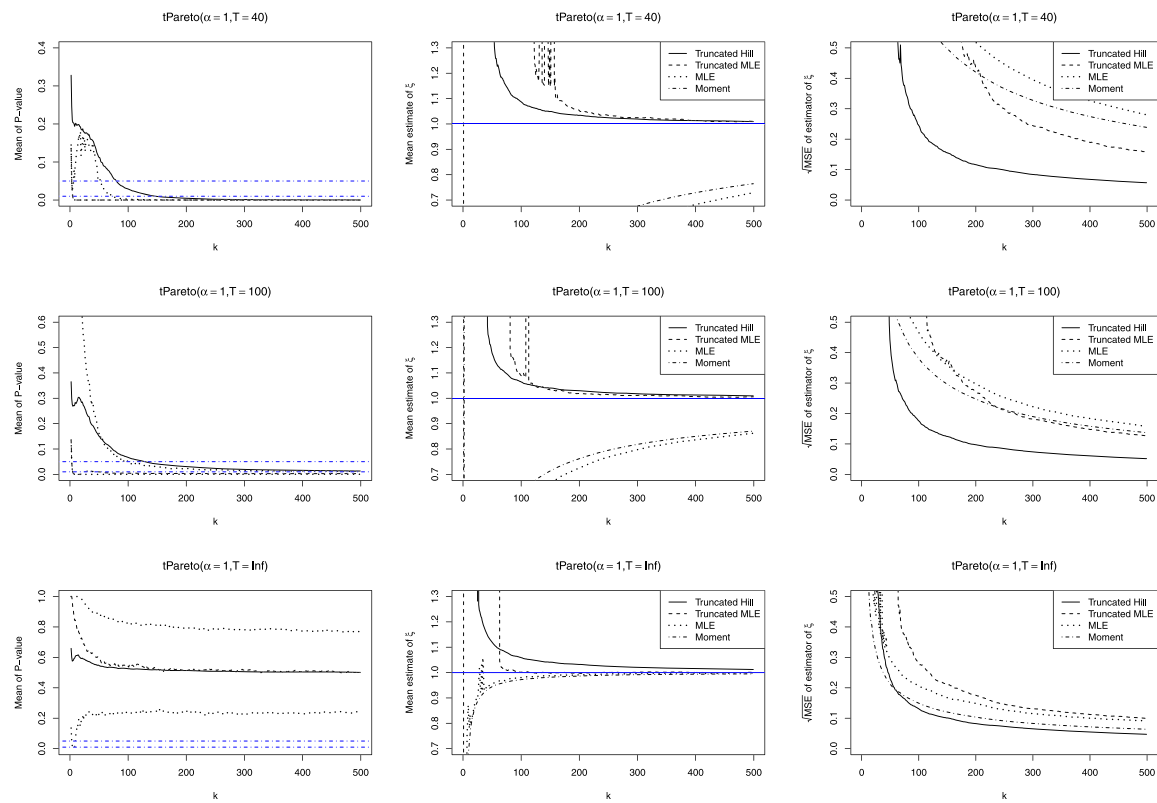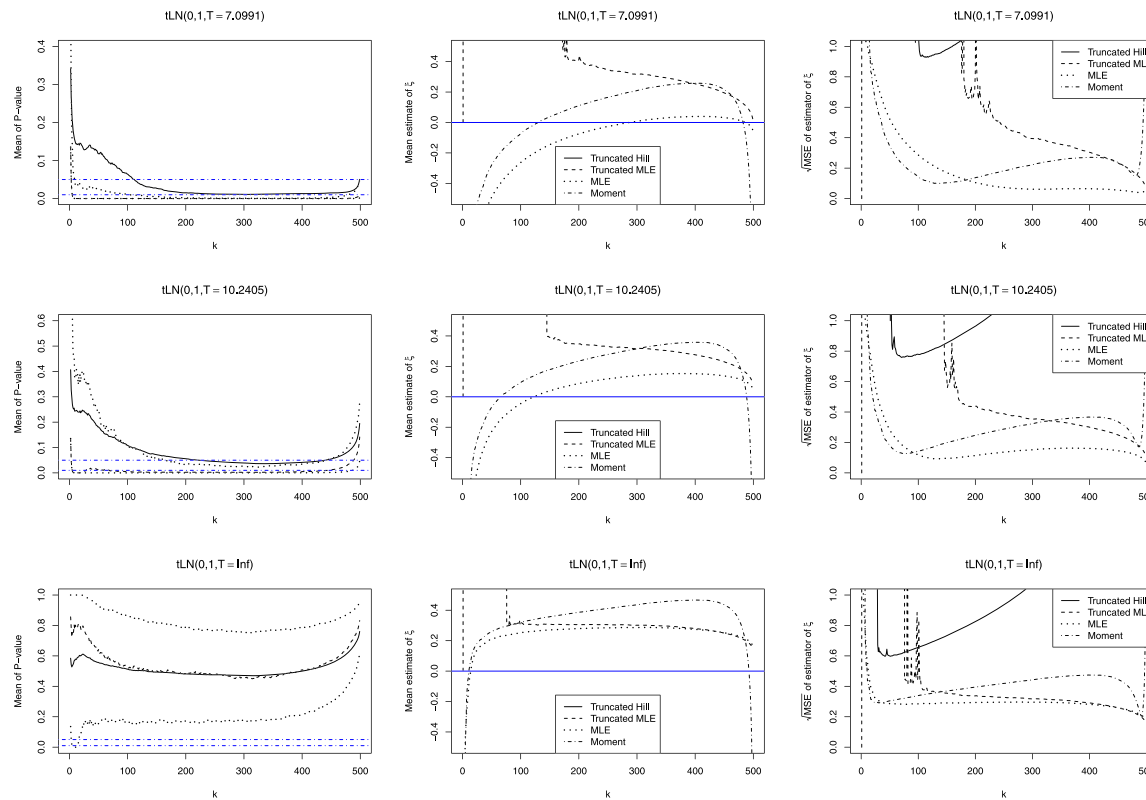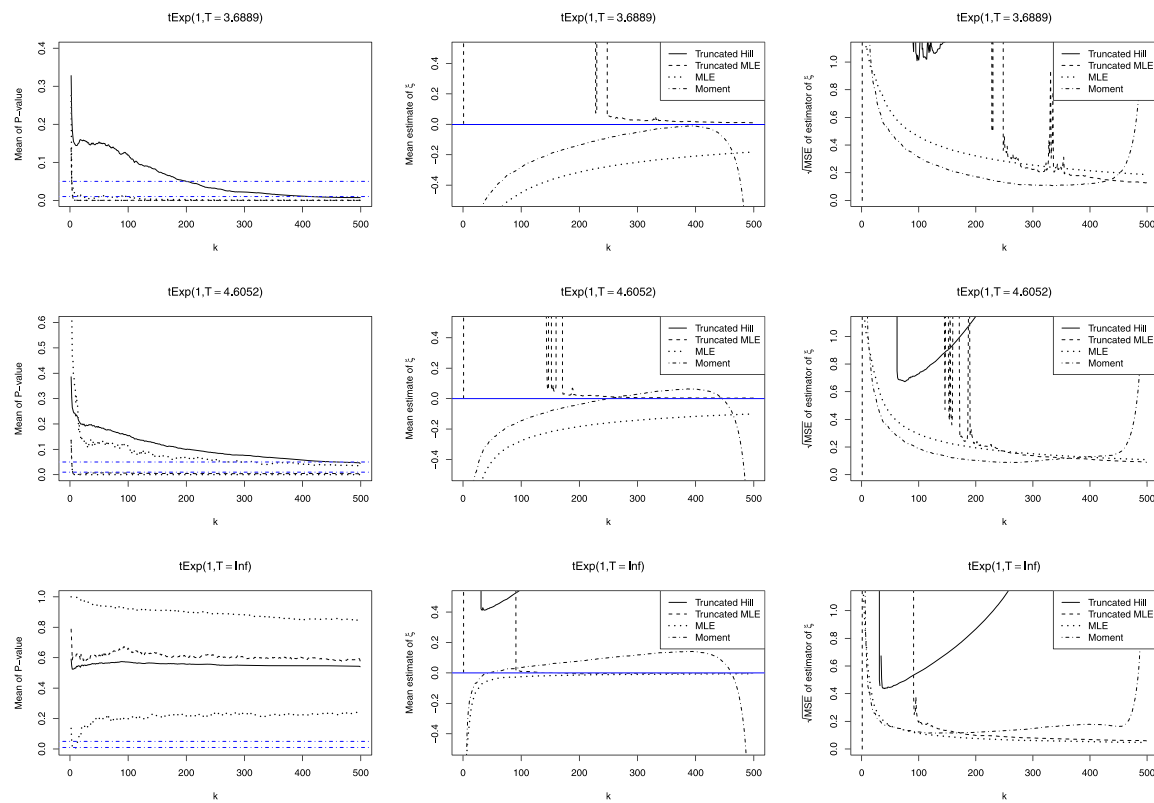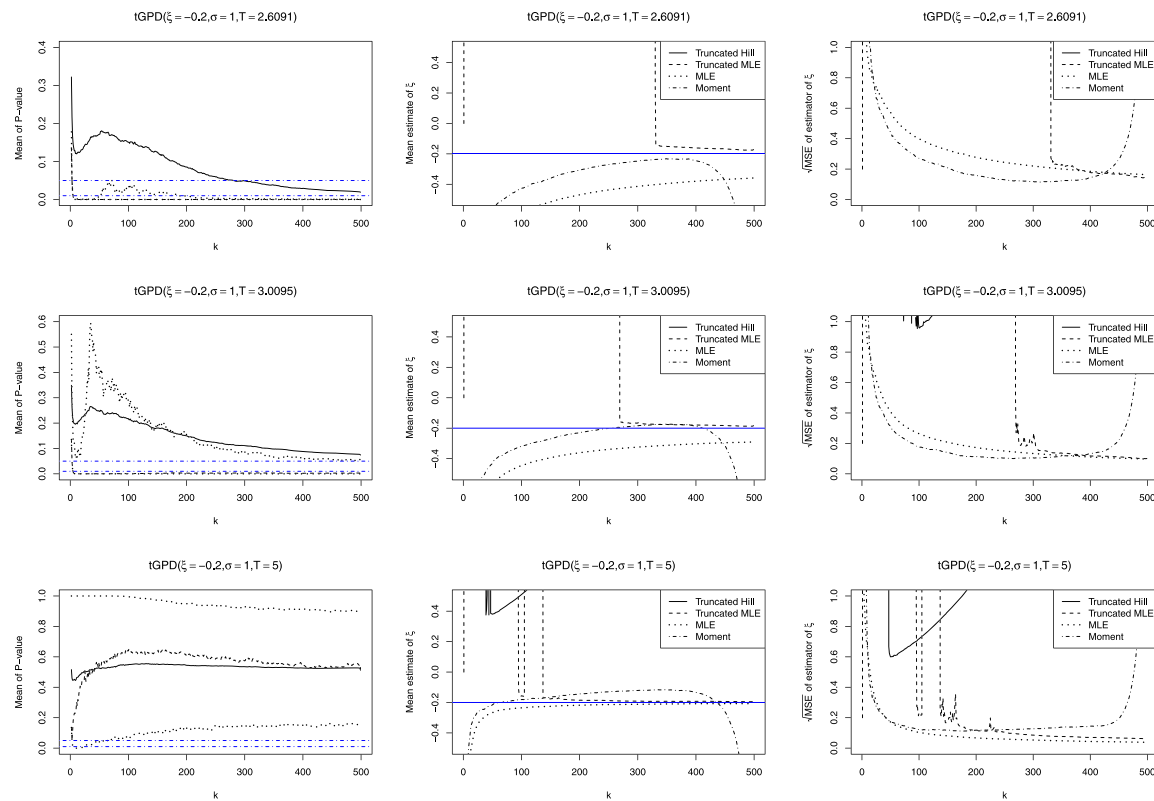
Appendix: Simulation results

FIG 7. *Means and boxplots of P-values for test (left), means (middle) and root MSE (right) of* $\hat{\xi}_k^+$, $\hat{\xi}_k$, $\hat{\xi}_k^\infty$ *and* $\hat{\xi}_k^M$ *from the standard Pareto distribution truncated at* $Q_Y(0.975)$ *(top),* $Q_Y(0.99)$ *(middle) and non truncated (bottom).*

FIG 8. *Means and boxplots of P-values for test (left), means (middle) and root MSE (right) of* $\hat{\xi}_k^+$, $\hat{\xi}_k$, $\hat{\xi}_k^\infty$ *and* $\hat{\xi}_k^M$ *from the standard lognormal distribution truncated at* $Q_Y(0.975)$ *(top),* $Q_Y(0.99)$ *(middle) and non truncated (bottom).*

FIG 9. *Means and boxplots of P-values for test (left), means (middle) and root MSE (right) of $\hat{\xi}_k^+$, $\hat{\xi}_k$, $\hat{\xi}_k^\infty$ and $\hat{\xi}_k^M$ from the standard exponential distribution truncated at $Q_Y(0.975)$ (top), $Q_Y(0.99)$ (middle) and non truncated (bottom).*

FIG 10. *Means and boxplots of P-values for test (left), means (middle) and root MSE (right) of $\hat{\xi}_k^+$, $\hat{\xi}_k$, $\hat{\xi}_k^\infty$ and $\hat{\xi}_k^M$ from GPD(-0.2,1) truncated at $Q_Y(0.975)$ (top), $Q_Y(0.99)$ (middle) and $Q_Y(1)$ (bottom).*
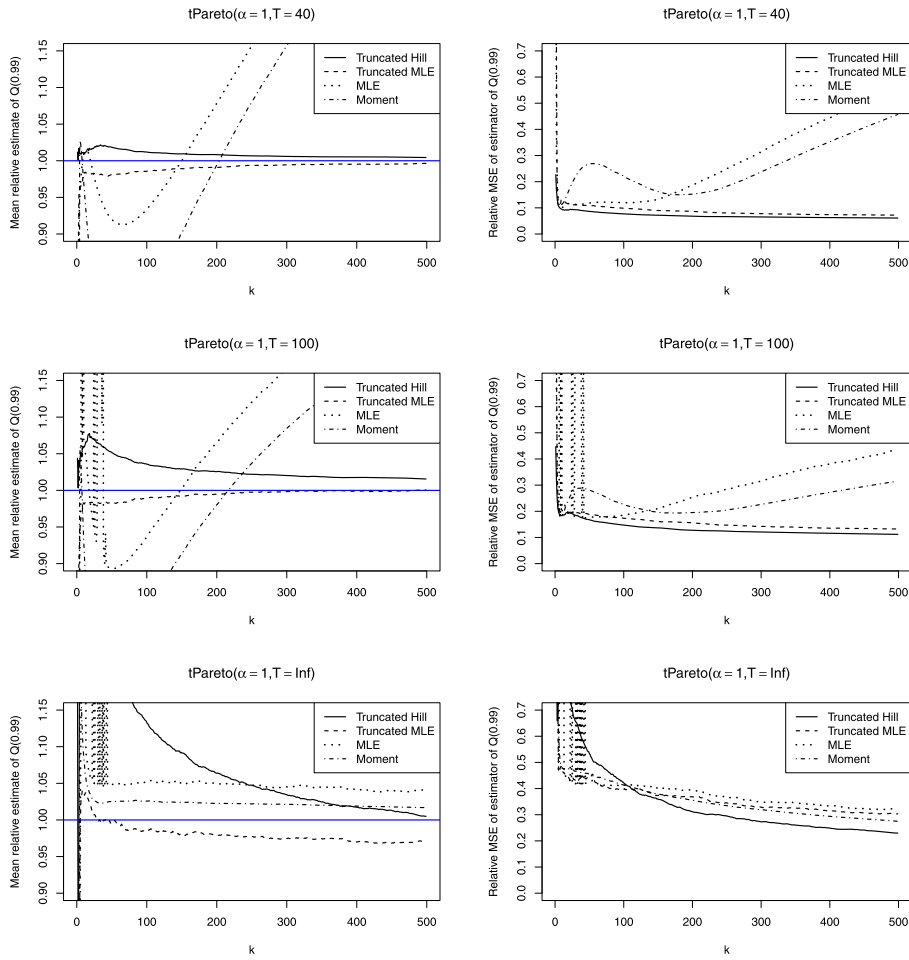
FIG 11. *Mean deviations of $\hat{Q}_{T,k}^{+}(1-p)/Q_T(1-p)$, $\hat{Q}_{T,k}(1-p)/Q_T(1-p)$, $\hat{Q}_k^{\infty}(1-p)/Q_T(1-p)$, $\hat{Q}_k^M(1-p)/Q_T(1-p)$ and corresponding MSE with $p = 0.01$ for the standard Pareto distribution truncated at $Q_Y(0.975)$ (top), $Q_Y(0.99)$ (middle) and non truncated (bottom).*
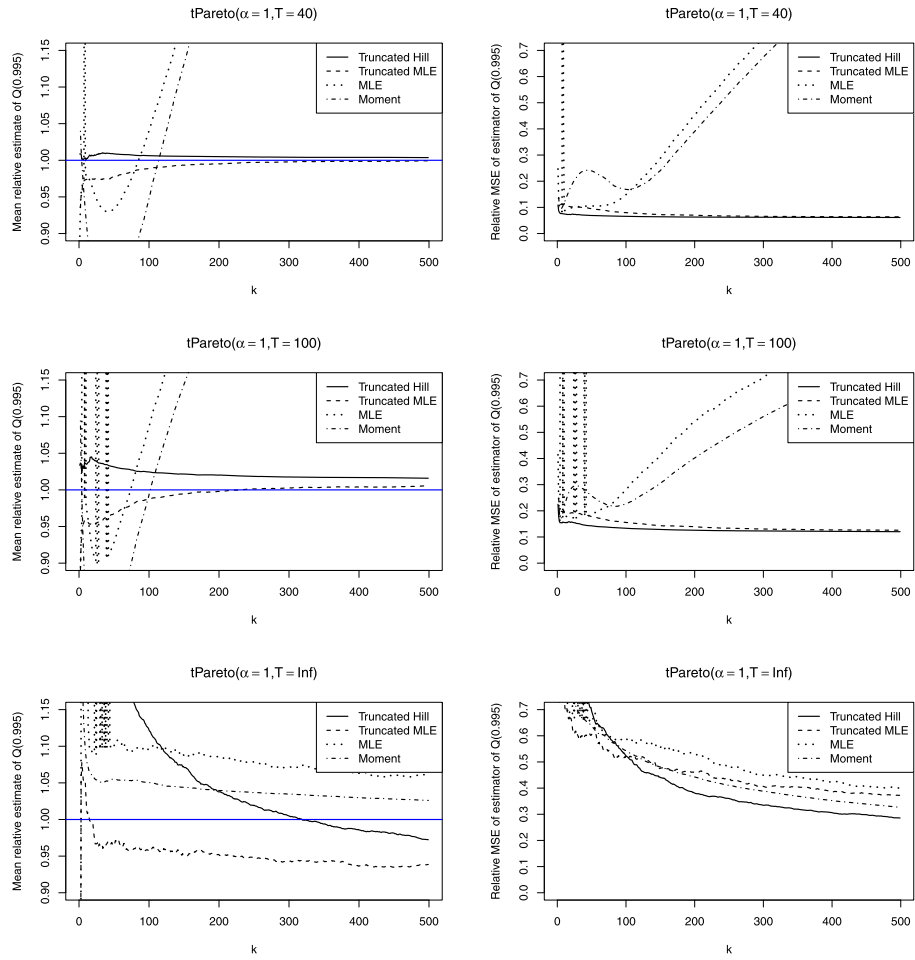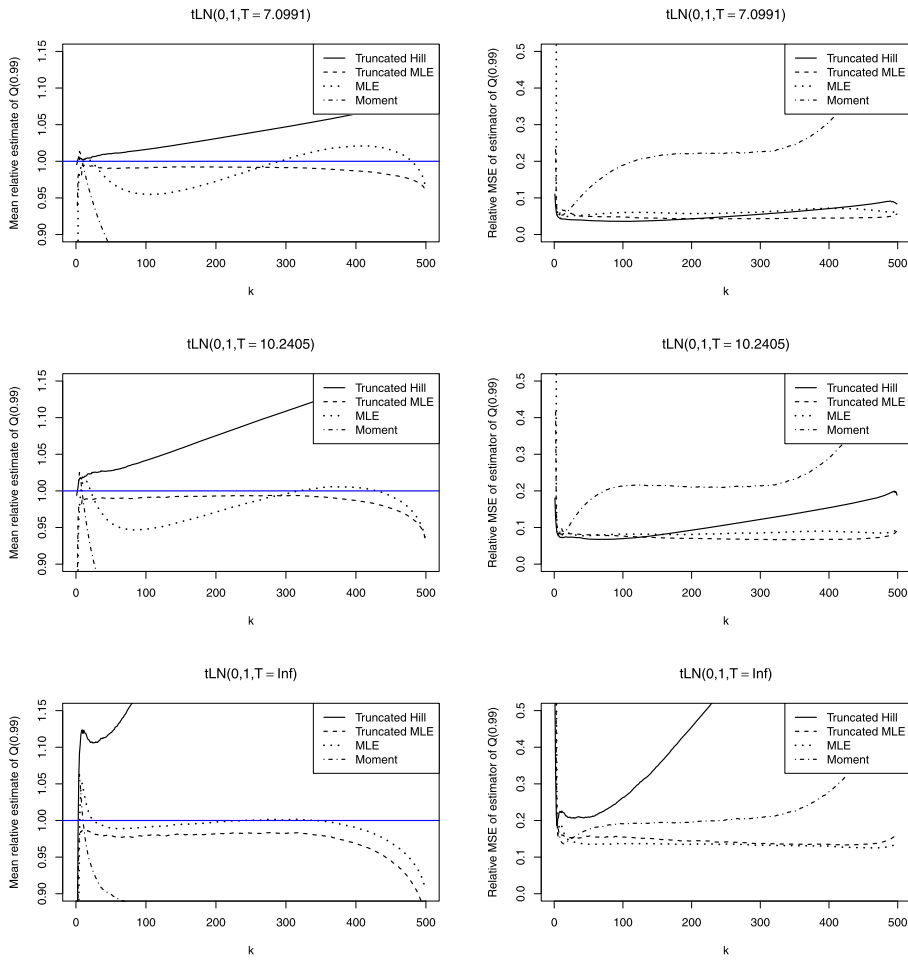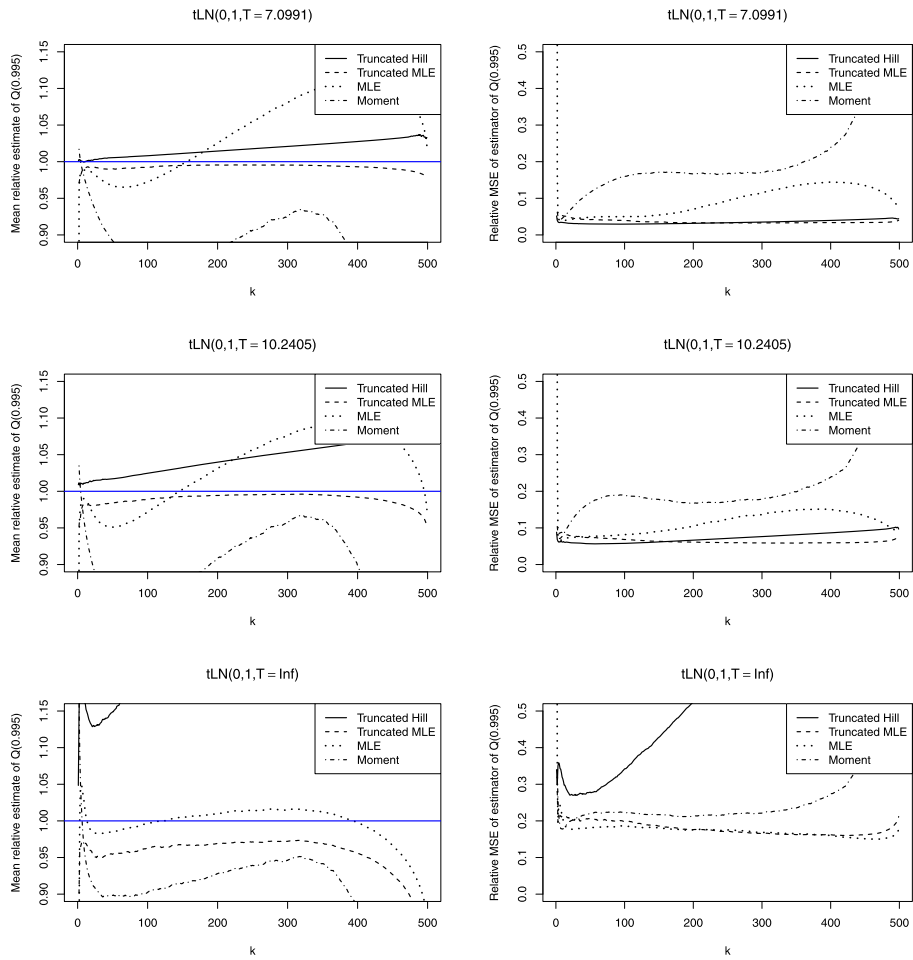
FIG 12. *Mean deviations of $\hat{Q}^{+}_{T,k}(1-p)/Q_T(1-p)$, $\hat{Q}_{T,k}(1-p)/Q_T(1-p)$, $\hat{Q}^{\infty}_k(1-p)/Q_T(1-p)$, $\hat{Q}^M_k(1-p)/Q_T(1-p)$ and corresponding MSE with $p = 0.005$ for the standard Pareto distribution truncated at $Q_Y(0.975)$ (top), $Q_Y(0.99)$ (middle) and non truncated (bottom).*

FIG 13. *Mean deviations of $\hat{Q}^{+}_{T,k}(1-p)/Q_T(1-p)$, $\hat{Q}_{T,k}(1-p)/Q_T(1-p)$, $\hat{Q}^{\infty}_k(1-p)/Q_T(1-p)$, $\hat{Q}^{M}_k(1-p)/Q_T(1-p)$ and corresponding MSE with $p = 0.01$ for the standard lognormal distribution truncated at $Q_Y(0.975)$ (top), $Q_Y(0.99)$ (middle) and non truncated (bottom).*
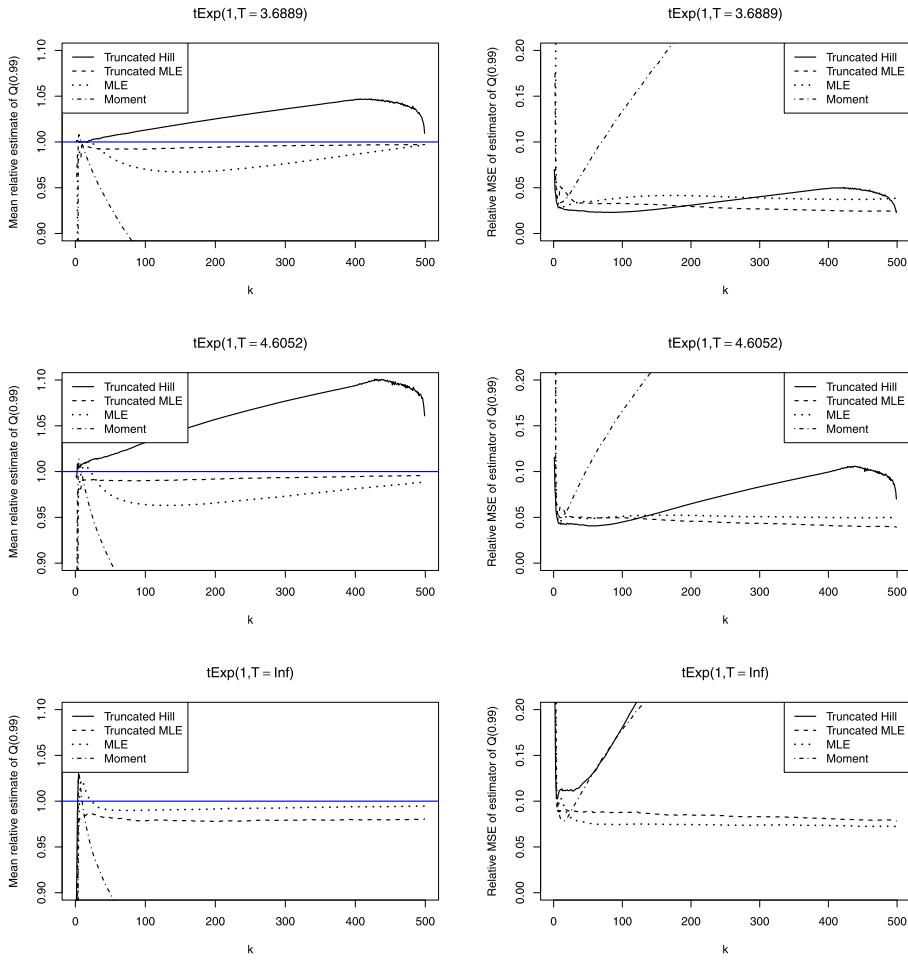
FIG 14. *Mean deviations of $\hat{Q}_{T,k}^{+}(1-p)/Q_T(1-p)$, $\hat{Q}_{T,k}(1-p)/Q_T(1-p)$, $\hat{Q}_k^{\infty}(1-p)/Q_T(1-p)$, $\hat{Q}_k^M(1-p)/Q_T(1-p)$ and corresponding MSE with $p = 0.005$ for the standard lognormal distribution truncated at $Q_Y(0.975)$ (top), $Q_Y(0.99)$ (middle) and non truncated (bottom).*
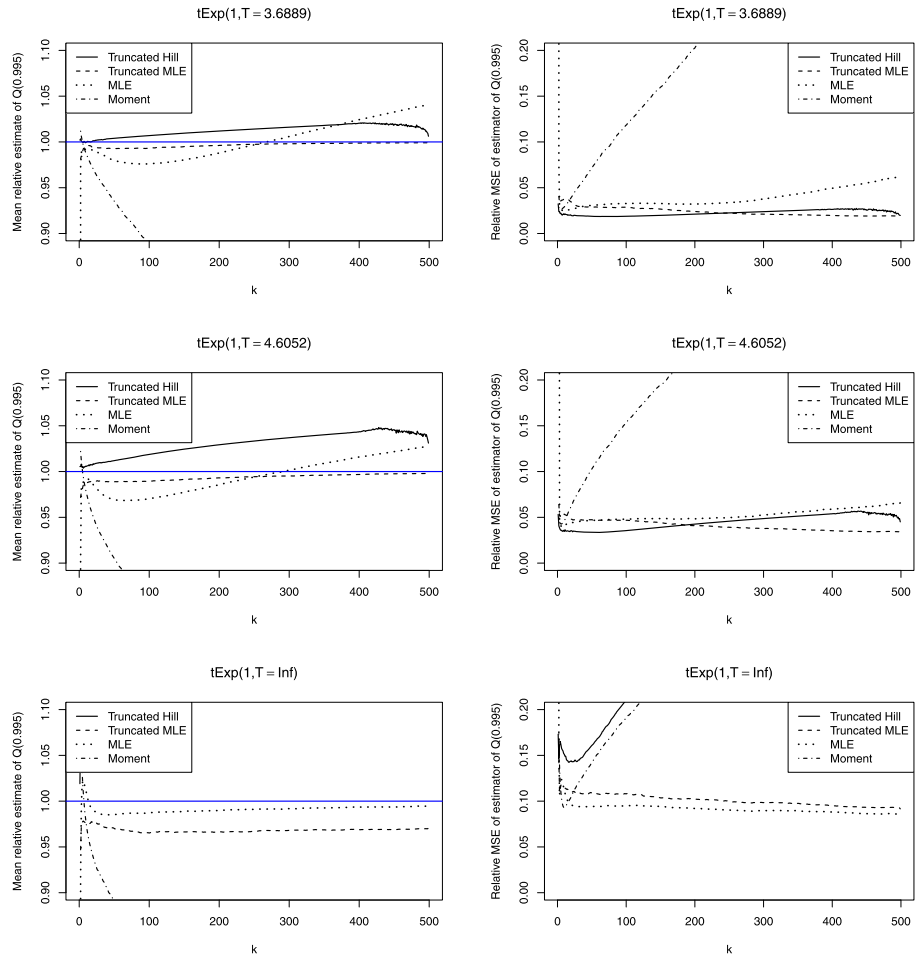
FIG 15. *Mean deviations of* $\hat{Q}_{T,k}^{+}(1-p)/Q_T(1-p)$, $\hat{Q}_{T,k}(1-p)/Q_T(1-p)$, $\hat{Q}_{k}^{\infty}(1-p)/Q_T(1-p)$, $\hat{Q}_{k}^{M}(1-p)/Q_T(1-p)$ *and corresponding MSE with* $p = 0.01$ *for the standard exponential distribution truncated at* $Q_Y(0.975)$ *(top),* $Q_Y(0.99)$ *(middle) and non truncated (bottom).*

FIG 16. *Mean deviations of $\hat{Q}_{T,k}^{+}(1-p)/Q_T(1-p)$, $\hat{Q}_{T,k}(1-p)/Q_T(1-p)$, $\hat{Q}_k^{\infty}(1-p)/Q_T(1-p)$, $\hat{Q}_k^{M}(1-p)/Q_T(1-p)$ and corresponding MSE with $p = 0.005$ for the standard exponential distribution truncated at $Q_Y(0.975)$ (top), $Q_Y(0.99)$ (middle) and non truncated (bottom).*
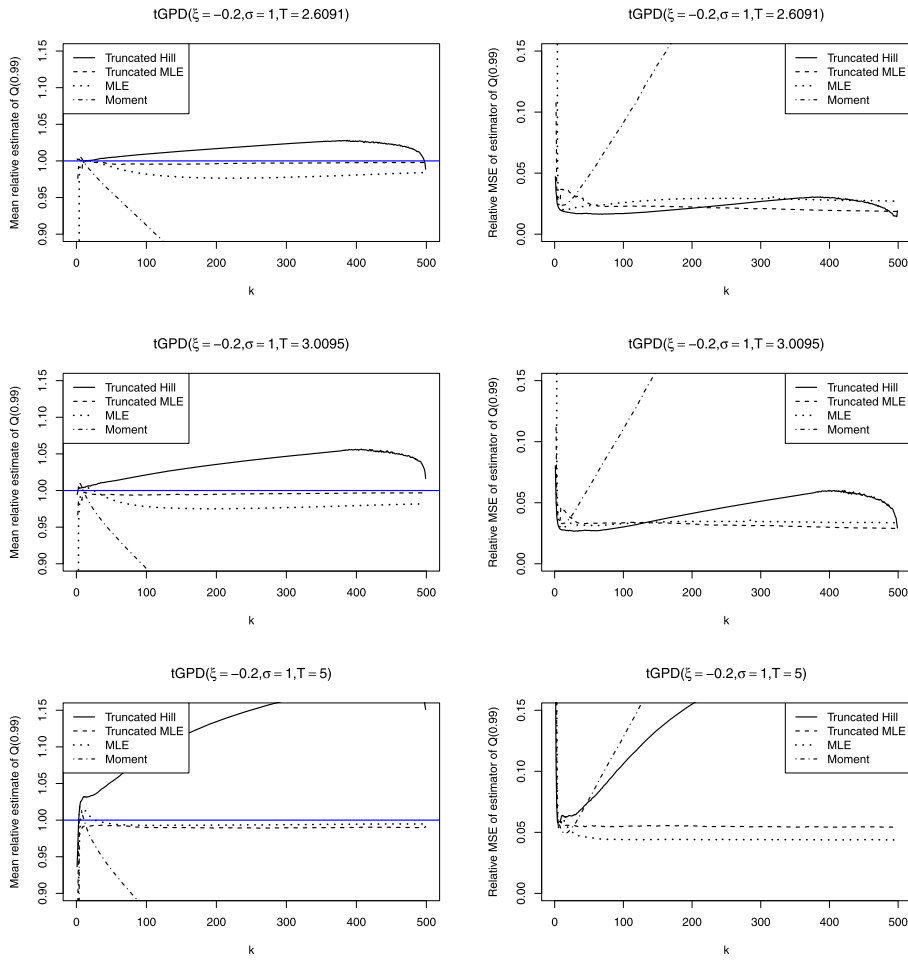
FIG 17. *Mean deviations of $\hat{Q}_{T,k}^{+}(1-p)/Q_T(1-p)$, $\hat{Q}_{T,k}(1-p)/Q_T(1-p)$, $\hat{Q}_k^{\infty}(1-p)/Q_T(1-p)$, $\hat{Q}_k^M(1-p)/Q_T(1-p)$ and corresponding MSE with $p = 0.01$ for GPD(-0.2,1) truncated at $Q_Y(0.975)$ (top), $Q_Y(0.99)$ (middle) and $Q_Y(1)$ (bottom).*
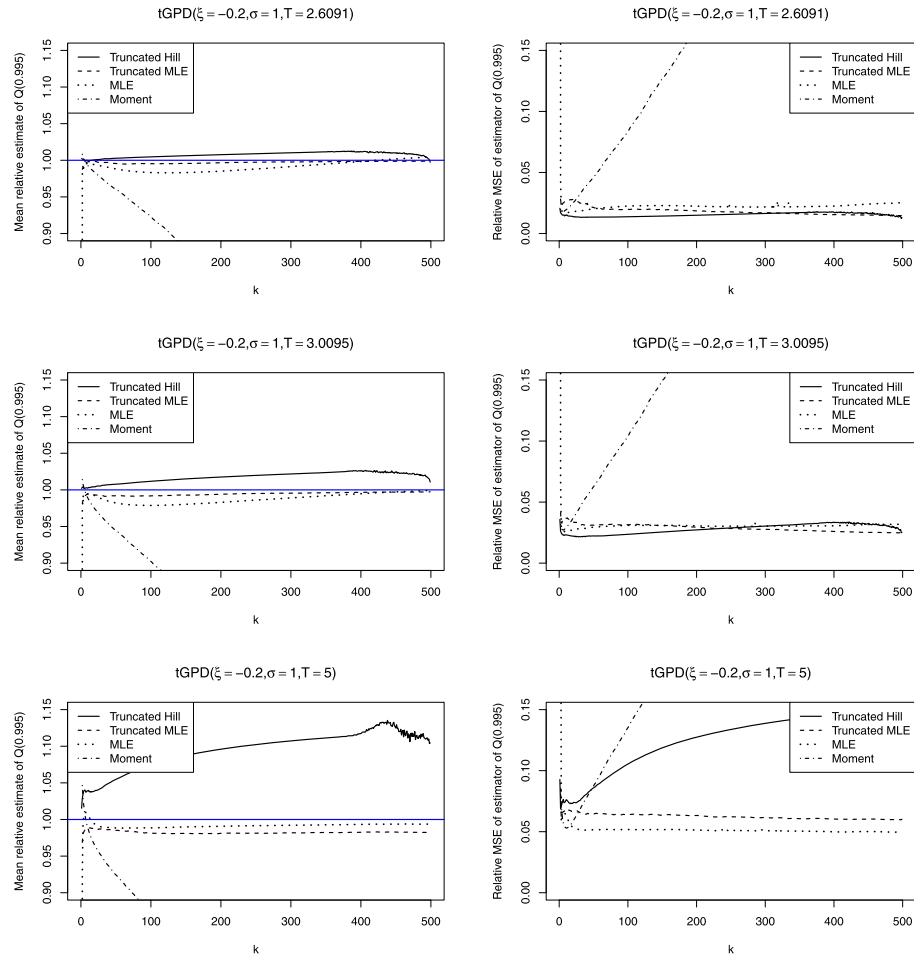
FIG 18. *Mean deviations of $\hat{Q}^+_{T,k}(1-p)/Q_T(1-p)$, $\hat{Q}_{T,k}(1-p)/Q_T(1-p)$, $\hat{Q}^\infty_k(1-p)/Q_T(1-p)$, $\hat{Q}^M_k(1-p)/Q_T(1-p)$ and corresponding MSE with $p = 0.005$ for GPD(-0.2,1) truncated at $Q_Y(0.975)$ (top), $Q_Y(0.99)$ (middle) and $Q_Y(1)$ (bottom).*

Acknowledgements

References

Aban, I. B., Meerschaert, M. M. and Panorska, A. K. (2006). Parameter Estimation for the Truncated Pareto Distribution. *J. Amer. Statist. Assoc.* **101** 270–277. MR2268044

Beirlant, J., Fraga Alves, M. I. and Gomes, M. I. (2016). Tail Fitting for Truncated and Non-truncated Pareto-type Distributions. *Extremes* **19** 429–462. MR3535961

Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes: Theory and Applications.* John Wiley & Sons, Ltd. MR2108013

Chakrabarty, A. and Samorodnitsky, G. (2012). Understanding Heavy Tails in a Bounded World, or, is a Truncated Heavy Tail Heavy or not? *Stoch. Models* **28** 109–143. MR2889004

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer-Verlag. MR1932132

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: an Introduction.* Springer Science and Business Media, LLC. MR2234156

Dekkers, A., Einmahl, J. and de Haan, L. (1989). A Moment Estimator for the Index of an Extreme-value Distribution. *Ann. Statist.* **17** 1795–1832. MR1026315

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag. MR1458613

Gutenberg, B. and Richter, C. F. (1956). Earthquake Magnitude, Intensity, Energy and Acceleration. *Bull. Seismol. Soc. Am.* **46** 105–145.

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *Ann. Statist.* **3** 1163–1174. MR0378204

Kijko, A. and Singh, M. (2011). Statistical Tools for Maximum Possible Earthquake Estimation. *Acta Geophys.* **59** 674–700.

Verster, A., de Waal, D., Schall, R. and Prins, C. (2012). A Truncated Pareto Model to Estimate the Under Recovery of Large Diamonds. *Math. Geosci.* **44** 91–100.

Willems, P. (2009). A Time Series Tool to Support the Multi-criteria Performance Evaluation of Rainfall-runoff Models. *Environ. Model. Softw.* **24** 311–321.