Extreme Value Analysis Without the Largest Values: What Can Be Done?

Jingjing Zou *
Department of Statistics, Columbia University,
Richard A. Davis
Department of Statistics, Columbia University
and
Gennady Samorodnitsky
School of Operations Research and Information Engineering, Cornell University

April 6, 2017

Abstract

In this paper we are concerned with the analysis of heavy-tailed data when a portion of the extreme values are unavailable. This research was motivated by an analysis of the degree distributions in a large social network. The degree distributions of such networks tend to have power law behavior in the tails. We focus on the Hill estimator, which plays a starring role in heavy-tailed modeling. The Hill estimator for this data exhibited a smooth and increasing "sample path" as a function of the number of upper order statistics used in constructing the estimator. This behavior became more apparent as we artificially removed more of the upper order statistics. Building on this observation, we introduce a new parameterization into the Hill estimator that is a function of δ and θ, that correspond, respectively, to the proportion of extreme values that are unavailable and the proportion of upper order statistics used in the estimation. As a function of (δ, θ), we establish functional convergence of the

*The authors would like to thank Zhi-Li Zhang for providing the Google+ data. This research is funded by ARO MURI grant W911NF-12-1-0385.

normalized Hill estimator to a Gaussian random field. An estimation procedure is developed based on the limit theory to estimate the number of missing extremes and extreme value parameters including the tail index and the bias of Hill's estimate. We illustrate how this approach works in both simulations and real data examples.

Keywords: Hill estimator; Heavy-tailed distributions; Missing extremes; Functional convergence

1 Introduction

In studying data exhibiting heavy-tailed behavior, a widely used model is the family of distributions that are regular varying. A distribution F is regular varying if

$$\frac{\bar{F}(tx)}{\bar{F}(t)} \to x^{-\alpha} \tag{1}$$

as $t \to \infty$ for all $x > 0$, where $\alpha > 0$ and $\bar{F}(t) = 1 - F(t)$ is the survival function. The parameter α is called the tail index or the extreme value index, and it controls the heaviness of the tail of the distribution. This is perhaps the most important parameter in extreme value theory and a great deal of research has been devoted to its estimation. The most used and studied estimate of α is based on the Hill estimator for its reciprocal $\gamma = 1/\alpha$ (see Hill 1975, Drees et al. 2000 and de Haan and Ferreira 2006 for further discussion on this estimator). The Hill estimator is defined by

$$H_n(k) = \frac{1}{k} \sum_{i=1}^{k} \log X_{(n-i+1)} - \log X_{(n-k)},$$

where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are the order statistics of the sample $X_1, X_2, \ldots, X_n \sim F(x)$. As an illustration, the left panel of Figure 1 shows the Hill plot of 1000 independent and identically distributed (iid) observations from a Pareto distribution with $\gamma = 2$ ($F(x) = 1 - x^{-0.5}$ for $x \geq 1$ and 0 otherwise).

If the largest several observations in the data are removed, the Hill curve behaves very differently. For example, when the 100 largest observations of the previous Pareto sample have been removed, the Hill plot renders a much smoother curve that is generally increasing (see the right panel of Figure 1).

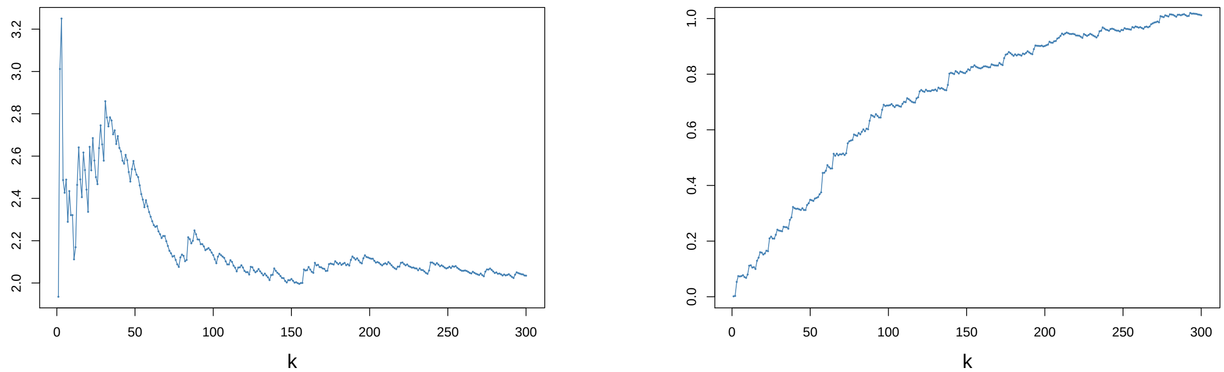

Figure 1: Hill plot of iid Pareto ($\alpha = 0.5$) variables ($n = 1000$). x-axis: number k of upper order statistics used in the calculation. y-axis: $H_n(k)$. Left: without removal. Right: top 100 removed

A similar phenomenon is observed when we study the tail behavior of the in- and out-degrees in a large social network, which in fact is the motivation for this research. We looked at data from a snapshot of Google+, the social network owned and operated by Google, taken on October 19, 2012. The data contain 76,438,791 nodes (registered users) and 1,442,504,499 edges (directed connections). The in-degree of each user is the number of other users following the user and the out-degree is the number of others followed by the user. The degree distributions in natural and social networks are often heavy-tailed (see Newman 2010). The resulting Hill plot for the in-degrees of the Google+ data (the first plot in Figure 2) resembles the curve of the Hill plot for the Pareto observations with the largest extremes removed. This raises the question of whether some extreme in-degrees of the Google+ data are also unobserved. For example, some users with extremely large in-degrees may have been excluded from the data. This pattern of a smooth curve becomes even more pronounced when we apply an additional removal of the top 500 and 1000 values of the in-degree (the second and the third plots in Figure 2).

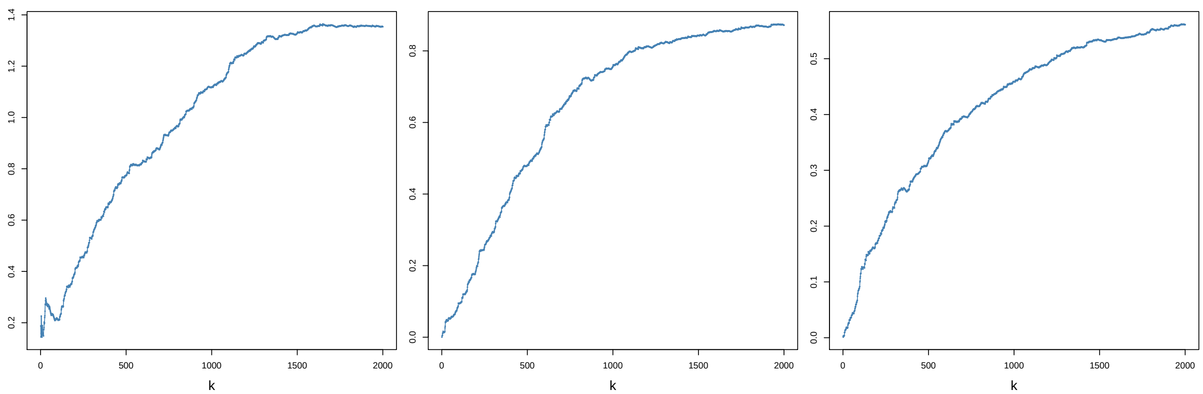

Figure 2: Hill plots of in-degrees of the Google+ network. Left: without removal. Middle: 500 largest values removed. Right: 1000 largest values removed

In order to understand the behavior of the Hill curves of samples in which some of the top extreme values have been removed, we introduce a new parametrization to the Hill estimator. Specifically, we define the Hill estimator without the extremes (HEWE) as a function of parameters δ and θ, which are, respectively, the proportion of the extreme values that are unavailable and the proportion of upper order statistics used in the estimation. This new parametrization allows one to examine the missing of extreme values both visually and theoretically. The Hill estimator curve of the data without the top extremes exhibits a strikingly smooth and increasing pattern, in contrast to the fluctuating shapes when no extremes are missing. And the differences in the shape of the curves are explained by the functional properties of the limiting process of the HEWE. Under a second-order regular varying condition, we show that the HEWE, suitably normalized, converges in distribution to a continuous Gaussian random field with mean zero and covariance depending on δ and parameters of the distribution F including the tail index α.

Based on the likelihood function of the limiting random field, an estimation procedure is developed for δ and the parameters of the distribution, in particular, the tail index α. The proposed approach may also have value in assessing the fidelity of the data to the heavy-tailed assumptions. Specifically, one would expect consistency of the estimation of the tail index when more extremes are artificially removed from the data.

There have been recent works (Aban et al. 2006, Beirlant et al. 2016a,b) that involve adapting classical extreme value theory to the case of truncated Pareto distributions. The truncation is modeled via an unknown threshold parameter and the probability of an observation exceeding the threshold is zero. Maximum likelihood estimators (MLE) are derived for the threshold and the tail index.

Our focus here is to study the path behavior of the HEWE if any arbitrary number of largest values are unavailable. Moreover, the estimation procedure we propose has a built-in mechanism to compensate for the bias introduced by non-Pareto heavy-tailed distributions. Ultimately, the HEWE provides a graphical and theoretical method for estimation and assessment of modeling assumptions. In addition, we feel the proposed approach may shed some useful insight on classical extreme value theory even when extreme values are not missing in the observed data. It is possible to remove a number of top extreme values artificially and study the effect of the artificial removal on the estimation of the tail index. In this case we know the true value of δ.

This paper is organized as follows. Section 2 introduces the HEWE process and states the main result of this paper dealing with the functional convergence of the HEWE to a continuous Gaussian random field. Section 3 explains the details of the estimation procedure based on the asymptotic results. Section 4 demonstrates how our estimation procedure works on simulated data from both Pareto and non-Pareto distributions. Section 5 applies our procedure to several interesting real data sets. All the proofs are postponed to the Appendix.

2 Functional Convergence of HEWE

In this section we set up the framework for studying the reparametrized Hill estimator. To start, let $X_1, X_2, \ldots$ be iid random variables with distribution function F satisfying the regular varying condition (1). Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ denote the order statistics of $X_1, \ldots, X_n$. For integer $k_n \in \{1, \ldots, n\}$, the HEWE process is defined by setting for $\delta \geq 0$ and $\theta > 0$

$$H_n(\delta,\theta) = \begin{cases} \frac{1}{\lfloor \theta k_n \rfloor} \sum_{i=1}^{\lfloor \theta k_n \rfloor} \log X_{(n-\lfloor \delta k_n \rfloor - i+1)} - \log X_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)}, & \theta \geq 1/k_n, \\ 0, & \theta < 1/k_n. \end{cases} \tag{2}$$

The HEWE will play a key role in estimating relevant parameters such as δ and α. To see the idea behind this definition, imagine that the top $\lfloor \delta k_n \rfloor$ observations are not available in the data set and the Hill estimator is computed based on $\lfloor \theta k_n \rfloor$ extreme order statistics of the remaining observations. Viewed as a function of the observable part of the sample, H_n is the usual Hill estimator based on the $\lfloor \theta k_n \rfloor$ upper order statistics. A special case is when $\delta = 0$ and no extreme values are missing, then $H_n(0, \theta)$ corresponds to the usual Hill estimator based on the upper $\lfloor \theta k_n \rfloor$ observations.

In order to obtain the functional convergence of $H_n(\delta, \theta)$, a second-order regular variation condition, which provides a rate of convergence in (1) is needed. This condition can be found, for example, in de Haan and Ferreira (2006), and it states that for $x > 0$,

$$\lim_{t\to\infty} \frac{\frac{\bar{F}(tx)}{\bar{F}(t)} - x^{-\alpha}}{A(\frac{1}{\bar{F}(t)})} = x^{-\alpha} \frac{x^{\alpha\cdot\rho} - 1}{\rho/\alpha}, \tag{3}$$

where $\rho \leq 0$ and A is a positive or negative function with $\lim_{t\to\infty} A(t) = 0$. Assume that the sequence $k_n \to \infty$ used to define H_n satisfies

$$\lim_{n\to\infty} \sqrt{k_n} A(n/k_n) = \lambda, \tag{4}$$

where λ is a finite constant. Note condition (4) implies that $n/k_n \to \infty$.

Distributions that satisfy the second-order condition include the Cauchy, Student's t_ν, stable, Weibull and extreme value distributions (for more discussion on the second-order condition, see, for example, Drees 1998 and Drees et al. 2000). In fact, any distribution with $\bar{F}(x) = c_1 x^{-\alpha} + c_2 x^{-\alpha+\alpha\rho}(1 + o(1))$ as $x \to \infty$, where $c_1 > 0$, $c_2 \neq 0$, $\alpha > 0$ and $\rho < 0$, satisfies the second-order condition with the indicated values of α and ρ (de Haan and Ferreira 2006).

Pareto distributions with tail index $\alpha > 0$ ($\bar{F}(x) = x^{-\alpha}$ for $x \geq 1$ and zero otherwise), however, do not satisfy the second-order condition, as the numerator on the left side of (3) is zero when t is large enough. As will be seen later, the results can be readily extended to the case of Pareto distributions by replacing terms involving ρ with zero.

We now state the main result of this paper which establishes the functional convergence of the HEWE to a Gaussian random field.

Theorem 2.1. Assume the second-order condition (3) holds and (4) is satisfied for a given sequence k_n and λ. Then as $n \to \infty$,

$$\sqrt{k_n}\left(H_n(\cdot,\cdot) - \frac{g(\cdot,\cdot)}{\alpha}\right) - b_\rho(\cdot,\cdot) \xrightarrow{d} \frac{1}{\alpha} G(\cdot,\cdot)$$

in $D([0,\infty) \times (0,\infty))$, where

$$g(\delta,\theta) = \begin{cases} 1, & \delta = 0, \\ 1 - \frac{\delta}{\theta}\log\left(\frac{\theta}{\delta} + 1\right), & \delta > 0, \end{cases}$$

$$b_\rho(\delta,\theta) = \begin{cases} \frac{\lambda}{1-\rho}\frac{1}{\theta^\rho}, & \delta = 0, \\ \frac{1+(\theta/\delta)\rho-(\theta/\delta+1)^\rho}{(\theta/\delta)(1-\rho)\rho}\frac{\lambda}{(\delta+\theta)^\rho}, & \delta > 0, \end{cases}$$

and G is a continuous Gaussian random field with mean zero and the following covariance function. If $\delta_1 \vee \delta_2 > 0$, then

$$
\begin{aligned}
&\operatorname{Cov}\big(G(\delta_1,\theta_1), G(\delta_2,\theta_2)\big) \\
&= \frac{1}{\theta_1\theta_2}\bigg[(\delta_1+\theta_1)\wedge(\delta_2+\theta_2) - (\delta_1\vee\delta_2) \\
&\qquad - (\delta_1+\delta_2)\log\left(\frac{(\delta_1+\theta_1)\wedge(\delta_2+\theta_2)}{\delta_1\vee\delta_2}\right) + \frac{\delta_1\delta_2}{\delta_1\vee\delta_2} - \frac{\delta_1\delta_2}{(\delta_1+\theta_1)\wedge(\delta_2+\theta_2)}\bigg].
\end{aligned}
$$

If $\delta_1 = \delta_2 = 0$,

$$
\operatorname{Cov}\big(G(0,\theta_1), G(0,\theta_2)\big) = \frac{1}{\theta_1 \vee \theta_2}.
$$

Remark. For fixed θ, the functions g and b_ρ are continuous at $\delta = 0$. For iid Pareto variables $X_1, X_2, \ldots$ with tail index $\alpha > 0$, the result of Theorem 2.1 still holds with the bias term b_ρ replaced by zero.

It is demonstrated next that the parameters, especially α and δ, are identifiable via the path of the reparametrized Hill estimator. Figure 3 shows the Hill estimates of the same sample from the Pareto distribution with $\alpha = 0.5$ as in Figure 1. We choose $k_n = 100$ and $\delta = 1$ so that the top 100 observations are removed from the original sample. In the left panel of Figure 3, the Hill estimates are overlaid with the mean curves of the Gaussian random field $g(\delta,\theta)/\alpha$ with different values of δ while fixing the true value of $\alpha = 0.5$. The right panel of Figure 3 shows the mean curves with different values of α while fixing the true value $\delta = 1$. In both plots, the Hill plot is closest to the mean curve corresponding to the true value of the parameter.

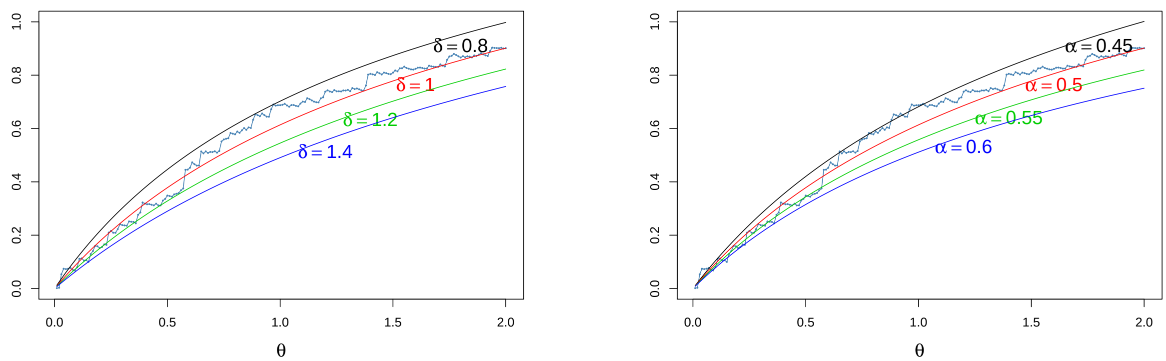

Figure 3: Fitting mean curves with different values of parameters to the Hill plot for the Pareto sample as in Figure 1. Left: fixing $\alpha = 0.5$. Right: fixing $\delta = 1$

In order to demonstrate the variability generated by the limiting Gaussian random field, we compare the Hill plots for samples from Pareto and Cauchy distributions with their Gaussian process approximations given by Theorem 2.1. Figure 4 presents the Hill plots for the same Pareto sample as in Figures 1 and 3, without removal of extremes (left) and with the top 100 observations removed (right), along with 50 independent realizations from the corresponding Gaussian processes with bias $b_\rho \equiv 0$.

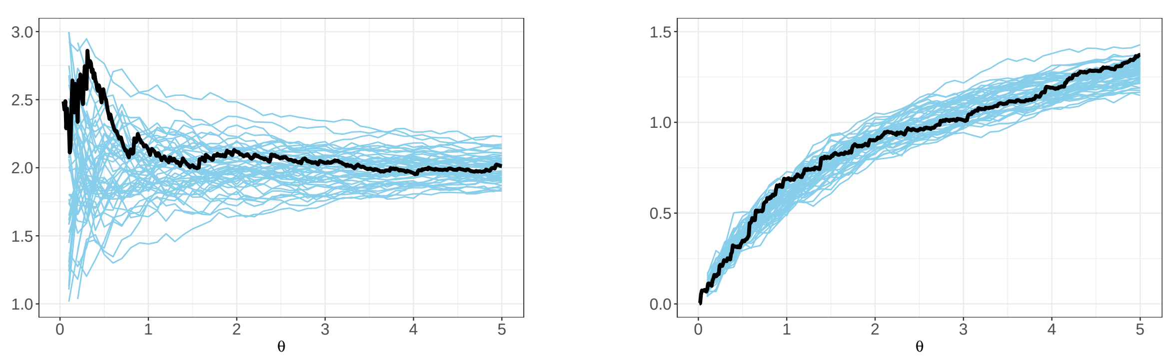

Figure 4: Observed Hill plots for the Pareto sample (bold lines) and realizations from corresponding Gaussian processes (thin lines). Left: with the original sample. Right: top 100 extreme values removed

Figure 5 shows the Hill plots for a Cauchy sample ($n = 1000$, $k_n = 100$, $\alpha = 1$ and $\rho = -2$), without removal of extremes and with the top 100 extremes removed, along with 50 independent realizations from the corresponding Gaussian processes with non-zero b_ρ.

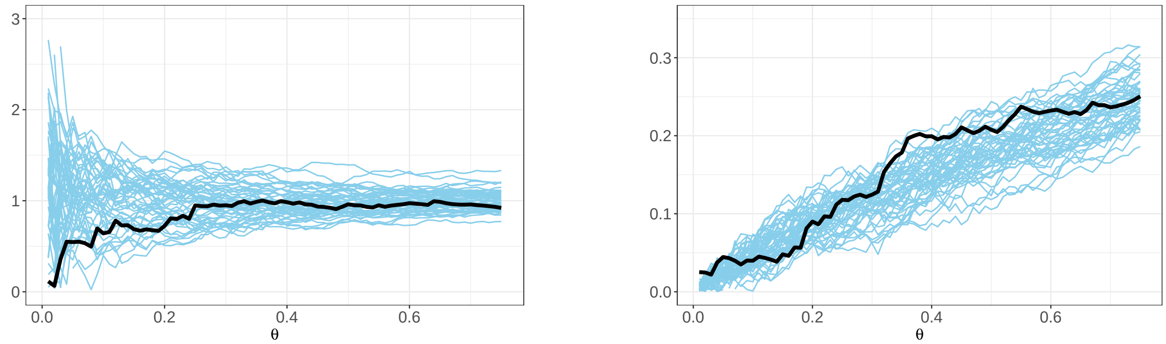

Figure 5: Observed Hill plots for a Cauchy sample (bold lines) and realizations from corresponding Gaussian processes (thin lines). Left: with the original sample. Right: top 100 extreme values removed

3 Parameter Estimation

Let $X_1, X_2, \dots X_n$ be a sample from a distribution F satisfying the second-order regular variation condition (3), and let $X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}$ denote the increasing order statistics of $\{X_i\}$. Suppose the $\lfloor \delta k_n \rfloor$ largest observations are unobserved in the data. In this section, we develop an approximate maximum likelihood estimation procedure for the unknown parameters δ, α and ρ given the observed data. The procedure is based on the asymptotic distribution of the two-parameter Hill estimator $H_n(\delta, \theta)$. When δ is fixed, we use the single-parameter notation $H_n(\theta)$.

By Theorem (2.1), for fixed $(\theta_1, \dots, \theta_s)$ the joint distribution of $(H_n(\theta_1), \dots, H_n(\theta_s))$ can be approximated, when k_n is large, by a distribution with density function at $\mathbf{h} =$

$(h_1, \ldots, h_s)$ given by

$$\frac{1}{\sqrt{(2\pi)^s|\mathbf{\Sigma}_{\alpha,\delta}|}} \exp\left[-\frac{1}{2}\Big(\mathbf{h} - \frac{\mathbf{g}_\delta}{\alpha} - \frac{\mathbf{b}_{\delta,\rho}}{\sqrt{k_n}}\Big)^\top \mathbf{\Sigma}_{\alpha,\delta}^{-1}\Big(\mathbf{h} - \frac{\mathbf{g}_\delta}{\alpha} - \frac{\mathbf{b}_{\delta,\rho}}{\sqrt{k_n}}\Big)\right], \tag{5}$$

where

$$\{\mathbf{g}_\delta\}_i = \begin{cases} 1, & \delta = 0, \\ 1 - \frac{\delta}{\theta_i}\log\big(\frac{\theta_i}{\delta} + 1\big), & \delta > 0, \end{cases}$$

$$\{\mathbf{b}_{\delta,\rho}\}_i = \begin{cases} \frac{\lambda}{1-\rho}\frac{1}{\theta_i^\rho}, & \delta = 0, \\ \frac{1+(\theta_i/\delta)\rho-(\theta_i/\delta+1)^\rho}{(\theta_i/\delta)(1-\rho)\rho}\frac{\lambda}{(\delta+\theta_i)^\rho}, & \delta > 0, \end{cases}$$

and

$$\mathbf{\Sigma}_{\alpha,\delta}(i,j) = \begin{cases} \frac{1}{\alpha^2 k_n}\frac{1}{\theta_i \vee \theta_j}, & \delta = 0, \\ \frac{1}{\alpha^2 k_n}\frac{(\theta_i\wedge\theta_j)^2}{\delta\theta_i\theta_j} v\big(\frac{\theta_i\wedge\theta_j}{\delta}\big), & \delta > 0, \end{cases}$$

with

$$v(\theta) = \frac{1}{\theta} - \frac{2\log(\theta+1)}{\theta^2} + \frac{1}{\theta(\theta+1)}.$$

To simplify the calculation for the maximum likelihood estimator of α, δ and ρ, let

$$T_i = H_n(\theta_i) - \frac{\theta_{i-1}}{\theta_i}H_n(\theta_{i-1}),$$

where $\theta_0 = 0$ is introduced for convenience. Note that the T_i are asymptotically independent with the joint density function at $\mathbf{t} = (t_1, \ldots, t_s)$ being

$$\frac{1}{\sqrt{(2\pi)^s|\tilde{\mathbf{\Sigma}}_{\alpha,\delta}|}} \exp\left[-\frac{1}{2}\big(\mathbf{t} - \mathbf{m}\big)^\top \tilde{\mathbf{\Sigma}}_{\alpha,\delta}^{-1}\big(\mathbf{t} - \mathbf{m}\big)\right], \tag{6}$$

where

$$m_i = \frac{1}{\alpha}\Big(\{\mathbf{g}_\delta\}_i - \frac{\theta_{i-1}}{\theta_i}\{\mathbf{g}_\delta\}_{i-1}\Big) + \frac{1}{\sqrt{k_n}}\Big(\{\mathbf{b}_{\delta,\rho}\}_i - \frac{\theta_{i-1}}{\theta_i}\{\mathbf{b}_{\delta,\rho}\}_{i-1}\Big)$$

and $\tilde{\mathbf{\Sigma}}_{\alpha,\delta}$ is a diagonal matrix, in which

$$\tilde{\mathbf{\Sigma}}_{\alpha,\delta}(i,i) = \begin{cases} \frac{1}{\alpha^2 k_n}\left(\frac{1}{\theta_i} - \frac{\theta_{i-1}}{\theta_i^2}\right), & \delta = 0, \\ \frac{1}{\alpha^2 k_n \delta}\left(v\left(\frac{\theta_i}{\delta}\right) - \left(\frac{\theta_{i-1}}{\theta_i}\right)^2 v\left(\frac{\theta_{i-1}}{\delta}\right)\right), & \delta > 0. \end{cases}$$

The log-likelihood corresponding to the density (6) is

$$C + s\log(\alpha) + \frac{1}{2}\sum_{i=1}^{s}\log(w_i) - \frac{1}{2}\alpha^2 k_n \sum_{i=1}^{s} w_i(t_i - m_i)^2, \tag{7}$$

where C is a constant independent of α, δ and ρ. For $\delta > 0$,

$$w_i = \delta \Big/ \left(v\left(\frac{\theta_i}{\delta}\right) - \left(\frac{\theta_{i-1}}{\theta_i}\right)^2 v\left(\frac{\theta_{i-1}}{\delta}\right)\right).$$

For $\delta = 0$,

$$w_i = 1 \Big/ \left(\frac{1}{\theta_i} - \frac{\theta_{i-1}}{\theta_i^2}\right).$$

For fixed α and δ, the only part of the log-likelihood (7) that needs to be optimized is the weighted sum of squares

$$\sum_{i=1}^{s} w_i(t_i - m_i)^2, \tag{8}$$

and it is minimized over the values of ρ and λ. Note the value of λ depends on the choice of k_n through (4). When k_n is fixed, λ is viewed as an independent nuisance parameter and appears in m_i via

$$\frac{1}{\sqrt{k_n}}\left(\{\mathbf{b}_{\delta,\rho}\}_i - \frac{\theta_i}{\theta_{i-1}}\{\mathbf{b}_{\delta,\rho}\}_{i-1}\right) = \frac{\lambda}{\sqrt{k_n}}\{\mathbf{f}_{\delta,\rho}\}_i,$$

where

$$\{\mathbf{f}_{\delta,\rho}\}_i = \begin{cases} \frac{1}{1-\rho}\frac{1}{\theta_i^\rho} - \frac{\theta_{i-1}}{\theta_i}\frac{1}{1-\rho}\frac{1}{\theta_{i-1}^\rho}, & \delta = 0, \\ \frac{1+(\theta_i/\delta)\rho-(\theta_i/\delta+1)^\rho}{(\theta_i/\delta)(1-\rho)\rho}\frac{1}{(\delta+\theta_i)^\rho} - \frac{\theta_{i-1}}{\theta_i}\frac{1+(\theta_{i-1}/\delta)\rho-(\theta_{i-1}/\delta+1)^\rho}{(\theta_{i-1}/\delta)(1-\rho)\rho}\frac{1}{(\delta+\theta_{i-1})^\rho}, & \delta > 0. \end{cases}$$

Minimizing (8) over λ and ρ results in

$$\hat{\rho}_{\alpha,\delta} = \arg\min_{\rho \le 0} \sum_{i=1}^{s} w_i \left(t_i - \frac{1}{\alpha} \Big(\{\mathbf{g}_\delta\}_i - \frac{\theta_{i-1}}{\theta_i} \{\mathbf{g}_\delta\}_{i-1} \Big) - \frac{\hat{\lambda}_{\alpha,\delta,\rho}}{\sqrt{k_n}} \{\mathbf{f}_{\delta,\rho}\}_i \right)^2,$$

where

$$\hat{\lambda}_{\alpha,\delta,\rho} = \sqrt{k_n}\, \frac{\sum_{i=1}^{s} w_i \big(t_i - (\{\mathbf{g}_\delta\}_i - \frac{\theta_{i-1}}{\theta_i}\{\mathbf{g}_\delta\}_{i-1})/\alpha\big)\{\mathbf{f}_{\delta,\rho}\}_i}{\sum_{i=1}^{s} w_i \{\mathbf{f}_{\delta,\rho}\}_i^2}.$$

Note that this estimation approach, in which λ is viewed as a nuisance parameter, adjusts for the choice of k_n automatically. If a different k_n is selected, the estimate of λ will adapt to reflect this change.

Once we have found the optimal values of ρ and λ, we optimize the resulting expression in (7) by examining its values on a fine grid of (α, δ). Alternatively, an iterative procedure can be used, where in each step one of α, δ, ρ is updated given values of the other two parameters until convergence of the log-likelihood function.

4 Simulation Studies

In this section we test our procedure on simulated data. In each of the following simulations, we generate 200 independent samples of size n from a regular-varying distribution function with tail index α. Given a k_n, we remove the largest $\lfloor \delta k_n \rfloor$ observations from each of the original samples and apply the proposed method to the samples after the removal.

For comparison, we also apply the method in Beirlant et al. (2016a) to the same samples. In Beirlant et al. (2016a), α and the threshold T over which the observations are discarded are estimated with the MLE based on the truncated Pareto distribution. The odds ratio of the truncated observations under the un-truncated Pareto distribution is estimated by solving an equation involving the estimates of α and T. Finally, the number of truncated observations is calculated given the odds ratio and the observed sample size.

For each combination of distribution and parameters, we start from $\theta_1 = 5/k_n$ and let $\theta_i = \theta_{i-1} + 1/k_n$ for $1 < i \leq s$. We consider a sequence of different endpoints $\theta_s k_n$ to examine the influence of the range of order statistics included in the estimation. For each value of θ_s, we solve for the estimates of α and δ based on the asymptotic density of $(H_n(\theta_1), \ldots, H_n(\theta_s))$ following the procedure described in Section 3.

Simulations from both Pareto and non-Pareto distributions show that the proposed method provides reliable estimates of the tail index and performs particularly well in estimating the number of missing extremes. The advantages of the proposed method become more apparent in dealing with non-Pareto samples.

4.1 Pareto Samples

First we examine Pareto samples with $n = 500$ and $\alpha = 0.5$. Let $k_n = 50$ and $\delta = 1$ so that $\delta k_n = 50$ top extreme observations are removed from the original data. Figures 6 and 7 show the averaged estimates of α and δk_n as well as the estimated mean squared errors (MSE) with different $\theta_s k_n$. Estimates by the proposed method are plotted in solid lines while those by the method in Beirlant et al. (2016a) are in dashed lines. The proposed method overestimates the tail index α, especially when the number of upper order statistics included in the estimation is small. This is not unexpected, as the method does not assume the data are from a Pareto distribution and thus does not benefit from the extra information that the bias term in the likelihood should be zero. However, the proposed method estimates the number of missing extreme values accurately, and the estimation is robust to different numbers of upper order statistics included.

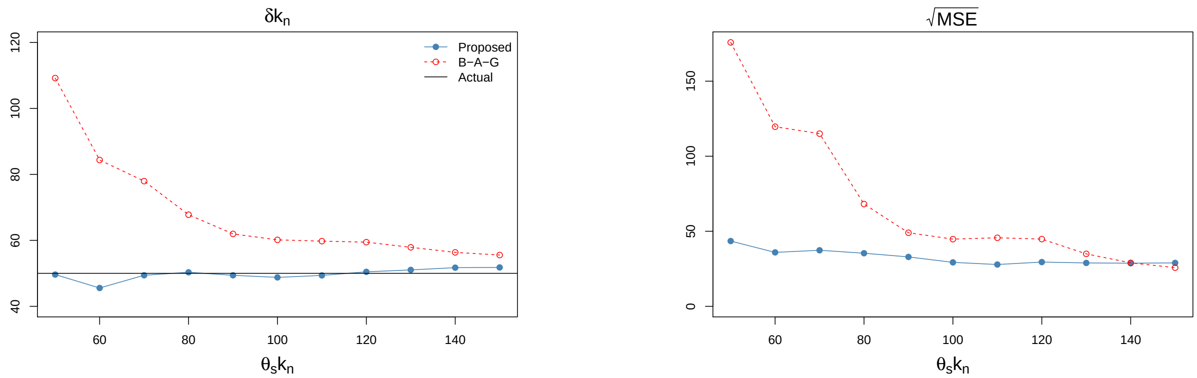

Figure 6: Estimated number of missing extremes and $\sqrt{\text{MSE}}$ for Pareto samples. $n = 500$, $\alpha = 0.5$, $k_n = 50$, $\delta = 1$

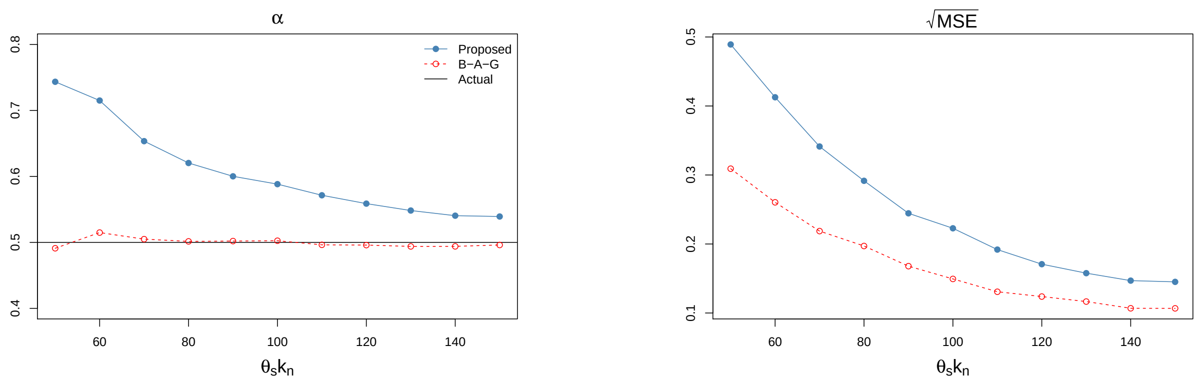

Figure 7: Estimated tail index and $\sqrt{\text{MSE}}$ for Pareto samples. $n = 500$, $\alpha = 0.5$, $k_n = 50$, $\delta = 1$

We also examine the efficacy of the estimation procedure for 200 independent Pareto samples without any extreme values missing ($\delta = 0$). Figure 8 shows that both methods give accurate estimates of the tail index and are able to estimate the number of missing extremes to be close to zero.

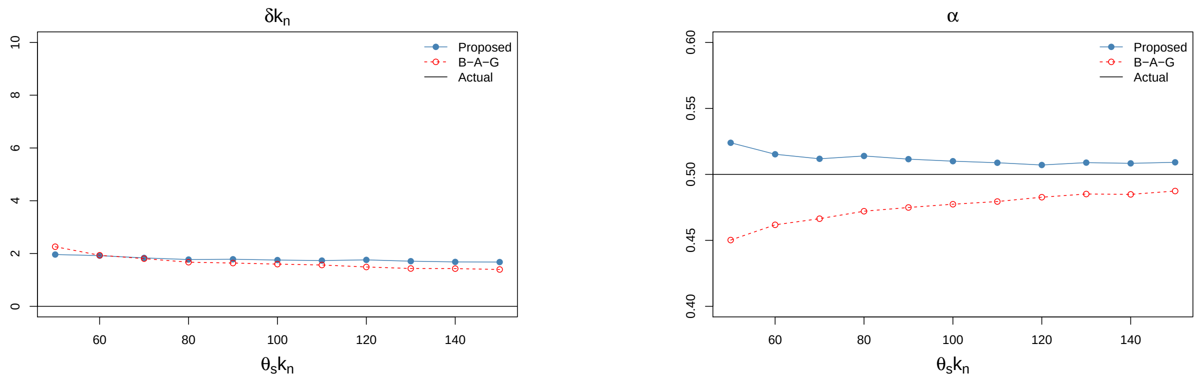

Figure 8: Estimated number of missing extremes and tail index for Pareto samples. $n = 500$, $\alpha = 0.5$, $k_n = 50$, $\delta = 0$

4.2 Non-Pareto Samples

Next we examine the scenarios when the data are not from Pareto distributions. Observations used here are generated from Cauchy and Student's t-distributions. The following results show that the proposed method continues to perform well in estimating the number of missing extremes, even for distributions whose tail indices are more challenging to estimate when the top extremes are unobserved.

4.2.1 Cauchy Samples

Figures 9 and 10 show averaged estimates for 200 independent Cauchy samples with the largest 100 observations removed from each sample.

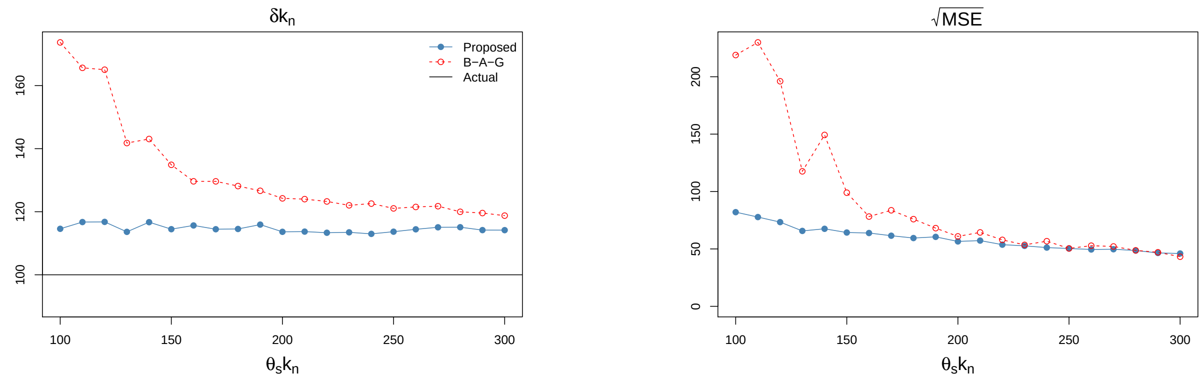

Figure 9: Estimated number of missing extremes and $\sqrt{\text{MSE}}$ for Cauchy samples. $n = 2000$, $\alpha = 1$, $k_n = 100$, $\delta = 1$

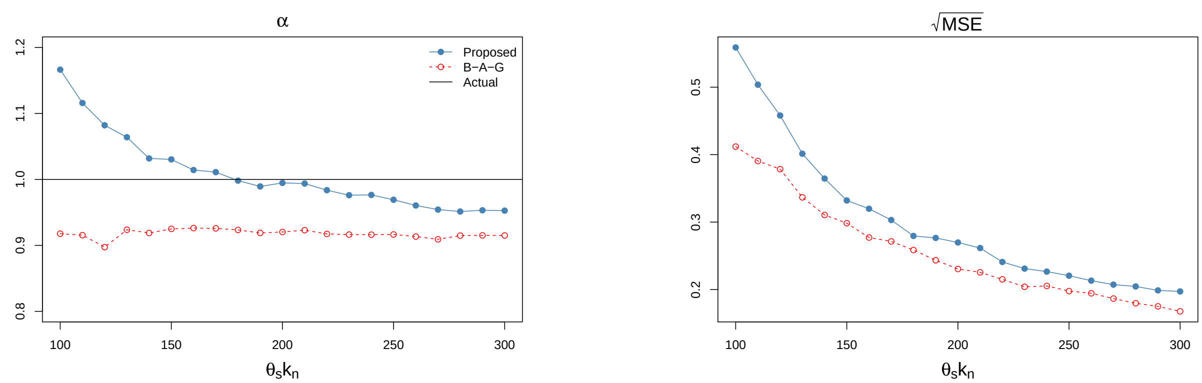

Figure 10: Estimated tail index and $\sqrt{\text{MSE}}$ for Cauchy samples. $n = 2000$, $\alpha = 1$, $k_n = 100$, $\delta = 1$

Figure 11 shows the estimates for 200 independent Cauchy samples without any extremes missing. Both methods produce accurate results for the zero number of missing extremes and the tail index.

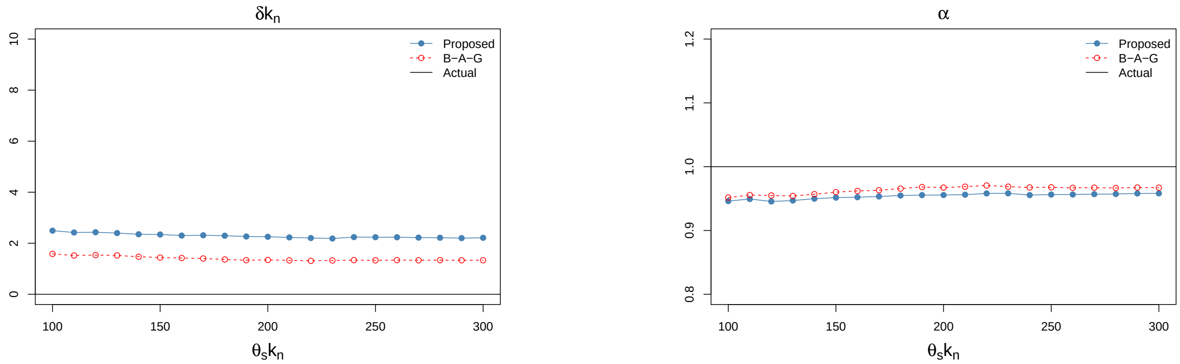

Figure 11: Estimated number of missing extremes and tail index for Cauchy samples. $n = 2000$, $\alpha = 1$, $k_n = 100$, $\delta = 0$

4.2.2 Student's $t_{2.5}$ Samples

Figures 12 and 13 show the estimates for 200 independent samples from the Student's t-distribution with degrees of freedom $df = 2.5$. The tail index $\alpha = df$. In each sample there are $n = 10000$ observations originally. Let $k_n = 200$ and $\delta = 1$ so that the largest 200 observations have been removed from each of the original samples.

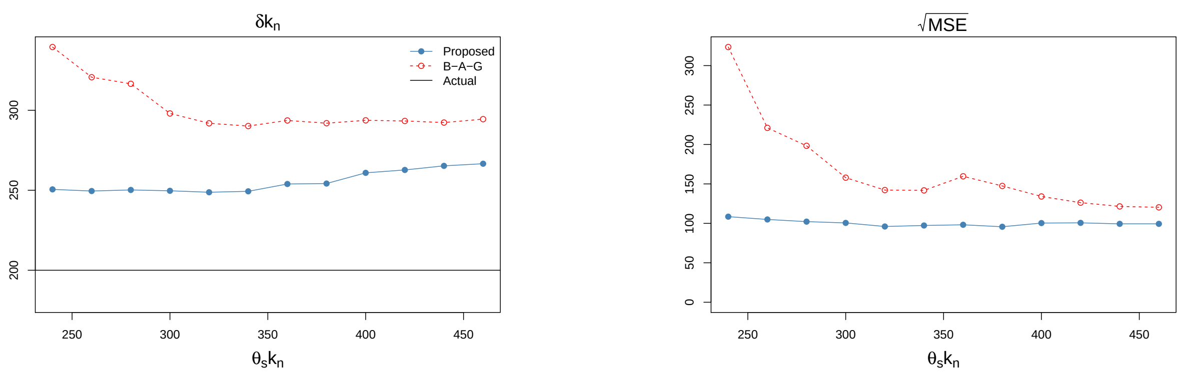

Figure 12: Estimated number of missing extremes and $\sqrt{\text{MSE}}$ for Student's $t_{2.5}$ samples. $n = 10000$, $\alpha = 2.5$, $k_n = 200$, $\delta = 1$

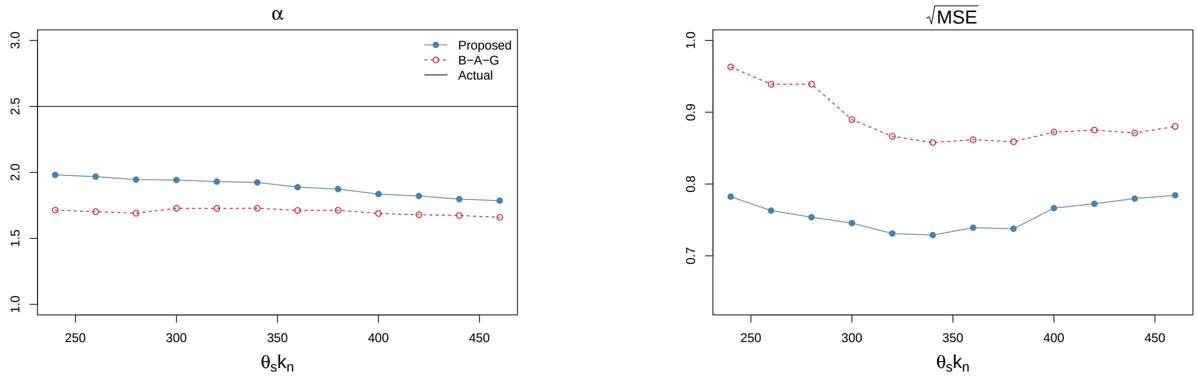

Figure 13: Estimated tail index and $\sqrt{\text{MSE}}$ for Student's $t_{2.5}$ samples. $n = 10000$, $\alpha = 2.5$, $k_n = 200$, $\delta = 1$

5 Applications

We now apply the proposed method to real data. In practice, the number of missing extreme values and the reason for their absence are usually unknown. The consistency of an estimation procedure can be tested by artificially removing a number of additional extremes from the observed data. Consistency requires that, in a certain range, such additional removal should not have a major effect on the estimated tail index. Further, the estimated number of the originally missing upper order statistics should stay, approximately, the same after accounting for the artificially removed observations. Here we examine a massive Google+ social network dataset and a moderate-sized earthquake fatality dataset, and in both cases the proposed procedure provides reasonable results.

5.1 Google+

We first apply our method to the data from the Google+ social network introduced in Section 1. The data contain one of the largest weakly connected components of a snapshot of the network taken on October 19, 2012. A weakly connected component of the network is created by treating the network as undirected and finding all nodes that can be reached from a randomly selected initial node. There are 76,438,791 nodes and 1,442,504,499 edges in this component. The quantities of interest are the in- and out-degrees of nodes in the network, which often exhibit heavy-tailed properties (see, for example, Newman 2010).

We use, as the data set for estimation purposes, the largest 5000 values of the in-degree. We choose $k_n = 200$. Next, we repeat the estimation procedure after artificially removing 400 largest of the 5000 values of the in-degree. In the estimation, we start from $\theta_1 = 1/k_n$ and let $\theta_i = \theta_{i-1} + 1/k_n$ for $1 < i \leq s$. As in the simulation studies, we consider a sequence of different endpoints $\theta_s k_n$ and obtain estimates corresponding to different values of $\theta_s k_n$. For comparison, we also apply the estimation procedure of Beirlant et al. (2016a) to the dataset.

Figures 14 and 15 show, respectively, the estimates of the number of missing extremes and the tail index of the in-degree, before and after the artificial removal. It can be seen by comparing the plots on the left and right panels of Figure 14 that the estimates by the proposed method reflect reasonably well the additional removal of 400 top values. The tail index is mostly estimated to be in the range of $0.5 - 0.6$ and the estimates are reasonably consistent before and after the artificial removal (Figure 15).

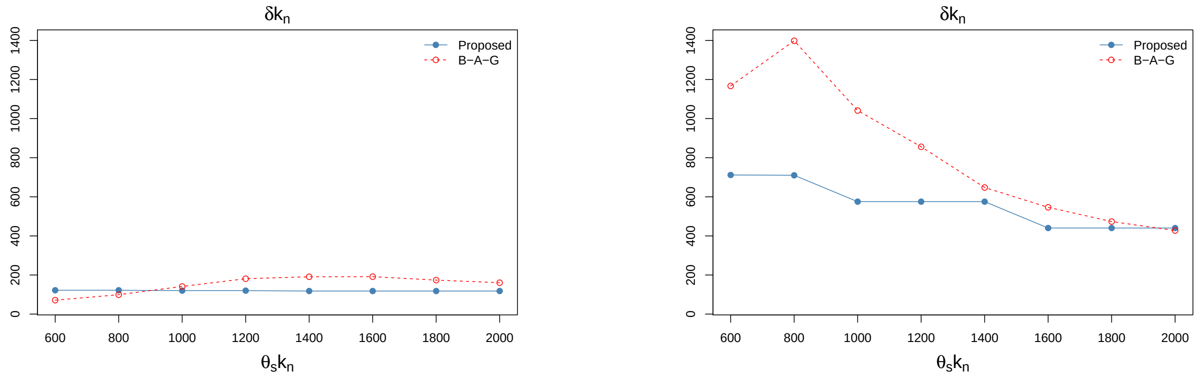

Figure 14: Estimated number of missing extremes. Left: with the original 5000 observations. Right: top 400 values removed

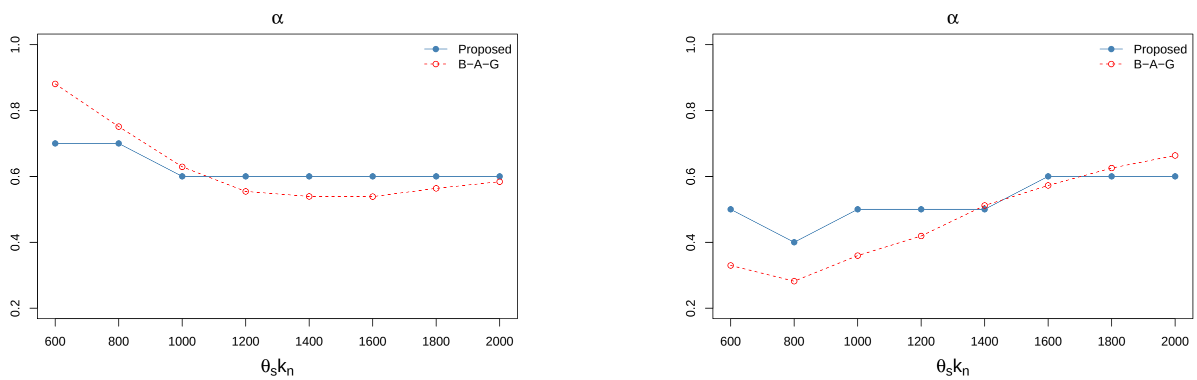

Figure 15: Estimated tail index. Left: with the original 5000 observations. Right: top 400 values removed

5.2 Earthquakes

While power-law distributions are widely used to model natural disasters such as earthquakes, forest fires and floods, some studies (Burroughs and Tebbens 2001a,b, 2002, Clark 2013, Beirlant et al. 2016a,b) have observed evidence of truncation in the data available for such events. Causes for the truncation are complex. Possible explanations include physical limitations on the magnitude of the events (Clark 2013), spatial and temporal sampling limitations and changes in the mechanisms of the events (Burroughs and Tebbens 2001a,b, 2002). In addition, improved detection and rescue techniques might have led to reduction in disaster-related fatalities occurred in recent years.

We apply our method to the dataset of earthquake fatalities (`http://earthquake.usgs.gov/earthquakes/world/world_deaths.php`) published by the U.S. Geological Survey, which was also used for demonstration in Beirlant et al. (2016a). The dataset is of moderate sample size. It contains information of 125 earthquakes causing 1,000 or more deaths from 1900 to 2014. In the estimation procedure we choose $k_n = 10$. Initially the procedure is applied to the original data set. Then we repeat the procedure after artificially removing 10 largest of the 125 values. In the estimation, we start from $\theta_1 = 1/k_n$ and let $\theta_i = \theta_{i-1} + 1/k_n$ for $1 < i \leq s$. We consider a sequence of different endpoints $\theta_s k_n$ and estimate the number of missing extremes and the tail index with different values of $\theta_s k_n$. Since the top k order statistics in the data after removing the top 10 extreme values are the top $k+10$ in the original data without the 10 largest observations, in comparing results before and after the removal, the range of $\theta_s k_n$ for the data after the removal is shifted to the left by 10.

Figures 16 and 17 show the estimates of the number of missing extremes and the tail index of the fatalities. After removing the top 10 earthquakes with the most fatalities, the estimates by the proposed method reflect reasonably well the additional removal (see the left and right panels of Figure 16). The estimates of the tail index are reasonably consistent and remain to be in the range of $0.25 - 0.3$ after the additional removal (Figure 17).

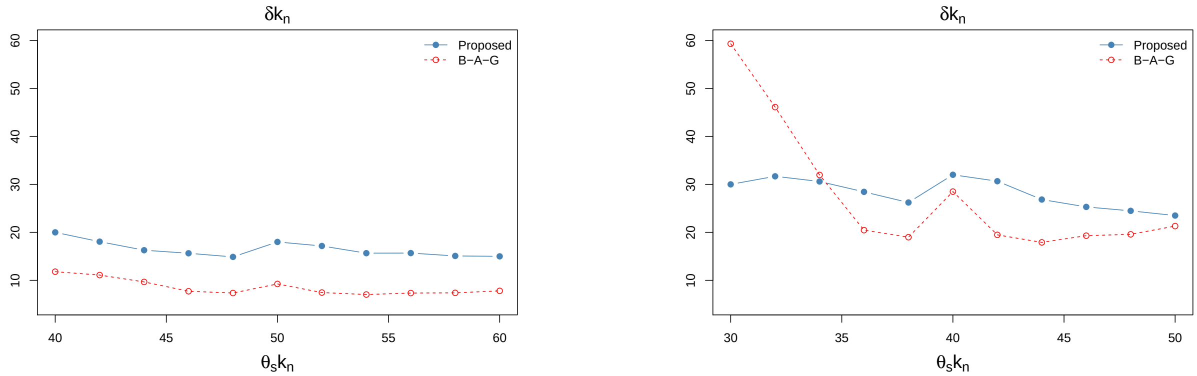

Figure 16: Estimated number of missing extremes. Left: with the original 125 observations. Right: with top 10 values removed

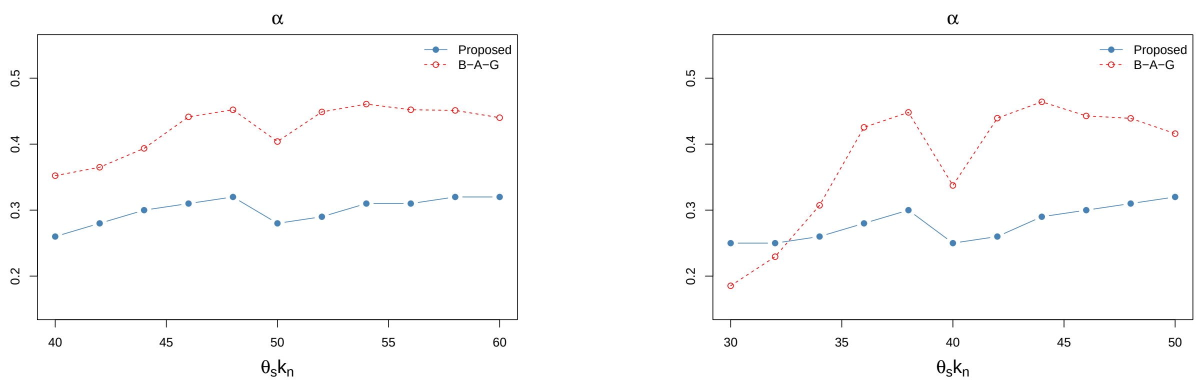

Figure 17: Estimated tail index. Left: with the original 125 observations. Right: with top 10 values removed

Appendix

In the following we address the technical details of the proof of Theorem 2.1. Before proving the main result, we establish some preliminary results. Suppose $X_1, X_2, \ldots$ are iid Pareto random variables with distribution function $F(x) = 1 - x^{-\alpha}$ for $x \geq 1$ and 0 otherwise, where $\alpha \in (0, \infty)$. Since $E_i := \alpha \log X_i$ are iid exponential random variables with mean 1, we have

$$H_n(\delta, \theta) = \frac{1}{\alpha} \frac{1}{\lfloor \theta k_n \rfloor} \sum_{i=1}^{\lfloor \theta k_n \rfloor} E_{(n - \lfloor \delta k_n \rfloor - i + 1)} - E_{(n - \lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)}$$

where $E_{(1)} \leq E_{(2)} \leq \cdots \leq E_{(n)}$ are increasing order statistics of $E_i, \ldots, E_n$. Applying Rényi's representation (de Haan and Ferreira 2006),

$$\{E_{(i)}\}_{i=1}^n \stackrel{d}{=} \left\{ \sum_{j=1}^{i} \frac{1}{n - j + 1} E_j \right\}_{i=1}^{n}, \tag{9}$$

so that for all $\delta \geq 0$ and $\theta \geq 1/k_n$,

$$\begin{aligned} H_n(\delta, \theta) &= \frac{1}{\alpha} \frac{1}{\lfloor \theta k_n \rfloor} \sum_{i=1}^{\lfloor \theta k_n \rfloor} E_{(n - \lfloor \delta k_n \rfloor - i + 1)} - E_{(n - \lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)} \\ &\stackrel{d}{=} \frac{1}{\alpha} \frac{1}{\lfloor \theta k_n \rfloor} \sum_{i=1}^{\lfloor \theta k_n \rfloor} \sum_{j=n - \lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor + 1}^{n - \lfloor \delta k_n \rfloor - i + 1} \frac{1}{n - j + 1} E_j \\ &= \frac{1}{\alpha} \frac{1}{\lfloor \theta k_n \rfloor} \sum_{i=n - \lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor + 1}^{n - \lfloor \delta k_n \rfloor} \sum_{j=n - \lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor + 1}^{i} \frac{1}{n - j + 1} E_j \\ &= \frac{1}{\alpha} \frac{1}{\lfloor \theta k_n \rfloor} \sum_{j=n - \lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor + 1}^{n - \lfloor \delta k_n \rfloor} \frac{n - \lfloor \delta k_n \rfloor - j + 1}{n - j + 1} E_j. \end{aligned} \tag{10}$$

Lemma 5.1. Suppose $X_1, X_2, \ldots$ are iid Pareto random variables with distribution function $F(x) = 1 - x^{-\alpha}$ for $x \geq 1$ and 0 otherwise, where $\alpha \in (0, \infty)$. Let

$$W_n(\delta, \theta) = \alpha \sqrt{k_n} (H_n(\delta, \theta) - E H_n(\delta, \theta)), \tag{11}$$

then as $n \to \infty$,

$$W_n(\cdot,\cdot) \stackrel{\text{fidi}}{\to} G(\cdot,\cdot),$$

where $\stackrel{\text{fidi}}{\to}$ is convergence in finite dimensional distributions and G is as in Theorem (2.1).

Proof. By (10), the distribution of the process $\{W_n(\delta,\theta)\}$ is the same as the distribution of the process

$$\frac{\sqrt{k_n}}{\lfloor \theta k_n \rfloor} \sum_{j=n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor + 1}^{n-\lfloor \delta k_n \rfloor} \frac{n - \lfloor \delta k_n \rfloor - j + 1}{n - j + 1}(E_j - 1).$$

For any $\theta > 0$, $\delta \geq 0$ and $\epsilon > 0$,

$$E\Bigg[\sum_{j=n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor + 1}^{n-\lfloor \delta k_n \rfloor} \left(\frac{n - \lfloor \delta k_n \rfloor - j + 1}{n - j + 1} \right)^{2+\epsilon} (E_j - 1)^{2+\epsilon} \Bigg] \leq \theta k_n E(E_1 - 1)^{2+\epsilon} = C k_n,$$

where $C = \theta E(E_1 - 1)^{2+\epsilon}$ is a finite constant, and

$$\begin{aligned} &\mathrm{Var}\Bigg(\sum_{j=n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor + 1}^{n-\lfloor \delta k_n \rfloor} \frac{n - \lfloor \delta k_n \rfloor - j + 1}{n - j + 1}(E_j - 1) \Bigg) = \mathrm{Var}\Big(\sum_{j=1}^{\lfloor \theta k_n \rfloor} \frac{j}{j + \lfloor \delta k_n \rfloor} E_j \Big) \\ &= \sum_{j=1}^{\lfloor \theta k_n \rfloor} \left(\frac{j}{j + \lfloor \delta k_n \rfloor} \right)^2 = \lfloor \theta k_n \rfloor - 2\lfloor \delta k_n \rfloor \sum_{j=1}^{\lfloor \theta k_n \rfloor} \frac{1}{j + \lfloor \delta k_n \rfloor} + \lfloor \delta k_n \rfloor^2 \sum_{j=1}^{\lfloor \theta k_n \rfloor} \left(\frac{1}{j + \lfloor \delta k_n \rfloor} \right)^2. \quad (12) \end{aligned}$$

If $\delta = 0$, then (12) is $\lfloor \theta k_n \rfloor$. If $\delta > 0$, then as $n \to \infty$,

$$\begin{aligned} \sum_{j=1}^{\lfloor \theta k_n \rfloor} \frac{1}{j + \lfloor \delta k_n \rfloor} &= \sum_{j=\lfloor \delta k_n \rfloor + 1}^{\lfloor \delta k_n \rfloor + \lfloor \theta k_n \rfloor} \frac{1}{j} \to \log\left(\frac{\theta}{\delta} + 1 \right), \\ \sum_{j=1}^{\lfloor \theta k_n \rfloor} \left(\frac{1}{j + \lfloor \delta k_n \rfloor} \right)^2 &= \sum_{j=\lfloor \delta k_n \rfloor + 1}^{\lfloor \delta k_n \rfloor + \lfloor \theta k_n \rfloor} \frac{1}{j^2} \sim \frac{1}{\lfloor \delta k_n \rfloor} - \frac{1}{\lfloor \delta k_n \rfloor + \lfloor \theta k_n \rfloor}, \end{aligned}$$

and hence there exists a finite constant C', such that

$$\mathrm{Var}\Bigg(\sum_{j=n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor + 1}^{n-\lfloor \delta k_n \rfloor} \frac{n - \lfloor \delta k_n \rfloor - j + 1}{n - j + 1}(E_j - 1) \Bigg) \sim C' k_n.$$

Since $Ck_n/(C'k_n)^{1+\epsilon/2} \to 0$ as $n \to \infty$, it follows by the Lyapunov central limit theorem that

$$\frac{\sqrt{k_n}}{\lfloor \theta k_n \rfloor} \sum_{j=n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor + 1}^{n-\lfloor \delta k_n \rfloor} \frac{n - \lfloor \delta k_n \rfloor - j + 1}{n - j + 1}(E_j - 1) \tag{13}$$

converges weakly to a Normal distribution with mean zero. Further by the multivariate Lyapunov central limit theorem, the finite dimensional distributions of $W_n(\delta, \theta)$ converges to multivariate Normal distribution with mean zero. Assume that $\delta_1 > \delta_2$ and put $\delta_l + \theta_l = (\delta_1 + \theta_1) \wedge (\delta_2 + \theta_2)$. Then the covariance

$$\begin{aligned}
&\operatorname{Cov}(W_n(\delta_1, \theta_1), W_n(\delta_2, \theta_2)) \\
&= \frac{k_n}{\lfloor \theta_1 k_n \rfloor \lfloor \theta_2 k_n \rfloor} E\Big[\sum_{j=n-\lfloor \delta_1 k_n \rfloor - \lfloor \theta_1 k_n \rfloor + 1}^{n-\lfloor \delta_1 k_n \rfloor} \frac{n - \lfloor \delta_1 k_n \rfloor - j + 1}{n - j + 1}(E_j - 1) \\
&\qquad\qquad \cdot \sum_{j=n-\lfloor \delta_2 k_n \rfloor - \lfloor \theta_2 k_n \rfloor + 1}^{n-\lfloor \delta_2 k_n \rfloor} \frac{n - \lfloor \delta_2 k_n \rfloor - j + 1}{n - j + 1}(E_j - 1) \Big] \\
&= \frac{k_n}{\lfloor \theta_1 k_n \rfloor \lfloor \theta_2 k_n \rfloor} \sum_{j=n-\lfloor \delta_l k_n \rfloor - \lfloor \theta_l k_n \rfloor + 1}^{n-\lfloor \delta_1 k_n \rfloor} \Big(1 - \frac{\lfloor \delta_1 k_n \rfloor}{n - j + 1}\Big)\Big(1 - \frac{\lfloor \delta_2 k_n \rfloor}{n - j + 1}\Big) + o(1) \\
&= \frac{k_n}{\lfloor \theta_1 k_n \rfloor \lfloor \theta_2 k_n \rfloor} \Bigg((\lfloor \delta_l k_n \rfloor + \lfloor \theta_l k_n \rfloor - \lfloor \delta_1 k_n \rfloor) \\
&\qquad\qquad - (\lfloor \delta_1 k_n \rfloor + \lfloor \delta_2 k_n \rfloor) \sum_{j=n-\lfloor \delta_l k_n \rfloor - \lfloor \theta_l k_n \rfloor + 1}^{n-\lfloor \delta_1 k_n \rfloor} \frac{1}{n - j + 1} \\
&\qquad\qquad + \lfloor \delta_1 k_n \rfloor \lfloor \delta_2 k_n \rfloor \sum_{j=n-\lfloor \delta_l k_n \rfloor - \lfloor \theta_l k_n \rfloor + 1}^{n-\lfloor \delta_1 k_n \rfloor} \frac{1}{(n - j + 1)^2} \Bigg) + o(1) \\
&\sim \frac{1}{\theta_1 \theta_2} \Bigg(\delta_l + \theta_l - \delta_1 - (\delta_1 + \delta_2) \log\Big(\frac{\delta_l + \theta_l}{\delta_1}\Big) + \delta_1 \delta_2 \Big(\frac{1}{\delta_1} - \frac{1}{\delta_l + \theta_l}\Big) \Bigg).
\end{aligned}$$

By Slutsky's theorem,

$$W_n(\delta, \theta) \overset{\text{fidi}}{\to} G(\delta, \theta).$$

$\square$

The following lemma states that the process $\{W_n(\delta, \theta)\}$ satisfies a sufficient condition for tightness given by Bickel and Wichura (1971).

Lemma 5.2. There exists a constant C, such that for all $k_n \in \mathbb{N}$ and non-negative integers M_1, M_2, N_1, N_2 satisfying $M_1 < M_2$ and $\epsilon k_n < N_1 < N_2$, where $\epsilon > 0$ is a fixed constant,

$$E(|W(B)|^6) \leq C\lambda(B)^{3/2}. \tag{14}$$

Here λ is the Lebesgue measure, $B = (M_1/k_n, M_2/k_n] \times (N_1/k_n, N_2/k_n]$ and

$$W(B) = W_n\Big(\frac{M_2}{k_n}, \frac{N_2}{k_n}\Big) - W_n\Big(\frac{M_1}{k_n}, \frac{N_2}{k_n}\Big) - \Big(\frac{M_2}{k_n}, \frac{N_1}{k_n}\Big) + W_n\Big(\frac{M_1}{k_n}, \frac{N_1}{k_n}\Big).$$

In addition,

$$E\bigg|W_n\bigg(\frac{M_2}{k_n}, \frac{N_1}{k_n}\bigg) - W_n\bigg(\frac{M_1}{k_n}, \frac{N_1}{k_n}\bigg)\bigg|^6 \leq C\bigg(\frac{M_2 - M_1}{k_n}\bigg)^{3/2} \tag{15}$$

and

$$E\bigg|W_n\bigg(\frac{M_1}{k_n}, \frac{N_2}{k_n}\bigg) - W_n\bigg(\frac{M_1}{k_n}, \frac{N_1}{k_n}\bigg)\bigg|^6 \leq C\bigg(\frac{N_2 - N_1}{k_n}\bigg)^{3/2}. \tag{16}$$

Proof. By (10),

$$\bigg\{W_n\Big(\frac{M_p}{k_n}, \frac{N_q}{k_n}\Big)\bigg\}_{p,q=1,2} \stackrel{d}{=} \bigg\{\frac{\sqrt{k_n}}{N_q}\sum_{j=n-M_p-N_q+1}^{n-M_p} \frac{n - M_p - j + 1}{n - j + 1}(E_j - 1)\bigg\}_{p,q=1,2},$$

where $\{E_j\}$ are iid standard exponential variables. For simplicity, let $\tilde{E}_j = E_j - 1$ and

$$C_{p,q}(j) = \frac{1}{N_q}\frac{n - M_p - j + 1}{n - j + 1}.$$

First assume that $M_2 - M_1 < N_2 - N_1$ and $M_2 - M_1 < N_1$. Then we have

$$W(B) \stackrel{d}{=} \sqrt{k_n}\bigg[\sum_{j=n-M_2-N_2+1}^{n-M_2} C_{2,2}(j)\tilde{E}_j - \sum_{j=n-M_2-N_1+1}^{n-M_2} C_{2,1}(j)\tilde{E}_j$$

$$
\begin{aligned}
&\qquad - \sum_{j=n-M_1-N_2+1}^{n-M_1} C_{1,2}(j)\tilde{E}_j + \sum_{j=n-M_1-N_1+1}^{n-M_1} C_{1,1}(j)\tilde{E}_j\Big] \\
&= \sqrt{k_n}\Big[\sum_{j=n-M_2-N_2+1}^{n-M_1-N_2} C_{2,2}(j)\tilde{E}_j && (17) \\
&\qquad + \sum_{j=n-M_1-N_2+1}^{n-M_2-N_1} [C_{2,2}(j) - C_{1,2}(j)]\tilde{E}_j && (18) \\
&\qquad + \sum_{j=n-M_2-N_1+1}^{n-M_1-N_1} [C_{2,2}(j) - C_{1,2}(j) - C_{2,1}(j)]\tilde{E}_j && (19) \\
&\qquad + \sum_{j=n-M_1-N_1+1}^{n-M_2} [C_{2,2}(j) - C_{1,2}(j) - C_{2,1}(j) + C_{1,1}(j)]\tilde{E}_j && (20) \\
&\qquad + \sum_{j=n-M_2+1}^{n-M_1} [C_{1,1}(j) - C_{1,2}(j)]\tilde{E}_j\Big], && (21)
\end{aligned}
$$

and the ranges of the sums in (17) - (21) are disjoint. To show (14), we need to examine the upper bound of

$$
\frac{k_n^3(E|W(B)|^6)}{(M_2-M_1)^{3/2}(N_2-N_1)^{3/2}}. \tag{22}
$$

To further simplify the notation, introduce the following coefficients

$$
d_j = \begin{cases}
\frac{k_n C_{2,2}(j)}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}, & n-M_2-N_2+1 \le j \le n-M_1-N_2, \\
\frac{k_n (C_{2,2}(j)-C_{1,2}(j))}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}, & n-M_1-N_2+1 \le j \le n-M_2-N_1, \\
\frac{k_n (C_{2,2}(j)-C_{1,2}(j)-C_{2,1}(j))}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}, & n-M_2-N_1+1 \le j \le n-M_1-N_1, \\
\frac{k_n (C_{2,2}(j)-C_{1,2}(j)-C_{2,1}(j)+C_{1,1}(j))}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}, & n-M_1-N_1+1 \le j \le n-M_2, \\
\frac{k_n (C_{1,1}(j)-C_{1,2}(j))}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}, & n-M_2+1 \le j \le n-M_1,
\end{cases}
$$

so that (22) becomes

$$
E\Big[\sum_{j=n-M_2-N_2+1}^{n-M_1} d_j\tilde{E}_j\Big]^6,
$$

which, by convexity, is bounded by

$$
\begin{aligned}
&K \cdot E\Bigg[\Bigg(\sum_{j=n-M_2-N_2+1}^{n-M_1-N_2} d_j\tilde{E}_j\Bigg)^6 + \Bigg(\sum_{j=n-M_1-N_2+1}^{n-M_2-N_1} d_j\tilde{E}_j\Bigg)^6 + \Bigg(\sum_{j=n-M_2-N_1+1}^{n-M_1-N_1} d_j\tilde{E}_j\Bigg)^6 \\
&\quad + \Bigg(\sum_{j=n-M_1-N_1+1}^{n-M_2} d_j\tilde{E}_j\Bigg)^6 + \Bigg(\sum_{j=n-M_2+1}^{n-M_1} d_j\tilde{E}_j\Bigg)^6\Bigg], \qquad (23)
\end{aligned}
$$

where K is a constant independent of M_i, N_i and k_n. For $n-M_2-N_2+1 \le j \le n-M_1-N_2$,

$$
\begin{aligned}
|d_j| = \left|\frac{k_n C_{2,2}(j)}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}\right| &= \frac{k_n}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}\frac{1}{N_2}\frac{n-M_2-j+1}{n-j+1} \\
&\le \frac{k_n}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}\frac{1}{M_2+N_2} := \tilde{d}_1,
\end{aligned}
$$

for $n-M_1-N_2+1 \le j \le n-M_2-N_1$,

$$
\begin{aligned}
|d_j| = \left|\frac{k_n(C_{2,2}(j)-C_{1,2}(j))}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}\right| &= \left|\frac{k_n}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}\frac{1}{N_2}\frac{M_1-M_2}{n-j+1}\right| \\
&\le \frac{k_n}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}\frac{1}{N_2}\frac{M_2-M_1}{M_2+N_1+1} := \tilde{d}_2,
\end{aligned}
$$

for all $n-M_2-N_1+1 \le j \le n-M_1-N_1$,

$$
\begin{aligned}
|d_j| &= \left|\frac{k_n(C_{2,2}(j)-C_{1,2}(j)-C_{2,1}(j))}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}\right| \\
&= \frac{k_n}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}\left(\frac{1}{N_2}\frac{M_2-M_1}{n-j+1} + \frac{1}{N_1}\frac{n-M_2-j+1}{n-j+1}\right) \\
&\le \frac{k_n}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}\frac{1}{N_1}\frac{n-M_1-j+1}{n-j+1} \\
&\le \frac{k_n}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}\frac{1}{N_1}\frac{(M_2-M_1)+N_1}{M_2+N_1} := \tilde{d}_3
\end{aligned}
$$

and for $n-M_2+1 \le j \le n-M_1$,

$$
|d_j| = \left|\frac{k_n(C_{1,1}(j)-C_{1,2}(j))}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}}\right|
$$

$$= \frac{k_n}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}} \frac{n-M_1-j+1}{n-j+1}\left(\frac{1}{N_1}-\frac{1}{N_2}\right)$$
$$\leq \frac{k_n}{(M_2-M_1)^{1/4}(N_2-N_1)^{1/4}} \frac{M_2-M_1}{M_2}\frac{N_2-N_1}{N_1N_2} := \tilde{d}_5.$$

Thus (23) is bounded by

$$K \cdot \left[\tilde{d}_1^6 E\left(\sum_{j=n-M_2-N_2+1}^{n-M_1-N_2} \tilde{E}_j\right)^6 + \tilde{d}_2^6 E\left(\sum_{j=n-M_1-N_2+1}^{n-M_2-N_1} \tilde{E}_j\right)^6 + \tilde{d}_3^6 E\left(\sum_{j=n-M_2-N_1+1}^{n-M_1-N_1} \tilde{E}_j\right)^6 \right.$$
$$\left. + E\left(\sum_{j=n-M_1-N_1+1}^{n-M_2} d_j\tilde{E}_j\right)^6 + \tilde{d}_5^6 E\left(\sum_{j=n-M_2+1}^{n-M_1} \tilde{E}_j\right)^6\right], \tag{24}$$

Denote the number of $\tilde{E}_j$ in the sum with coefficient d_j by s_i, then $s_1 = s_3 = s_5 = M_2 - M_1$ and $s_2 = (N_2 - N_1) - (M_2 - M_1)$. Expanding terms in (24), by the independence of the $\tilde{E}_j$ and that $E(\tilde{E}_j) = 0$, the non-zero terms are those consist of the second and higher moments of $\{\tilde{E}_j\}$ only. Therefore,

$$\tilde{d}_1^6 E\left(\sum_{j=n-M_2-N_2+1}^{n-M_1-N_2} \tilde{E}_j\right)^6 \leq 6!\,\tilde{d}_1^6\Big(E(\tilde{E}_j)^6 + s_1^2E(\tilde{E}_j)^2E(\tilde{E}_j)^4 + s_1^2(E(\tilde{E}_j)^3)^2 + s_1^3(E(\tilde{E}_j)^2)^3\Big),$$

and to show it is bounded by a constant, it suffices to show $s_1\tilde{d}_1^2$ is bounded by a constant, as $\tilde{d}_1$ is bounded and the moments of $\tilde{E}_j$ are finite. For $i = 1$,

$$s_1\tilde{d}_1^2 = (M_2-M_1)\frac{k_n^2}{(M_2-M_1)^{1/2}(N_2-N_1)^{1/2}}\frac{1}{(M_2+N_2)^2} \leq \frac{1}{\epsilon^2}.$$

The same argument applies to $i = 2, 3, 5$. For $i = 2$,

$$s_2\tilde{d}_2^2 = ((N_2-N_1)-(M_2-M_1))\frac{k_n^2}{(M_2-M_1)^{1/2}(N_2-N_1)^{1/2}}\frac{1}{N_2^2}\frac{(M_2-M_1)^2}{(M_2+N_1+1)^2}$$
$$\leq (N_2-N_1)\frac{k_n^2}{(M_2-M_1)^{1/2}(N_2-N_1)^{1/2}}\frac{1}{N_2^2}\frac{(M_2-M_1)^2}{(M_2+N_1+1)^2} \leq \frac{1}{\epsilon^2},$$

for $i = 3$,

$$s_3\tilde{d}_3^2 = (M_2-M_1)\frac{k_n^2}{(M_2-M_1)^{1/2}(N_2-N_1)^{1/2}}\frac{1}{N_1^2}\frac{((M_2-M_1)+N_1)^2}{(M_2+N_1)^2} \leq \frac{1}{\epsilon^2},$$

and for $i = 5$,

$$s_5 \tilde{d}_5^2 = (M_2 - M_1) \frac{k_n^2}{(M_2 - M_1)^{1/2}(N_2 - N_1)^{1/2}} \frac{(M_2 - M_1)^2}{M_2^2} \frac{(N_2 - N_1)^2}{N_1^2 N_2^2} \leq \frac{1}{\epsilon^2}.$$

Finally, note

$$E\bigg(\sum_{j=n-M_1-N_1+1}^{n-M_2} d_j \tilde{E}_j \bigg)^6 \tag{25}$$

$$\leq 6! \bigg(\sum_j d_j^6 E(\tilde{E}_j)^6 + \sum_{i,j} d_i^3 d_j^3 (E(\tilde{E}_i)^3)^2 + \sum_{i,j} d_i^2 d_j^4 E(\tilde{E}_i)^2 E(\tilde{E}_j)^4 + \sum_{i,j,k} d_i^2 d_j^2 d_k^2 (E(\tilde{E}_i)^2)^3 \bigg),$$

where

$$\begin{aligned} d_j &= \frac{k_n(C_{2,2}(j) - C_{1,2}(j) - C_{2,1}(j) + C_{1,1}(j))}{(M_2 - M_1)^{1/4}(N_2 - N_1)^{1/4}} \\ &= \frac{k_n}{(M_2 - M_1)^{1/4}(N_2 - N_1)^{1/4}} \frac{M_2 - M_1}{n - j + 1} \left(\frac{1}{N_1} - \frac{1}{N_2} \right) \end{aligned}$$

for $n - M_1 - N_1 + 1 \leq j \leq n - M_2$. Therefore (25) is bounded by a constant times

$$\begin{aligned} &\sum_j d_j^6 + \sum_{i,j} (d_i^3 d_j^3 + d_i^2 d_j^4) + \sum_{i,j,k} d_i^2 d_j^2 d_k^2 \\ &= \left(\frac{k_n(M_2 - M_1)}{(M_2 - M_1)^{1/4}(N_2 - N_1)^{1/4}} \frac{N_2 - N_1}{N_1 N_2} \right)^6 \left(\sum_{j=M_2+1}^{N_1+M_1} \frac{1}{j^6} + \sum_{i,j} \left(\frac{1}{i^3} \frac{1}{j^3} + \frac{1}{i^2} \frac{1}{j^4} \right) + \sum_{i,j,k} \frac{1}{i^2} \frac{1}{j^2} \frac{1}{k^2} \right) \\ &\leq \left(\frac{1}{\epsilon} \frac{(M_2 - M_1)^{3/4}}{N_2^{1/4}} \right)^6 \left(\frac{N_1}{(M_2 + 1)^6} + \frac{7}{12 M_2^4} + \frac{1}{M_2^3} \right) \leq \frac{3}{\epsilon^6}, \end{aligned}$$

as $1/(j+1)^k \leq \int_j^{j+1} (1/t^k) dt$ for $k > 1$. Therefore (24) is bounded and the condition (14) holds.

If $M_2 - M_1 < N_2 - N_1$ and $M_2 - M_1 \geq N_1$ (if equation holds then $E\big(\sum_{j=n-M_2}^{n-M_1-N_1+1} d_j \tilde{E}_j \big)^6$ disappears), then in the above calculations, terms that are different are

$$|d_j| \leq \frac{k_n}{(M_2 - M_1)^{1/4}(N_2 - N_1)^{1/4}} \frac{1}{N_1} \frac{(M_2 - M_1) + N_1}{M_2 + N_1} := \tilde{d}_3$$

for $n - M_2 - N_1 + 1 \le j \le n - M_2$ with $s_3 = N_1$, and

$$|d_j| \le \frac{k_n}{(M_2 - M_1)^{1/4}(N_2 - N_1)^{1/4}} \frac{1}{N_2} \frac{M_2 - M_1}{M_2} := \tilde{d}_4$$

for $n - M_2 + 1 \le j \le n - M_1 - N_1$ with $s_4 = (M_2 - M_1) - N_1$, and

$$|d_j| \le \frac{k_n}{(M_2 - M_1)^{1/4}(N_2 - N_1)^{1/4}} \frac{1}{M_1 + N_1} \frac{N_2 - N_1}{N_2} := \tilde{d}_5$$

for $n - M_1 - N_1 + 1 \le j \le n - M_1$ with $s_5 = N_1$. It can be shown $s_i \tilde{d}_i^2 \le 1/\epsilon^2$ still holds for each coefficient. The case of $M_2 - M_1 \ge N_2 - N_1$ and conditions (15) and (16) can be shown similarly.

$\square$

Now we are ready for the proof of the main result.

Proof of Theorem (2.1). It suffices to show that for all $0 < m < M$ the weak convergence holds on the Skorokhod space $D([0, M] \times [m, M])$. Details about the structure of $D([0, M] \times [m, M])$ can be found in Straf (1972).

Let U be the left-continuous inverse function of $1/\bar{F}$. The second-order regular varying condition (3) implies that (Drees 1998 and de Haan and Ferreira 2006)

$$\lim_{t \to \infty} \frac{x^{-1/\alpha} \frac{U(tx)}{U(t)} - 1}{A(t)} = \frac{x^\rho - 1}{\rho},$$

which is equivalent to

$$\lim_{t \to \infty} \frac{\log U(tx) - \log U(t) - \log(x)/\alpha}{A(t)} = \frac{x^\rho - 1}{\rho}.$$

Moreover, there exists $A_0(t) \sim A(t)$ that is regular varying with index ρ (denoted by $|A_0| \in \mathrm{RV}(\rho)$), such that for all $\epsilon > 0$, there exists $t_0 = t_0(\epsilon)$, and for all $t \ge t_0$ and $x \ge 1$,

$$\left| \frac{\log U(tx) - \log U(t) - \log(x)/\alpha}{A_0(t)} - \frac{x^\rho - 1}{\rho} \right| \le \epsilon x^{\rho + \epsilon}. \tag{26}$$

Let $Y_i = e^{E_i}$, where $E_1, E_2, \ldots$ are iid standard exponential random variables. Note $U(Y_{(i)}) \overset{\mathrm{d}}{=} X_{(i)}$, and thus

$$\{H_n(\delta,\theta)\} \overset{d}{=} \left\{ \frac{1}{\lfloor \theta k_n \rfloor} \sum_{i=1}^{\lfloor \theta k_n \rfloor} \log U(Y_{(n-\lfloor \delta k_n \rfloor - i+1)}) - \log U(Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)}) \right\}.$$

Without loss of generality, we replace $X_{(i)}$ with $U(Y_{(i)})$ in the following arguments.

Let $t = \min_{(\delta,\theta)\in[0,M]\times[m,M]} Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)}$. Since $Y_{(n-\lfloor \delta k_n \rfloor - i+1)}/Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)} \geq 1$ for all $i = 1, \ldots, \lfloor \theta k_n \rfloor$, (26) implies that on $\{\min_{(\delta,\theta)\in[0,M]\times[m,M]} Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)} \geq t_0\}$, for all $(\delta,\theta) \in [0,M]\times[m,M]$,

$$\begin{aligned} &\Bigg| \alpha\sqrt{k_n} H_n(\delta,\theta) - \sqrt{k_n} H_n^E(\delta,\theta) \\ &\quad - \frac{\alpha}{\rho}\sqrt{k_n} A_0(Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)}) \frac{1}{\lfloor \theta k_n \rfloor} \sum_{i=1}^{\lfloor \theta k_n \rfloor} \left[\left(\frac{Y_{(n-\lfloor \delta k_n \rfloor - i+1)}}{Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)}} \right)^{\rho} - 1 \right] \Bigg| \\ &\leq \epsilon\, \alpha \sqrt{k_n} |A_0(Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)})| \frac{1}{\lfloor \theta k_n \rfloor} \sum_{i=1}^{\lfloor \theta k_n \rfloor} \left(\frac{Y_{(n-\lfloor \delta k_n \rfloor - i+1)}}{Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)}} \right)^{\rho+\epsilon}, \end{aligned}$$

where

$$H_n^E(\delta,\theta) = \frac{1}{\lfloor \theta k_n \rfloor} \sum_{i=1}^{\lfloor \theta k_n \rfloor} E_{(n-\lfloor \delta k_n \rfloor - i+1)} - E_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)}.$$

Let $W_n(\delta,\theta) = \alpha\sqrt{k_n}(H_n^E(\delta,\theta) - EH_n^E(\delta,\theta))$, it follows that

$$\begin{aligned} &\left| \sqrt{k_n}\big(\alpha H_n(\delta,\theta) - g(\delta,\theta)\big) - \alpha b_\rho(\delta,\theta) - W_n(\delta,\theta) \right| \\ &\leq \sqrt{k_n} \big| E(H_n^E(\delta,\theta)) - g(\delta,\theta) \big| \qquad (27) \\ &\quad + \alpha \left| \frac{1}{\rho}\sqrt{k_n} A_0(Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)}) \frac{1}{\lfloor \theta k_n \rfloor} \sum_{i=1}^{\lfloor \theta k_n \rfloor} \left[\left(\frac{Y_{(n-\lfloor \delta k_n \rfloor - i+1)}}{Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)}} \right)^{\rho} - 1 \right] - b_\rho(\delta,\theta) \right| \qquad (28) \\ &\quad + \epsilon \cdot \alpha \sqrt{k_n}\, |A_0(Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)})| \frac{1}{\lfloor \theta k_n \rfloor} \sum_{i=1}^{\lfloor \theta k_n \rfloor} \left(\frac{Y_{(n-\lfloor \delta k_n \rfloor - i+1)}}{Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)}} \right)^{\rho+\epsilon}. \qquad (29) \end{aligned}$$

Now we show (27)-(29) convergence to zero uniformly in $(\delta,\theta)\in[0,M]\times[m,M]$. For (27), by (10),

$$
\begin{aligned}
&\sqrt{k_n}|E(H_n^E(\delta,\theta))-g(\delta,\theta)|\\
&=\sqrt{k_n}\left|\frac{1}{\lfloor\theta k_n\rfloor}\sum_{j=n-\lfloor\delta k_n\rfloor-\lfloor\theta k_n\rfloor+1}^{n-\lfloor\delta k_n\rfloor}\frac{n-\lfloor\delta k_n\rfloor-j+1}{n-j+1}-\left(1-\frac{\delta}{\theta}\log\left(\frac{\theta}{\delta}+1\right)\right)\right|\\
&=\sqrt{k_n}\left|\frac{\delta}{\theta}\log\left(\frac{\theta}{\delta}+1\right)-\frac{\lfloor\delta k_n\rfloor}{\lfloor\theta k_n\rfloor}\sum_{j=\lfloor\delta k_n\rfloor+1}^{\lfloor\delta k_n\rfloor+\lfloor\theta k_n\rfloor}\frac{1}{j}\right|. \qquad (30)
\end{aligned}
$$

If $\delta k_n\geq 1$, then (30) is bounded by

$$
\sqrt{k_n}\left|\frac{\delta}{\theta}-\frac{\lfloor\delta k_n\rfloor}{\lfloor\theta k_n\rfloor}\right|\log\left(\frac{\theta}{\delta}+1\right)+\sqrt{k_n}\frac{\lfloor\delta k_n\rfloor}{\lfloor\theta k_n\rfloor}\left|\log\left(\frac{\theta}{\delta}+1\right)-\sum_{j=\lfloor\delta k_n\rfloor+1}^{\lfloor\delta k_n\rfloor+\lfloor\theta k_n\rfloor}\frac{1}{j}\right|. \qquad (31)
$$

For the first part of (31),

$$
\sqrt{k_n}\left|\frac{\delta}{\theta}-\frac{\lfloor\delta k_n\rfloor}{\lfloor\theta k_n\rfloor}\right|\log\left(\frac{\theta}{\delta}+1\right)\leq\sqrt{k_n}\frac{\delta}{\theta}\frac{1}{\delta k_n\wedge\lfloor\theta k_n\rfloor}\log\left(\frac{\theta}{\delta}+1\right),
$$

which converges uniformly to zero. For the second part,

$$
\begin{aligned}
&\sqrt{k_n}\frac{\lfloor\delta k_n\rfloor}{\lfloor\theta k_n\rfloor}\left[\log\left(\frac{\theta}{\delta}+1\right)-\sum_{j=\lfloor\delta k_n\rfloor+1}^{\lfloor\delta k_n\rfloor+\lfloor\theta k_n\rfloor}\frac{1}{j}\right] \qquad (32)\\
&\leq\sqrt{k_n}\frac{\lfloor\delta k_n\rfloor}{\lfloor\theta k_n\rfloor}\left|\log\left(\frac{\theta}{\delta}+1\right)-\log\left(\frac{\lfloor\delta k_n\rfloor+\lfloor\theta k_n\rfloor}{\lfloor\delta k_n\rfloor}\right)\right|+\left|\log\left(\frac{\lfloor\delta k_n\rfloor+\lfloor\theta k_n\rfloor}{\lfloor\delta k_n\rfloor}\right)-\sum_{j=\lfloor\delta k_n\rfloor+1}^{\lfloor\delta k_n\rfloor+\lfloor\theta k_n\rfloor}\frac{1}{j}\right|,
\end{aligned}
$$

where

$$
\left|\log\left(\frac{\theta}{\delta}+1\right)-\log\left(\frac{\lfloor\delta k_n\rfloor+\lfloor\theta k_n\rfloor}{\lfloor\delta k_n\rfloor}\right)\right|\leq(1+o(1))\left(\frac{2}{\lfloor\delta k_n\rfloor+\lfloor\theta k_n\rfloor}+\frac{1}{\delta k_n}\right)
$$

and

$$
\left|\log\left(\frac{\lfloor\delta k_n\rfloor+\lfloor\theta k_n\rfloor}{\lfloor\delta k_n\rfloor}\right)-\sum_{j=\lfloor\delta k_n\rfloor+1}^{\lfloor\delta k_n\rfloor+\lfloor\theta k_n\rfloor}\frac{1}{j}\right|\leq\frac{1}{\lfloor\delta k_n\rfloor}-\frac{1}{\lfloor\delta k_n\rfloor+\lfloor\theta k_n\rfloor}
$$

by the error bound of the Riemann sum. Therefore (32) converges to zero uniformly. If $\delta k_n < 1$, then for $k_n \geq 1/m$,

$$\sqrt{k_n}\frac{\delta}{\theta}\log\left(\frac{\theta}{\delta}+1\right) < \frac{\sqrt{\delta}}{\theta}\log\left(\frac{\theta}{\delta}+1\right)$$

and thus converges to zero uniformly as $n \to \infty$.

Next we show (28) converges to zero uniformly in probability. Since $|A_0| \in \mathrm{RV}(\rho)$, by Potter's inequalities (de Haan and Ferreira 2006), for any $\tilde{\epsilon} > 0$, there exists $\tilde{t}_0 > 0$, such that whenever $n/(\delta k_n + \theta k_n) > \tilde{t}_0$ and $Y_{(n-\lfloor\theta k_n\rfloor-\lfloor\theta k_n\rfloor)} > \tilde{t}_0$,

$$(1-\tilde{\epsilon})(\tilde{A}_{\delta,\theta}^{\rho+\tilde{\epsilon}} \wedge \tilde{A}_{\delta,\theta}^{\rho-\tilde{\epsilon}}) < \frac{A_0(Y_{(n-\lfloor\delta k_n\rfloor-\lfloor\theta k_n\rfloor)})}{A_0(\frac{n}{\delta k_n+\theta k_n})} < (1+\tilde{\epsilon})(\tilde{A}_{\delta,\theta}^{\rho+\tilde{\epsilon}} \vee \tilde{A}_{\delta,\theta}^{\rho-\tilde{\epsilon}}), \tag{33}$$

where

$$\tilde{A}_{\delta,\theta} = \frac{\delta k_n + \theta k_n}{n} Y_{(n-\lfloor\delta k_n\rfloor-\lfloor\theta k_n\rfloor)}.$$

By Lemma 2.4.10 of de Haan and Ferreira (2006), for all $\beta > 0$, given $n \to \infty$, $K \to \infty$ and $K/n \to 0$,

$$\sup_{K^{-1}\leq s\leq 1} s^{1/2+\beta}\left|\sqrt{K}\left(\frac{Ks}{n}Y_{(n-[Ks])}-1\right) - \frac{B_n(s)}{s}\right| = o_P(1), \tag{34}$$

where $\{B_n(s)\}$ is a sequence of Brownian motions. Let $K = 2Mk_n$, $s = (\lfloor\delta k_n\rfloor + \lfloor\theta k_n\rfloor)/K$, then $(mk_n - 1)/(2Mk_n) \leq s \leq 1$. Consider k_n large enough such that $(mk_n - 1)/(2Mk_n) \geq m/3M > K^{-1}$, then for all $0 < \epsilon' < \epsilon$,

$$\begin{aligned}
&P\left(\sup_{\delta,\theta}\left|\frac{\lfloor\delta k_n\rfloor + \lfloor\theta k_n\rfloor}{n}Y_{(n-\lfloor\delta k_n\rfloor-\lfloor\theta k_n\rfloor)} - 1\right| > \epsilon\right) \\
&\leq P\left(\sup_{m/3M\leq s\leq 1}\sqrt{K}\left|\frac{Ks}{n}Y_{(n-[Ks])} - 1\right| > \epsilon\sqrt{K}\right) \\
&\leq P\left(\sup_{m/3M\leq s\leq 1}\left|\frac{B_n(s)}{s}\right| > \epsilon\sqrt{K} - \epsilon'\right)
\end{aligned}$$

$$+ P\left(\sup_{m/3M\le s\le 1}\left|\sqrt{K}\left(\frac{Ks}{n}Y_{(n-[Ks])}-1\right)-\frac{B_n(s)}{s}\right|\ge \epsilon'\right).$$

Since $\sup_{m/3M\le s\le 1}|B(s)|<\infty, a.s.$, as $K\to\infty$,

$$P\left(\sup_{m/3M\le s\le 1}\left|\frac{B_n(s)}{s}\right|>\epsilon\sqrt{K}-\epsilon'\right)\le P\left(\sup_{m/3M\le s\le 1}|B_n(s)|>(\epsilon\sqrt{K}-\epsilon')\frac{m}{3M}\right)\to 0.$$

By (34),

$$\begin{aligned}&P\left(\sup_{m/3M\le s\le 1}\left|\sqrt{K}\left(\frac{Ks}{n}Y_{(n-[Ks])}-1\right)-\frac{B_n(s)}{s}\right|\ge \epsilon'\right)\\ &\le P\left(\sup_{m/3M\le s\le 1}s^{1/2+\beta}\left|\sqrt{K}\left(\frac{Ks}{n}Y_{(n-[Ks])}-1\right)-\frac{B_n(s)}{s}\right|\ge \left(\frac{m}{3M}\right)^{1/2+\beta}\epsilon'\right)\to 0.\end{aligned}$$

Since $(\lfloor\delta k_n\rfloor+\lfloor\theta k_n\rfloor)/(\delta k_n+\theta k_n)\to 1$ uniformly, it follows that

$$\sup_{\delta,\theta}\left|\frac{\delta k_n+\theta k_n}{n}Y_{(n-\lfloor\delta k_n\rfloor-\lfloor\theta k_n\rfloor)}-1\right|=o_P(1).$$

Therefore by (33),

$$\sup_{\delta,\theta}\sqrt{k_n}\left|\frac{A_0(Y_{(n-\lfloor\delta k_n\rfloor-\lfloor\theta k_n\rfloor)})}{A_0\left(\frac{n}{\delta k_n+\theta k_n}\right)}\right|=1+o_P(1),$$

where

$$\sqrt{k_n}\left|A_0\left(\frac{n}{\delta k_n+\theta k_n}\right)\right|\sim\sqrt{k_n}\left|A_0\left(\frac{n}{k_n}\right)\right|(\delta+\theta)^{-\rho}\to\frac{\lambda}{(\delta+\theta)^{\rho}}$$

uniformly in (δ,θ) (de Haan and Ferreira 2006). Similarly,

$$\sup_{\delta,\theta}\left|\frac{1}{\lfloor\theta k_n\rfloor}\sum_{i=1}^{\lfloor\theta k_n\rfloor}\left[\left(\frac{Y_{(n-\lfloor\delta k_n\rfloor-i+1)}}{Y_{(n-\lfloor\delta k_n\rfloor-\lfloor\theta k_n\rfloor)}}\right)^{\rho}-1\right]-\frac{1}{\lfloor\theta k_n\rfloor}\sum_{i=1}^{\lfloor\theta k_n\rfloor}\left[\left(\frac{\delta k_n+\theta k_n}{\delta k_n+i-1}\right)^{\rho}-1\right]\right|=o_P(1),$$

and the Riemann sum

$$\frac{1}{\lfloor\theta k_n\rfloor}\sum_{i=1}^{\lfloor\theta k_n\rfloor}\left[\left(\frac{\delta k_n+\theta k_n}{\delta k_n+i-1}\right)^{\rho}-1\right]\to\int_0^1\left(\frac{\delta/\theta+1}{\delta/\theta+x}\right)^{\rho}dx-1,$$

which is

$$\frac{1 + (\theta/\delta)\rho - (\theta/\delta + 1)^{\rho}}{(\theta/\delta)(1-\rho)}$$

if $\delta > 0$ and $\rho/(1-\rho)$ if $\delta = 0$. The error bounded is given by

$$\frac{1}{\lfloor \theta k_n \rfloor}\left[1 - \left(\frac{\delta/\theta + 1}{\delta/\theta}\right)^{\rho}\right] \leq \frac{1}{\lfloor \theta k_n \rfloor},$$

which converges to zero uniformly.

It can be shown along the same lines for (29) that

$$\sup_{\delta,\theta}\left|\frac{1}{\lfloor \theta k_n \rfloor}\sum_{i=1}^{\lfloor \theta k_n \rfloor}\left(\frac{Y_{(n-\lfloor \delta k_n \rfloor - i+1)}}{Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)}}\right)^{\rho+\epsilon} - \tilde{b}_{\rho,\epsilon}(\delta,\theta)\right| = o_P(1),$$

where

$$\tilde{b}_{\rho,\epsilon}(\delta,\theta) = \frac{(\theta/\delta + 1) - (\theta/\delta + 1)^{\rho+\epsilon}}{(\theta/\delta)(1-\rho-\epsilon)}$$

if $\delta > 0$ and $1/(1-\rho-\epsilon)$ if $\delta = 0$, which is bounded on $(\delta,\theta) \in [0, M] \times [m, M]$ when ϵ is small enough. Therefore, (29) converges to zero uniformly in probability when $\epsilon \to 0$.

Since

$$P\Big(\min_{(\delta,\theta)\in[0,M]\times[m,M]} Y_{(n-\lfloor \delta k_n \rfloor - \lfloor \theta k_n \rfloor)} \geq t_0\Big) \to 1$$

as $n \to \infty$, given the convergence results for (27)-(29), we have that for all $\tilde{\epsilon} > 0$,

$$P\left(\sup_{(\delta,\theta)\in[0,M]\times[m,M]}\left|\sqrt{k_n}\big(\alpha H_n(\delta,\theta) - g(\delta,\theta)\big) - \alpha b_{\rho}(\delta,\theta) - W_n(\delta,\theta)\right| > \tilde{\epsilon}\right) \to 0.$$

By Lemma (5.1) and Lemma (5.2), $W_n(\cdot,\cdot) \xrightarrow{\text{d}} G(\cdot,\cdot)$ (Bickel and Wichura 1971), and the desired weak convergence follows.

□

SUPPLEMENTARY MATERIAL

Technical proofs: Detailed proof of Lemma 5.2.

R Code for simulations and real data examples: Code for R algorithms used to produce illustrations in Sections 1 and 2 and estimation results in 4 and 5. (.r files)

Earthquake fatality data set: Data set used in the illustration in section 5. (comma-separated values (CSV) file)

References

Aban, I. B., Meerschaert, M. M., and Panorska, A. K. (2006). Parameter Estimation for the Truncated Pareto Distribution. *Journal of the American Statistical Association*, 101(473):270–277.

Beirlant, J., Alves, I. F., and Gomes, I. (2016a). Tail Fitting for Truncated and Non-Truncated Pareto-Type Distributions. *Extremes*, 19(3):429–462.

Beirlant, J., Alves, I. F., and Reynkens, T. (2016b). Fitting Tails Affected by Truncation. *arXiv.org*, page arXiv:1606.02090.

Bickel, P. J. and Wichura, M. J. (1971). Convergence Criteria for Multiparameter Stochastic Processes and Some Applications. *Annals of Mathematical Statistics*, 42(5):1656–&.

Burroughs, S. M. and Tebbens, S. F. (2001a). Upper-Truncated Power Law Distributions. *Fractals*, 09(02):209–222.

Burroughs, S. M. and Tebbens, S. F. (2001b). Upper-Truncated Power Laws in Natural Systems. *Pure and Applied Geophysics*, 158(4):741–757.

Burroughs, S. M. and Tebbens, S. F. (2002). The Upper-Truncated Power Law Applied to Earthquake Cumulative Frequency-Magnitude Distributions: Evidence for a Time-Independent Scaling Parameter. *Bulletin of the Seismological Society of America*, 92(8):2983–2993.

Clark, D. R. (2013). A Note on the Upper-Truncated Pareto Distribution. *Casualty Actuarial Society E-Forum*.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York.

Drees, H. (1998). On Smooth Statistical Tail Functionals. *Scandinavian Journal of Statistics*, 25(1):187–210.

Drees, H., de Haan, L., and Resnick, S. (2000). How to Make a Hill Plot. *The Annals of Statistics*, 28(1):254–274.

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, 3(5):1163–1174.

Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.

Straf, M. L. (1972). Weak Convergence of Stochastic Processes With Several Parameters. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 187–221. Univ. California Press, Berkeley, Calif.