ESTIMATING THE FIRST AND SECOND ORDER PARAMETERS OF A HEAVY TAILED DISTRIBUTION

LIANG PENG[1] AND YONGCHENG QI[2*]

Georgia Institute of Technology and University of Minnesota Duluth

Summary

We suggest censored maximum likelihood estimators for the first and second order parameters of a heavy tailed distribution by incorporating the second order regular variation into the censored likelihood function. This approach is different from the bias-reduced MLE proposed by Feuerverger & Hall (1999). In comparison with Feuerverger & Hall (1999), we derive the joint asymptotic limit for the first and second order parameters under a weaker assumption. We also demonstrate through a simulation study that our estimator for the first order parameter is better than that proposed by Feuerverger & Hall (1999) although these two estimators have the same asymptotic variances.

Key words: Bias, censored likelihood function, Hill estimator, second order regular variation, tail index.

1. Introduction

In order to estimate high quantiles or extreme tail probabilities of an unknown distribution function, we have to estimate beyond the observations, so extra assumptions on the underlying distribution function are needed. One approach is to assume that the underlying distribution has a heavy tail; see Hall & Weissman (1997), Danielsson, Hartman & de Vries (1998), Danielsson & de Vries (1997), and Embrechts, Resnick & Samorodnitsky (1998). Thus, estimating the

* Author to whom correspondence should be addressed.

[1]School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA. Email: peng@math.gatech.edu

[2]Department of Mathematics and Statistics, University of Minnesota Duluth, CCTR 140, 1117 University Drive Duluth, MN 55812, USA. Email: yqi@d.umn.edu. Fax: (218) 7268399

tail index of a heavy tailed distribution is of both practical and methodological importance, and many different estimators have been proposed. See, e.g., Hill (1975), Hall (1982b), Csörgő, Deheuvels & Mason (1985), Csörgő & Viharos (1997), and de Haan & Peng (1998). Since we make inference about the tail quantity, we can only employ upper k order statistics of a sample size n, where $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. When k is small, the variance of the tail index estimator is large. However, the use of large k will introduce a big bias in the estimation, so the choice of k plays an important role. Recently, several procedures have been proposed to choose the optimal k in the sense of asymptotic minimal mean squared error; see Hall (1990), Dekkers & de Haan (1993), Beirlant et al. (1996), Drees & Kaufmann (1998), and Danielsson et al. (2001). Since the optimal choice of k depends on the second order regular variation parameter, which is usually hard to estimate accurately, some new estimators are proposed to reduce the bias term (see Beirlant et al. (1999) and Guillou & Hall (2001)).

Suppose $X_1, \ldots, X_n$ are i.i.d. random variables with common distribution function F which satisfies

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\alpha}, \quad x > 0, \tag{1}$$

where $\alpha > 0$ is termed as the tail index or first order regular variation parameter. One of the well-known estimators for the index α is Hill estimator (Hill, 1975) defined as

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

where $X_{n,1} \le \cdots \le X_{n,n}$ denote the order statistics of the random variables $X_1, \cdots, X_n$. Let $Y_i = i\log(X_{n,n-i+1}/X_{n,n-i})$ for $i = 1, \cdots, k$. Then it can be shown that, for any fixed $k \ge 1$, as $n \to \infty$

$$(Y_i, 1 \le i \le k) \xrightarrow{d} (\alpha^{-1}W_i, 1 \le i \le k),$$

where the $(W_i, 1 \le i \le k)$ are independent exponential random variables with mean one (see, e.g, Weissman (1978)). Therefore, the Hill estimator can be viewed as the asymptotic sample mean of the random variables $(\alpha^{-1}W_i, 1 \le i \le k)$. For the consistency of $\hat{\alpha}_H(k)$ we refer to Mason (1982). In order to derive the asymptotic normality of $\hat{\alpha}_H(k)$, we need a stricter condition than (1). Suppose that as $x \to \infty$,

$$1 - F(x) = cx^{-\alpha} + dx^{-\beta} + o(x^{-\beta}), \tag{2}$$

where $c > 0$, $d \ne 0$, and $\beta > \alpha > 0$. Here β is called the second order regular variation parameter. Note that (2) is a special case of the general second order regular variation (see de Haan & Stadtmüller (1996)). Under condition (2) it can be shown that, if $\sqrt{k}(n/k)^{1-\beta/\alpha} \to \lambda \in [0, \infty)$,

$$\sqrt{k}(\hat{\alpha}_H(k) - \alpha) \xrightarrow{d} N\big(\lambda\alpha\beta^{-1}(\beta-\alpha)dc^{-\beta/\alpha}, \alpha^2\big), \tag{3}$$

(see Hall & Welsh (1985) or de Haan & Peng (1998)). Hence the optimal choice of sample fraction is

$$k^* = \big(2^{-1}\alpha\beta^2(\beta-\alpha)^{-3}d^{-2}c^{2\beta/\alpha}\big)^{\frac{\alpha}{2\beta-\alpha}} n^{\frac{2\beta-2\alpha}{2\beta-\alpha}} \tag{4}$$

in the sense of minimal asymptotic mean squared error of the Hill estimator. By noting that

$$Y_i \sim W_i\alpha^{-1}\exp\big(D_1(\frac{i}{n})^{\beta_1}\big),$$

where $\beta_1 = \beta/\alpha - 1$ and $D_1 = (1-\beta/\alpha)c^{-\beta/\alpha}d$, Feuerverger & Hall (1999) regarded Y_i as exponential with mean $\alpha^{-1}\exp\big(D_1(i/n)^{\beta_1}\big)$ rather than α^{-1}, and then estimated α, D_1, β_1 by the maximum likelihood method. This results in the estimator

$$\hat{\alpha}_{FH}(k) = \Big(k^{-1}\sum_{i=1}^{k} Y_i \exp\big(-D_1(\frac{i}{n})^{\beta_1}\big)\Big)^{-1}, \tag{5}$$

where (D_1, β_1) was chosen to minimize

$$L_1(D_1,\beta_1) = k^{-1}D_1\sum_{i=1}^{k}(\frac{i}{n})^{\beta_1} + \log\Big(k^{-1}\sum_{i=1}^{k} Y_i\exp\big(-D_1(\frac{i}{n})^{\beta_1}\big)\Big). \tag{6}$$

This approach reduces bias by an order of magnitude without inflating the order of variance.

The determination of the optimal sample fraction k^* in (4) depends on both the first and the second order parameters, α and β, of the underlying distribution (2). Thus the estimation of the second order parameter is also desired in practice. In this paper we first derive the Hill estimator as the maximum likelihood estimator for left censored data, rather than based on an asymptotic exponential distribution, and then we can incorporate the second order regular variation into the censored likelihood, which introduces new estimators for the first and second order parameters. This new procedure allows a simultaneous estimation for both α and β, and permits a larger range of sample fraction for the new estimator of the first order parameter α without introducing any bias. The detailed methodology and main results are given in Section 2, which shows that our new estimator for α has the same asymptotic variance as $\hat{\alpha}_{FH}(k)$ defined in (5). In comparison with Feuerverger & Hall (1999), we are able to derive the joint asymptotic distribution for estimators of the first and second order parameters under a weaker assumption. A simulation study and a real application are presented in Section 3, where our new estimator for the first order parameter is shown to have a better performance than $\hat{\alpha}_{FH}(k)$ in Feuerverger & Hall (1999), although these two estimators have the same asymptotic variances. All proofs are deferred till Appendix.

2. Methodology

Let $T = T(n) \to \infty$ as $n \to \infty$, and define $\delta_i = I(X_i > T)$ for $i = 1, \cdots, n$. Since we can only use a part of upper order statistics to make inference, we view our observations as

$\big((X_i \vee T, \delta_i), 1 \le i \le n\big)$ instead of $(X_i, 1 \le i \le n)$. If we approximate $1 - F(x)$ by $cx^{-\alpha}$ as $x > T$, then the likelihood for $\big((X_i \vee T, \delta_i), 1 \le i \le n\big)$ is approximately

$$L(\alpha, c) = \Pi_{i=1}^n (c\alpha X_i^{-\alpha-1})^{\delta_i}(1 - cT^{-\alpha})^{1-\delta_i}.$$

Hence we have

$$(\tilde{\alpha}, \tilde{c}) = \text{argmax}_{\alpha>0,c>0} L(\alpha, c) = \Big(\frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n \delta_i(\log X_i - \log T)}, n^{-1}T^{\tilde{\alpha}}\sum_{i=1}^n \delta_i\Big).$$

So if T is chosen as $X_{n,n-k}$, then $\tilde{\alpha}$ becomes the Hill estimator $\hat{\alpha}_H(k)$. A somewhat similar approach was used by Hall (1982a) to derive the MLE for the endpoint of a distribution.

Next let us approximate $1 - F(x)$ by $cx^{-\alpha} + dx^{-\beta}$ as $x > T$. Then the likelihood for $\big((X_i \vee T, \delta_i), 1 \le i \le n\big)$ is approximately

$$L(\alpha, c, \beta, d) = \Pi_{i=1}^n (c\alpha X_i^{-\alpha-1} + d\beta X_i^{-\beta-1})^{\delta_i}(1 - cT^{-\alpha} - dT^{-\beta})^{1-\delta_i}.$$

Therefore our new estimators can be obtained as

$$(\bar{\alpha}, \bar{c}, \bar{\beta}, \bar{d}) = \text{argmax}_{(\alpha>0,c>0,\beta>\alpha,d\neq 0)} L(\alpha, c, \beta, d),$$

i.e., $(\bar{\alpha}, \bar{c}, \bar{\beta}, \bar{d})$ is the solution of the following equations (7) – (10):

$$\frac{\partial \log L(\alpha, c, \beta, d)}{\partial \alpha} = \sum_{i=1}^n \frac{\delta_i(cX_i^{-\alpha-1} - c\alpha X_i^{-\alpha-1}\log X_i)}{c\alpha X_i^{-\alpha-1} + d\beta X_i^{-\beta-1}} + \frac{cT^{-\alpha}\log T}{1 - cT^{-\alpha} - dT^{-\beta}}\sum_{i=1}^n (1-\delta_i) = 0, \quad (7)$$

$$\frac{\partial \log L(\alpha, c, \beta, d)}{\partial c} = \sum_{i=1}^n \frac{\delta_i \alpha X_i^{-\alpha-1}}{c\alpha X_i^{-\alpha-1} + d\beta X_i^{-\beta-1}} - \frac{T^{-\alpha}}{1 - cT^{-\alpha} - dT^{-\beta}}\sum_{i=1}^n (1-\delta_i) = 0, \quad (8)$$

$$\frac{\partial \log L(\alpha, c, \beta, d)}{\partial \beta} = \sum_{i=1}^n \frac{\delta_i(dX_i^{-\beta-1} - d\beta X_i^{-\beta-1}\log X_i)}{c\alpha X_i^{-\alpha-1} + d\beta X_i^{-\beta-1}} + \frac{dT^{-\beta}\log T}{1 - cT^{-\alpha} - dT^{-\beta}}\sum_{i=1}^n (1-\delta_i) = 0, \quad (9)$$

$$\frac{\partial \log L(\alpha, c, \beta, d)}{\partial d} = \sum_{i=1}^n \frac{\delta_i \beta X_i^{-\beta-1}}{c\alpha X_i^{-\alpha-1} + d\beta X_i^{-\beta-1}} - \frac{T^{-\beta}}{1 - cT^{-\alpha} - dT^{-\beta}}\sum_{i=1}^n (1-\delta_i) = 0 \quad (10)$$

under the constraints

$$\beta > \alpha > 0, \quad c > 0, \quad d \neq 0. \quad (11)$$

It follows from (7) – (10) that

$$\bar{c} = \frac{\alpha\beta T^{\alpha}}{\alpha - \beta}\Big(\frac{\sum_{i=1}^n \delta_i}{n\beta} - \frac{\sum_{i=1}^n \delta_i(\log X_i - \log T)}{n}\Big) \quad (12)$$

and

$$\bar{d} = \frac{\alpha\beta T^\beta}{\beta-\alpha}\Big(\frac{\sum_{i=1}^n \delta_i}{n\alpha} - \frac{\sum_{i=1}^n \delta_i(\log X_i - \log T)}{n}\Big). \tag{13}$$

See Appendix for detail. Write

$$\begin{cases} \bar{H}(\alpha) = \frac{\sum_{i=1}^n \delta_i}{n\alpha} - \frac{\sum_{i=1}^n \delta_i \log(X_i/T)}{n} \\ \bar{Q}_i(\alpha,\beta) = \alpha\big(\frac{\sum_{i=1}^n \delta_i}{n} + \frac{\alpha\beta}{\alpha-\beta}\bar{H}(\alpha)\big)(\frac{X_i}{T})^{\beta-\alpha} - \frac{\alpha\beta^2}{\alpha-\beta}\bar{H}(\alpha). \end{cases}$$

Substituting (12) and (13) into (10), we obtain

$$\frac{1}{n}\sum_{i=1}^n \delta_i \bar{Q}_i^{-1}(\alpha,\beta) = \beta^{-1}. \tag{14}$$

Substituting (12) and (13) into (9) and using (14) we have

$$\frac{1}{n}\sum_{i=1}^n \delta_i \bar{Q}_i^{-1}(\alpha,\beta)\log\frac{X_i}{T} = \beta^{-2}. \tag{15}$$

For simplicity, take $T = X_{n,n-k}$ and define

$$\begin{cases} H(\alpha) = \frac{1}{\alpha} - \frac{1}{k}\sum_{i=1}^k \log\frac{X_{n,n-i+1}}{X_{n,n-k}} \\ Q_i(\alpha,\beta) = \frac{\alpha}{\beta}\big(1 + \frac{\alpha\beta}{\alpha-\beta}H(\alpha)\big)(\frac{X_{n,n-i+1}}{X_{n,n-k}})^{\beta-\alpha} - \frac{\alpha\beta}{\alpha-\beta}H(\alpha). \end{cases}$$

Thus (14) and (15) become

$$\frac{1}{k}\sum_{i=1}^k Q_i^{-1}(\alpha,\beta) = 1 \tag{16}$$

and

$$\frac{1}{k}\sum_{i=1}^k Q_i^{-1}(\alpha,\beta)\log\frac{X_{n,n-i+1}}{X_{n,n-k}} = \beta^{-1}. \tag{17}$$

Put

$$\beta > \alpha > 0 \quad \text{and} \quad \beta > \hat{\alpha}_H(k). \tag{18}$$

The reason why we confine $\beta > \hat{\alpha}_H(k)$ is that $\beta = \hat{\alpha}_H(k)$ is an obvious solution to (16) for any fixed $\alpha < \hat{\alpha}_H(k)$. Note that for any fixed $\beta > \alpha_0$, where α_0 is the true parameter,

$$\begin{cases} \frac{1}{k}\sum_{i=1}^k Q_i^{-1}(\hat{\alpha}_H(k),\beta) \xrightarrow{p} 1 \\ \frac{1}{k}\sum_{i=1}^k Q_i^{-1}(\hat{\alpha}_H(k),\beta)\log\frac{X_{n,n-i+1}}{X_{n,n-k}} \xrightarrow{p} \beta^{-1} \end{cases}$$

as $k \to \infty$, $k/n \to 0$.

Let $U(x)$ denote the inverse function of $1/\big(1-F(x)\big)$. Then (2) implies that for any $x > 0$

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

i.e.,

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

where

$$A(t)=-\alpha^{-2}(\beta-\alpha)dc^{-\beta/\alpha}t^{1-\beta/\alpha}. \tag{19}$$

Suppose there exists a function $B(t)\to 0$, with constant sign near infinity, such that

$$\lim_{t\to\infty}\frac{\frac{\log U(tx)-\log U(t)-\alpha^{-1}\log x}{A(t)}-\frac{x^{1-\beta/\alpha}-1}{1-\beta/\alpha}}{B(t)}=\frac{1}{\rho}(\frac{x^{1-\beta/\alpha+\rho}-1}{1-\beta/\alpha+\rho}-\frac{x^{1-\beta/\alpha}-1}{1-\beta/\alpha})=h(x), \tag{20}$$

where $\rho\le 0$ may be called the third order regular variation parameter.

Our main result is as follows.

Theorem 1. *Suppose (20) holds with true parameters $\alpha_0>0$ and $\beta_0>\alpha_0$, and suppose $k=k(n)$ satisfies*

$$k\to\infty, \sqrt{k}|A(\frac{n}{k})|\to\infty, \sqrt{k}A^2(\frac{n}{k})\to 0, \sqrt{k}|A(\frac{n}{k})B(\frac{n}{k})|\to 0, \tag{21}$$

as $n\to\infty$. Assume there exists a solution to (16) - (18), say $(\hat\alpha(k),\hat\beta(k))$. Then

$$\Big(\sqrt{k}\big(\hat\alpha(k)-\alpha_0\big), \sqrt{k}A(\frac{n}{k})\big(\hat\beta(k)-\beta_0\big)\Big)\xrightarrow{d}(N_1,N_2),$$

where (N_1,N_2) is a bivariate normal random vector with $E(N_1)=E(N_2)=0$, $E(N_1^2)=\alpha_0^2\beta_0^4/(\beta_0-\alpha_0)^4$, $E(N_2^2)=\alpha_0(\beta_0-\alpha_0)^2/\big(\beta_0^2(2\beta_0-\alpha_0)\big)$, and $E(N_1N_2)=\alpha_0^2/(\beta_0-\alpha_0)$.

Remark 1. The condition $\sqrt{k}|A(n/k)|\to\infty$ ensures that there exists a consistent solution to (16) - (18) (See the proof of Theorem 1 in Appendix). We suspect that the theorem in Feuerverger & Hall (1999) requires the consistency of the estimator of β_0 since the expansion $D(i/n)^{\beta_1}$ in Feuerverger & Hall (1999, p.776) requires that $(\beta_1-\beta_1^0)\log(i/n)\to 0$ uniformly for $i=1,\cdots,k$, where $\beta_1=\beta/\alpha-1$ and $\beta_1^0=\beta_0/\alpha_0-1$.

Remark 2. By a tedious calculation we found that the complicated variance for $\hat\alpha_{FH}(k)$ given by Feuerverger & Hall (1999) is exactly $E(N_1^2)$, i.e., our new estimator $\hat\alpha(k)$ has the same asymptotic variance as $\hat\alpha_{FH}(k)$. Note that condition (4.1) in Feuerverger & Hall (1999) is slightly stronger than our condition (20). On the other hand, we expect that $\hat\alpha(k)$ behaves better than $\hat\alpha_{FH}(k)$ since $\hat\alpha(k)$ is based on censored likelihood function rather than an approximate exponential distribution like $\hat\alpha_{FH}(k)$. This is confirmed in Section 3.

Remark 3. Note that Feuerverger & Hall (1999) did not give the asymptotic variance for estimating the second order parameter β. There are a few consistent estimators for β in the literature, but as far as we know, no asymptotic properties for them are established. Remember that our estimator for β is a sort of maximum likelihood estimator, so it may be considered to be efficient.

Remark 4. In the case $\sqrt{k}A^2(n/k) \to \lambda_1 \in [0, \infty)$ and $\sqrt{k}A(n/k)B(n/k) \to \lambda_2 \in (-\infty, \infty)$, we could show, by a refinement of the proof of Theorem 1 in Appendix, that the limit in Theorem 1 has a bias term.

Remark 5. In comparison with the Hill estimator, the same conclusions as in Feuerverger & Hall (1999) can be drawn, i.e., our new estimator $\hat{\alpha}(k)$ allows to use a larger number of sample fraction k.

3. Simulation study and real application

3.1. Simulation study. In this subsection we report a simulation study which examined the finite sample properties of our estimator $\hat{\alpha}(k)$, and compare it with the $\hat{\alpha}_{FH}(k)$ proposed by Feuerverger & Hall (1999).

We generated 200 pseudorandom samples of size $n = 1000$ from one of the following two distributions (i) Burr(α, β) distribution, given by $F(x) = 1 - (1 + x^{\beta-\alpha})^{-\alpha/(\beta-\alpha)}$ $(x > 0)$; (ii) Fréchet(α) distribution, given by $F(x) = \exp(-x^{-\alpha})$ $(x > 0)$.

First we compare our estimator $\hat{\alpha}(k)$ with $\hat{\alpha}_{FH}(k)$ by employing a practical choice of $k = \min([k^* \log(k^*)/2], n/2)$ with the theoretical optimal value of k^* given in (4) for distributions Burr(0.5,0.7), Burr(0.5,1.0), Burr(2.0,3.0), Burr(2.0, 4.0), Fréchet(0.5) and Fréchet(2.0); see Table 1. Here we use the theoretical value of k^* rather than estimated value, since we will investigate the effect of the choice of sample fraction in our next comparison. Second, we compare $\hat{\alpha}(k)$ with $\hat{\alpha}_{FH}(k)$ by employing different choice of sample fraction for distributions Burr(0.5,0.7) and Burr(2.0,3.0); see Table 2. We can conclude from Tables 1 and 2 that our new estimator $\hat{\alpha}(k)$ is better than $\hat{\alpha}_{FH}(k)$, although both estimators have the same asymptotic variance. The reason is that our new estimator is based on censored likelihood function rather than an approximate exponential distribution like $\hat{\alpha}_{FH}(k)$.

Table 1: *Comparison by employing a practical choice of sample fraction.* Estimators $\hat{\alpha}(k)$ and $\hat{\alpha}_{FH}(k)$ are computed with $k = \min([k^* \log(k^*)/2], n/2)$, where k^* is given in (4). The corresponding standard errors are given between brackets.

	k^*	k	$\hat{\alpha}(k)$	$\hat{\alpha}_{FH}(k)$
Burr(0.5,0.7)	35	62	0.4099(0.0901)	0.3923(0.1189)
Burr(0.5,1.0)	125	301	0.4839(0.0616)	0.4809(0.0903)
Burr(2.0,3.0)	47	90	1.7268(0.3285)	1.6936(0.5064)
Burr(2.0,4.0)	125	301	1.9468(0.2505)	1.9240(0.3710)
Fréchet(0.5)	199	500	0.4930(0.0552)	0.4949(0.0765)
Fréchet(2.0)	199	500	1.9821(0.2203)	1.9805(0.3069)

Table 2: *Comparison by employing different choice of sample fraction.* Estimators $\hat{\alpha}(k)$ and $\hat{\alpha}_{FH}(k)$ are computed for $k = k^* + 10 * i$, $i = 1, \ldots, 10$, where k^* is given in (4). The corresponding standard errors are given between brackets.

k	$\hat{\alpha}(k)$ for Burr(0.5,0.7)	$\hat{\alpha}_{FH}(k)$ for Burr(0.5,0.7)	$\hat{\alpha}(k)$ for Burr(2.0,3.0)	$\hat{\alpha}_{FH}(k)$ for Burr(2.0,3.0)
$k^* + 10$	0.4275(0.1210)	0.4017(0.1334)	1.7558(0.4741)	1.6695(0.5359)
$k^* + 20$	0.4214(0.1015)	0.3990 (0.1206)	1.7469(0.4147)	1.6617(0.5359)
$k^* + 30$	0.4101(0.0877)	0.3906(0.1154)	1.8000(0.3873)	1.7263(0.4859)
$k^* + 40$	0.4125(.0784)	0.3959(0.1049)	1.7583(0.3366)	1.7350(0.4783)
$k^* + 50$	0.4081(0.0751)	0.3872(0.1074)	1.7428(0.3165)	1.6654(0.4165)
$k^* + 60$	0.4080(0.0688)	0.35956(0.0977)	1.7111(0.3038)	1.6514(0.4036)
$k^* + 70$	0.4150(0.0800)	0.3925(0.0961)	1.7084(0.2768)	1.6722(0.3903)
$k^* + 80$	0.4037(0.0560)	0.3902(0.0861)	1.7156(0.2866)	1.6637(0.3977)
$k^* + 90$	0.4019(0.0609)	0.3903(0.0926)	1.7298(0.2633)	1.6583(0.3281)
$k^* + 100$	0.3956(0.0570)	0.3856(0.0871)	1.7126(0.2518)	1.6682(0.3590)

3.2. Real application. The data set we shall analyze consists of 2156 Danish fire losses of over one million Danish Krone (DKK) from the years 1980 to 1990 inclusive (see Figure 1). The loss figure is a total loss figure for the event concerned, and includes damage to buildings and damage to furnish and personal property, as well as loss of profits. This Danish fire data set was analyzed by McNeil (1997). We compute $\hat{\alpha}(k)$ and $\hat{\alpha}_{FH}(k)$ for $k = 50 + i * 5$, $i = 1, \ldots, 100$; see Figure 2. We observe from Figure 2 that our new estimator $\hat{\alpha}(k)$ is much more robust than $\hat{\alpha}_{FH}(k)$ as the sample fraction k becomes large.

Appendix

A. Derivation of (12) and (13). It follows from (7)+(9) that

$$\begin{aligned} & c\sum_{i=1}^n \frac{\delta_i X_i^{-\alpha-1}}{c\alpha X_i^{-\alpha-1}+d\beta X_i^{-\beta-1}} + d\sum_{i=1}^n \frac{\delta_i X_i^{-\beta-1}}{c\alpha X_i^{-\alpha-1}+d\beta X_i^{-\beta-1}} - \sum_{i=1}^n \delta_i \log X_i \\ & + \frac{cT^{-\alpha}+dT^{-\beta}}{1-cT^{-\alpha}-dT^{-\beta}} \log T \sum_{i=1}^n (1-\delta_i) = 0. \end{aligned} \tag{22}$$

By $(8) * c/\alpha + (10) * d/\beta$, we have

$$\begin{aligned} & c\sum_{i=1}^n \frac{\delta_i X_i^{-\alpha-1}}{c\alpha X_i^{-\alpha-1}+d\beta X_i^{-\beta-1}} + d\sum_{i=1}^n \frac{\delta_i X_i^{-\beta-1}}{c\alpha X_i^{-\alpha-1}+d\beta X_i^{-\beta-1}} \\ & - \frac{T^{-\alpha}c/\alpha+T^{-\beta}d/\beta}{1-cT^{-\alpha}-dT^{-\beta}} \sum_{i=1}^n (1-\delta_i) = 0. \end{aligned} \tag{23}$$

By $(22) - (23)$ we have

$$\begin{aligned} & -\sum_{i=1}^n \delta_i \log X_i + \frac{cT^{-\alpha}+dT^{-\beta}}{1-cT^{-\alpha}-dT^{-\beta}} \log T \sum_{i=1}^n (1-\delta_i) \\ & + \frac{T^{-\alpha}c/\alpha+T^{-\beta}d/\beta}{1-cT^{-\alpha}-dT^{-\beta}} \sum_{i=1}^n (1-\delta_i) = 0. \end{aligned} \tag{24}$$

By $(8) * c + (10) * d$ we have

$$cT^{-\alpha} + dT^{-\beta} = \frac{\sum_{i=1}^{n} \delta_i}{n}. \tag{25}$$

Inserting (25) into (24), we obtain

$$\frac{T^{-\alpha}c}{\alpha} + \frac{T^{-\beta}d}{\beta} = \frac{\sum_{i=1}^{n} \delta_i(\log X_i - \log T)}{n}. \tag{26}$$

Hence, (12) and (13) follow from (25) and (26).

B. Proof of Theorem 1. Before we prove Theorem 1, we need two lemmas.

Lemma 1. *Suppose (20) holds. Then*

$$\lim_{t\to\infty} A(t)/B(t) = l_0 \in [-\infty, \infty]. \tag{27}$$

Furthermore, (i) if $l_0 = \pm\infty$, then

$$\lim_{t\to\infty} \frac{\frac{\frac{(tx)^{-1/\alpha}U(tx)}{t^{-1/\alpha}U(t)}-1}{A(t)} - \frac{x^{1-\beta/\alpha}-1}{1-\beta/\alpha}}{A(t)} = \frac{1}{2}(\frac{x^{1-\beta/\alpha}-1}{1-\beta/\alpha})^2; \tag{28}$$

(ii) if $l_0 \in (-\infty, \infty)$, then

$$\lim_{t\to\infty} \frac{\frac{\frac{(tx)^{-1/\alpha}U(tx)}{t^{-1/\alpha}U(t)}-1}{A(t)} - \frac{x^{1-\beta/\alpha}-1}{1-\beta/\alpha}}{B(t)} = h(x) + \frac{l_0}{2}(\frac{x^{1-\beta/\alpha}-1}{1-\beta/\alpha})^2. \tag{29}$$

In addition, for any $\epsilon > 0$ there exists $t_0 > 0$ such that for all $t \geq t_0, x \geq 1$,

$$|\frac{\frac{\log U(tx) - \log U(t) - \alpha^{-1}\log x}{A(t)} - \frac{x^{1-\beta/\alpha}-1}{1-\beta/\alpha}}{B(t)} - h(x)| \leq \epsilon(1 + x^{1-\beta/\alpha} + 2x^{1-\beta/\alpha+\rho+\epsilon}), \tag{30}$$

$$|\frac{\frac{\frac{(tx)^{-1/\alpha}U(tx)}{t^{-1/\alpha}U(t)}-1}{A(t)} - \frac{x^{1-\beta/\alpha}-1}{1-\beta/\alpha}}{A(t)} - \frac{1}{2}(\frac{x^{1-\beta/\alpha}-1}{1-\beta/\alpha})^2| \leq \epsilon(1 + x^{1-\beta/\alpha} + 2x^{2-2\beta/\alpha+\epsilon}) \tag{31}$$

in case $l_0 = \pm\infty$, and

$$|\frac{\frac{\frac{(tx)^{-1/\alpha}U(tx)}{t^{-1/\alpha}U(t)}-1}{A(t)} - \frac{x^{1-\beta/\alpha}-1}{1-\beta/\alpha}}{B(t)} - h(x) - \frac{l_0}{2}(\frac{x^{1-\beta/\alpha}-1}{1-\beta/\alpha})^2| \leq \epsilon(1 + x^{1-\beta/\alpha} + 2x^{1-\beta/\alpha+\rho+\epsilon}) \tag{32}$$

in case $l_0 \in (-\infty, \infty)$.

Proof. The relations (27) and (30) follow from Theorem A and Lemma 4.2 of Draisma et al. (1999), respectively. We can show (28) and (29) by expanding $\log\Big((tx)^{-1/\alpha}U(tx)/\big(t^{-1/\alpha}U(t)\big)\Big)$. Using an argument similar to the proof of Lemma 4.2 of Draisma et al. (1999) we can show (31) and (32).

Lemma 2. *Let $Y_{n,1}\le\cdots\le Y_{n,n}$ be the order statistics of a random sample of size n from the distribution function $1-x^{-1}(x>1)$. Assume $k=k(n)\to\infty$ and $k/n\to 0$ as $n\to\infty$. For any fixed $\gamma<0$ we define*

$$\begin{cases} P_1=\frac{1}{k}\sum_{i=1}^k\log\frac{Y_{n,n-i+1}}{Y_{n,n-k}}\\ P_2(\gamma)=\frac{1}{k}\sum_{i=1}^k(\frac{Y_{n,n-i+1}}{Y_{n,n-k}})^\gamma\\ P_3(\gamma)=\frac{1}{k}\sum_{i=1}^k(\frac{Y_{n,n-i+1}}{Y_{n,n-k}})^\gamma\log\frac{Y_{n,n-i+1}}{Y_{n,n-k}}.\end{cases}$$

Then

$$\sqrt{k}\{P_1-1,P_2(\gamma)-(1-\gamma)^{-1},P_3(\gamma)-(1-\gamma)^{-2}\}\xrightarrow{d}(N_3,N_4,N_5), \tag{33}$$

where (N_3,N_4,N_5) is a trivariate normal distribution with $E(N_3)=E(N_4)=E(N_5)=0$ and covariance matrix

$$V=V_\gamma=\begin{pmatrix} 1 & \frac{1}{(1-\gamma)^2}-\frac{1}{1-\gamma} & \frac{2}{(1-\gamma)^3}-\frac{1}{(1-\gamma)^2}\\ \frac{1}{(1-\gamma)^2}-\frac{1}{1-\gamma} & \frac{1}{1-2\gamma}-\frac{1}{(1-\gamma)^2} & \frac{1}{(1-2\gamma)^2}-\frac{1}{(1-\gamma)^3}\\ \frac{2}{(1-\gamma)^3}-\frac{1}{(1-\gamma)^2} & \frac{1}{(1-2\gamma)^2}-\frac{1}{(1-\gamma)^3} & \frac{1}{(1-2\gamma)^3}-\frac{1}{(1-\gamma)^4}\end{pmatrix}.$$

Proof. Similar to the proof of Lemma 3.4 of Dekkers et al. (1989).

Proof of Theorem 1. Put

$$\Delta_1(\gamma)=\frac{1}{k}\sum_{i=1}^k(\frac{X_{n,n-i+1}}{X_{n,n-k}})^\gamma$$

and

$$\Delta_2(\gamma)=\frac{1}{k}\sum_{i=1}^k(\frac{X_{n,n-i+1}}{X_{n,n-k}})^\gamma\log\frac{X_{n,n-i+1}}{X_{n,n-k}}.$$

Note that

$$\begin{aligned} Q_i^{-1}(\alpha,\beta) &= \tfrac{\beta}{\alpha}(\tfrac{X_{n,n-i+1}}{X_{n,n-k}})^{\alpha-\beta}-\tfrac{\beta^2}{\alpha-\beta}H(\alpha)(\tfrac{X_{n,n-i+1}}{X_{n,n-k}})^{\alpha-\beta}\\ &\quad+\tfrac{\beta^3}{\alpha(\alpha-\beta)}H(\alpha)(\tfrac{X_{n,n-i+1}}{X_{n,n-k}})^{2\alpha-2\beta}+O_p\big(H^2(\alpha)\big).\end{aligned}$$

Then (16) and (17) become

$$\begin{aligned} &\tfrac{\beta}{\alpha}\Delta_1(\alpha-\beta)-\tfrac{\beta^2}{\alpha-\beta}H(\alpha)\Delta_1(\alpha-\beta)+\tfrac{\beta^3}{\alpha(\alpha-\beta)}H(\alpha)\Delta_1(2\alpha-2\beta)\\ =\;&1+O_p\big(H^2(\alpha)\big)\end{aligned} \tag{34}$$

and

$$\begin{aligned} & \tfrac{\beta}{\alpha}\Delta_2(\alpha-\beta) - \tfrac{\beta^2}{\alpha-\beta}H(\alpha)\Delta_2(\alpha-\beta) + \tfrac{\beta^3}{\alpha(\alpha-\beta)}H(\alpha)\Delta_2(2\alpha-2\beta) \\ = & \ \beta^{-1} + O_p\big(H^2(\alpha)\big). \end{aligned} \tag{35}$$

Hence by Lemma 1,

$$H(\alpha) = \frac{\alpha_0-\alpha}{\alpha_0\alpha} - \frac{1}{\alpha_0}(P_1-1) - \frac{\alpha_0}{\beta_0}A(\frac{n}{k}) + O_p\big(|A(\frac{n}{k})|/\sqrt{k} + |A(\frac{n}{k})B(\frac{n}{k})|\big),$$

$$\begin{aligned} \Delta_1(\gamma) & = P_2(\gamma/\alpha_0) + \tfrac{\gamma\alpha_0}{\alpha_0-\beta_0}\big(P_2(1-\tfrac{\beta_0-\gamma}{\alpha_0}) - P_2(\tfrac{\gamma}{\alpha_0})\big)A(\tfrac{n}{k}) \\ & +O_p\big(|A(\tfrac{n}{k})|/\sqrt{k} + A^2(\tfrac{n}{k}) + |A(\tfrac{n}{k})B(\tfrac{n}{k})|\big), \\ \Delta_2(\gamma) & = \tfrac{1}{\alpha_0}P_3(\tfrac{\gamma}{\alpha_0}) + \tfrac{\gamma}{\alpha_0-\beta_0}\big(P_3(1-\tfrac{\beta_0-\gamma}{\alpha_0}) - P_3(\tfrac{\gamma}{\alpha_0})\big)A(\tfrac{n}{k}) \\ & +\tfrac{\alpha_0}{\alpha_0-\beta_0}\big(P_2(1-\tfrac{\beta_0-\gamma}{\alpha_0}) - P_2(\tfrac{\gamma}{\alpha_0})\big)A(\tfrac{n}{k}) \\ & +O_p\big(|A(\tfrac{n}{k})|/\sqrt{k} + A^2(\tfrac{n}{k}) + |A(\tfrac{n}{k})B(\tfrac{n}{k})|\big), \\ H(\alpha)\Delta_1(\gamma) & = \tfrac{\alpha_0-\alpha}{\alpha_0\alpha}\Delta_1(\gamma) - \tfrac{1}{\alpha_0}(P_1-1)P_2(\tfrac{\gamma}{\alpha_0}) - \tfrac{\alpha_0}{\beta_0}P_2(\tfrac{\gamma}{\alpha_0})A(\tfrac{n}{k}) \\ & +O_p\big(|A(\tfrac{n}{k})|/\sqrt{k} + A^2(\tfrac{n}{k}) + |A(\tfrac{n}{k})B(\tfrac{n}{k})|\big), \end{aligned}$$

and

$$\begin{aligned} H(\alpha)\Delta_2(\gamma) & = \tfrac{\alpha_0-\alpha}{\alpha_0\alpha}\Delta_2(\gamma) - \tfrac{1}{\alpha_0^2}(P_1-1)P_3(\tfrac{\gamma}{\alpha_0}) - \tfrac{1}{\beta_0}P_3(\tfrac{\gamma}{\alpha_0})A(\tfrac{n}{k}) \\ & +O_p\big(|A(\tfrac{n}{k})|/\sqrt{k} + A^2(\tfrac{n}{k}) + |A(\tfrac{n}{k})B(\tfrac{n}{k})|\big). \end{aligned}$$

So (34) and (35) become

$$\begin{aligned} & \alpha^{-1}\big(\Delta_1(\alpha-\beta) - \tfrac{\alpha_0}{\alpha_0-\alpha+\beta}\big) \\ & +\big(\tfrac{1}{\alpha_0} - \tfrac{1}{\alpha} + H(\alpha)\big)\big(-\tfrac{\beta\alpha_0}{(\alpha-\beta)(\alpha_0-\alpha+\beta)} + \tfrac{\beta^2\alpha_0}{\alpha(\alpha-\beta)(\alpha_0-2\alpha+2\beta)}\big) \\ & +(\alpha_0-\alpha)\big(-\tfrac{\alpha-\beta}{\alpha\beta(\alpha_0-\alpha+\beta)} - \tfrac{\beta}{\alpha(\alpha-\beta)(\alpha_0-\alpha+\beta)} + \tfrac{\beta^2}{\alpha^2(\alpha-\beta)(\alpha_0-2\alpha+2\beta)}\big) \\ & = O_p\big(H^2(\alpha) + \tfrac{|\alpha_0-\alpha|}{\sqrt{k}} + |\alpha_0-\alpha||A(\tfrac{n}{k})|\big) \end{aligned} \tag{36}$$

and

$$\begin{aligned} & \alpha^{-1}\big(\Delta_2(\alpha-\beta) - \tfrac{\alpha_0}{(\alpha_0-\alpha+\beta)^2}\big) \\ & +\big(\tfrac{1}{\alpha_0} - \tfrac{1}{\alpha} + H(\alpha)\big)\big(-\tfrac{\alpha_0\beta}{(\alpha-\beta)(\alpha_0-\alpha+\beta)^2} + \tfrac{\alpha_0\beta^2}{\alpha(\alpha-\beta)(\alpha_0-2\alpha+2\beta)^2}\big) \\ & +(\alpha_0-\alpha)\big(\tfrac{1}{\alpha(\alpha_0-\alpha+\beta)^2} - \tfrac{2}{\beta(\alpha_0-\alpha+\beta)^2} - \tfrac{\beta}{\alpha(\alpha-\beta)(\alpha_0-\alpha+\beta)^2} + \tfrac{\beta^2}{\alpha^2(\alpha-\beta)(\alpha_0-2\alpha+2\beta)^2}\big) \\ & = O_p\big(H^2(\alpha) + \tfrac{|\alpha_0-\alpha|}{\sqrt{k}} + |\alpha_0-\alpha||A(\tfrac{n}{k})| + (\alpha_0-\alpha)^2\big). \end{aligned} \tag{37}$$

Using those expansions, we can further reduce (36) and (37) to

$$\begin{aligned} & \alpha_0^{-1}\big(P_2(1-\beta/\alpha_0) - \alpha_0/\beta\big) + \tfrac{\beta-\alpha_0}{\alpha_0(2\beta-\alpha_0)}(P_1-1) \\ & +A(\tfrac{n}{k})\tfrac{\alpha_0(\beta-\alpha_0)^2(\beta-\beta_0)}{\beta_0\beta(2\beta-\alpha_0)(\beta_0-\alpha_0+\beta)} - \tfrac{(\alpha_0-\alpha)(\beta-\alpha_0)^3}{\beta^2\alpha_0^2(2\beta-\alpha_0)} \\ & = O_p\big(H^2(\alpha) + \tfrac{|\alpha_0-\alpha|}{\sqrt{k}} + |\alpha_0-\alpha||A(\tfrac{n}{k})| + |A(\tfrac{n}{k})B(\tfrac{n}{k})|\big) \end{aligned} \tag{38}$$

and

$$\begin{aligned} & \alpha_0^{-2}\big(P_3(1-\beta/\alpha_0) - \tfrac{\alpha_0^2}{\beta^2}\big) - \tfrac{-\beta^2+3\alpha_0\beta-\alpha_0^2}{\alpha_0\beta(2\beta-\alpha_0)^2}(P_1-1) \\ & +A(\tfrac{n}{k})\tfrac{\alpha_0(\alpha_0-\beta)(\beta-\beta_0)(2\alpha_0\beta_0\beta-\alpha_0^2\beta_0-4\alpha_0^2\beta+4\alpha_0\beta^2+\alpha_0^3-\beta^3+\beta_0\beta^2)}{\beta_0\beta^2(\beta_0-\alpha_0+\beta)^2(2\beta-\alpha_0)^2} \\ & +(\alpha_0-\alpha)\tfrac{(\beta-\alpha_0)^2(5\alpha_0\beta-2\alpha_0^2-\beta^2)}{\alpha_0^2\beta^3(2\beta-\alpha_0)^2} \\ & = O_p\big(H^2(\alpha) + \tfrac{|\alpha_0-\alpha|}{\sqrt{k}} + |\alpha_0-\alpha||A(\tfrac{n}{k})| + |A(\tfrac{n}{k})B(\tfrac{n}{k})| + (\alpha_0-\alpha)^2\big). \end{aligned} \tag{39}$$

To solve these two equations, we need to estimate the order of magnitude for both $\alpha - \alpha_0$ and $\beta - \beta_0$. First we can cancel the $A(n/k)$ term by subtracting (39) from (38), after multipling by an appropriate constant. Then we get that $\alpha - \alpha_0 = O_P(1/\sqrt{k})$ under (21), and thus, the right-hand sides of both (38) and (39) are of order $o_P(1/\sqrt{k})$. Since both $\big(P_2(1-\beta/\alpha_0)-\alpha_0/\beta\big)$ and $P_1 - 1$ are of order $O_P(1/\sqrt{k})$, we have from (37) that $A(n/k)(\beta-\beta_0) = O_P(1/\sqrt{k})$. In view of (21), we get $\beta - \beta_0 = O_P\Big(1/\big(\sqrt{k}A(n/k)\big)\Big) = o_P(1)$. Notice that $A(n/k)(\beta-\beta_0)^2 = O_P\Big(1/\big(kA(n/k)\big)\Big) = o_P(1/\sqrt{k})$. We can expand $P_2(1-\beta/\alpha_0) - \alpha_0/\beta$ at $\beta = \beta_0$ and obtain

$$\begin{aligned} & \big(P_2(1-\beta/\alpha_0) - \alpha_0/\beta\big) - \big(P_2(1-\beta_0/\alpha_0) - \alpha_0/\beta_0\big) \\ = & \big(P_3(1-\beta_0/\alpha_0) - \alpha_0^2/\beta_0^2\big)(\beta-\beta_0)\big(1+o_P(1)\big) \\ = & o_P(\tfrac{1}{\sqrt{k}}). \end{aligned} \tag{40}$$

A similar expansion is applied to $P_3(1-\beta/\alpha_0) - \alpha_0^2/\beta^2$. Now we expand the left-hand sides of both (38) and (39) at $\beta = \beta_0$ and get

$$\begin{aligned} & \alpha_0^{-1}\big(P_2(1-\beta_0/\alpha_0) - \alpha_0/\beta_0\big) + \tfrac{\beta_0-\alpha_0}{\alpha_0(2\beta_0-\alpha_0)}(P_1 - 1) \\ & + A(\tfrac{n}{k})\tfrac{\alpha_0(\beta_0-\alpha_0)^2(\beta-\beta_0)}{\beta_0^2(2\beta_0-\alpha_0)^2} - \tfrac{(\alpha_0-\alpha)(\beta_0-\alpha_0)^3}{\beta_0^2\alpha_0^2(2\beta_0-\alpha_0)} = o_P(\tfrac{1}{\sqrt{k}}) \end{aligned} \tag{41}$$

and

$$\begin{aligned} & \alpha_0^{-2}\big(P_3(1-\beta_0/\alpha_0) - \tfrac{\alpha_0^2}{\beta_0^2}\big) - \tfrac{-\beta_0^2+3\alpha_0\beta_0-\alpha_0^2}{\alpha_0\beta_0(2\beta_0-\alpha_0)^2}(P_1 - 1) \\ & + A(\tfrac{n}{k})\tfrac{\alpha_0(\alpha_0-\beta_0)(6\alpha_0\beta_0^2-5\alpha_0^2\beta_0-\alpha_0^3)(\beta-\beta_0)}{\beta_0^3(2\beta_0-\alpha_0)^4} \\ & + (\alpha_0-\alpha)\tfrac{(\beta_0-\alpha_0)^2(5\alpha_0\beta_0-2\alpha_0^2-\beta_0^2)}{\alpha_0^2\beta_0^3(2\beta_0-\alpha_0)^2} = o_P(\tfrac{1}{\sqrt{k}}). \end{aligned} \tag{42}$$

Set $\lambda = \beta_0/\alpha_0$ and solve the above two equations. Then $\hat{\alpha} - \alpha_0$ and $\hat{\beta} - \beta_0$ can be expressed as

$$\begin{aligned} & \hat{\alpha}(k) - \alpha_0 \\ = & \alpha_0\Big(\tfrac{\lambda^4}{(\lambda-1)^4}(P_1-1) + \tfrac{\lambda^2(2\lambda-1)(3\lambda-1)}{(\lambda-1)^5}\big(P_2(1-\lambda)-\lambda^{-1}\big) + \tfrac{\lambda^3(2\lambda-1)^2}{(\lambda-1)^4}\big(P_3(1-\lambda)-\lambda^{-2}\big)\Big) \\ & + o_P(\tfrac{1}{\sqrt{k}}) \end{aligned} \tag{43}$$

and

$$\begin{aligned} & A(\tfrac{n}{k})(\hat{\beta}(k) - \beta_0) \\ = & \tfrac{1}{\lambda-1}(P_1-1) + \tfrac{-\lambda^2+5\lambda-2}{(\lambda-1)^2}\big(P_2(1-\lambda)-\lambda^{-1}\big) + \tfrac{\lambda(2\lambda-1)}{\lambda-1}\big(P_3(1-\lambda)-\lambda^{-2}\big) + o_P(\tfrac{1}{\sqrt{k}}). \end{aligned} \tag{44}$$

Define

$$\Sigma = \begin{pmatrix} \frac{\alpha_0\lambda^4}{(\lambda-1)^4} & \frac{1}{\lambda-1} \\ \frac{\alpha_0\lambda^2(2\lambda-1)(3\lambda-1)}{(\lambda-1)^5} & \frac{-\lambda^2+5\lambda-2}{(\lambda-1)^2} \\ \frac{\alpha_0\lambda^3(2\lambda-1)^2}{(\lambda-1)^4} & \frac{\lambda(2\lambda-1)}{\lambda-1} \end{pmatrix}.$$

Then from Lemma 2 we have that $\Big(\sqrt{k}\big(\hat{\alpha}(k)-\alpha_0\big), \sqrt{k}A(n/k)\big(\hat{\beta}(k)-\beta_0\big)\Big)$ converges in distribution to a bivariate normal distribution with mean zero and covariance matrix $\Sigma' V_{1-\lambda}\Sigma$.

Tedious calculation shows that the variance matrix has a very simple form:

$$\Sigma' V_{1-\lambda}\Sigma = \begin{pmatrix} \frac{\alpha_0^2\lambda^4}{(\lambda-1)^4} & \frac{\alpha_0}{\lambda-1} \\ \frac{\alpha_0}{\lambda-1} & \frac{(\lambda-1)^2}{\lambda^2(2\lambda-1)} \end{pmatrix}.$$

That finishes the proof of the theorem.

Acknowledgement. The authors would like to thank the referees for their constructive suggestions which have led to improvement of the layout and readability of the paper. The authors are also grateful to Professor Ted Hill for his comments.

References

Beirlant, J., Dierckx, G., Goegebeur, Y. & Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes* **2(2)**, 177 - 200.

Beirlant, J., Teugels, J.L. & Vynckier, P. (1996). Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.* **91**, 1695 - 1667.

Csörgő, S., Deheuvels, P. & Mason, D. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.* **13**, 1050 - 1077.

Csörgő, S. & Viharos, L. (1997). Asymptotic normality of least-squares estimators of tail indices. *Bernoulli* **3(3)**, 351 - 370.

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

Danielsson, J., Hartmann, P. & de Vries, C.G. (1998). The cost of conservatism. *RISK* **11(10)**, 101 - 103.

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

Dekkers, A.L.M., Einmahl, J.H.J. & de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17(4)**, 1833 - 1855.

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

Draisma, G., de Haan, L., Peng, L. & Pereira, T.T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes* **2(4)**, 367 -404.

Drees, H. & Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes Appl.* **75(2)**, 149 - 172.

Embrechts, P., Resnick, S. & Samorodnitsky, G. (1998). Living on the Edge. *Risk* **11(1)**, 96 - 100.

Feuerverger, A. & Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.* **27**, 760 - 781.

Guillou, A. & Hall, P. (2001). A diagnostic for selecting the threshold in extreme-value analysis. *J. Roy. Statist. Soc. Ser. B.* **63(2)**, 293 - 305.

De Haan, L. & Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica* **52(1)**, 60 - 70.

De Haan, L. & Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Australian Math. Soc. (Series A)* **61**, 381 - 395.

Hall, P. (1982a). On estimating the endpoint of a distribution. *Ann. Statist.* **10(2)**, 556 - 568.

Hall, P. (1982b). On some simple estimators of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44**, 37 - 42.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and selecting parameter in nonparametric problems. *Journal of Multivariate Analysis* **32**, 177 - 203.

Hall, P. & Weissman, I. (1997). On the estimation of extreme tail probabilities. *Ann. Statist.* **25(3)**, 1311 - 1326.

Hall, P. & Welsh, A.H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13**, 331 - 341.

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163 - 1174.

Mason, D. (1982). Laws of large numbers for sums of extreme values. *Ann. Probab.* **10**, 754 - 764.

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

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73**, 812 - 815.

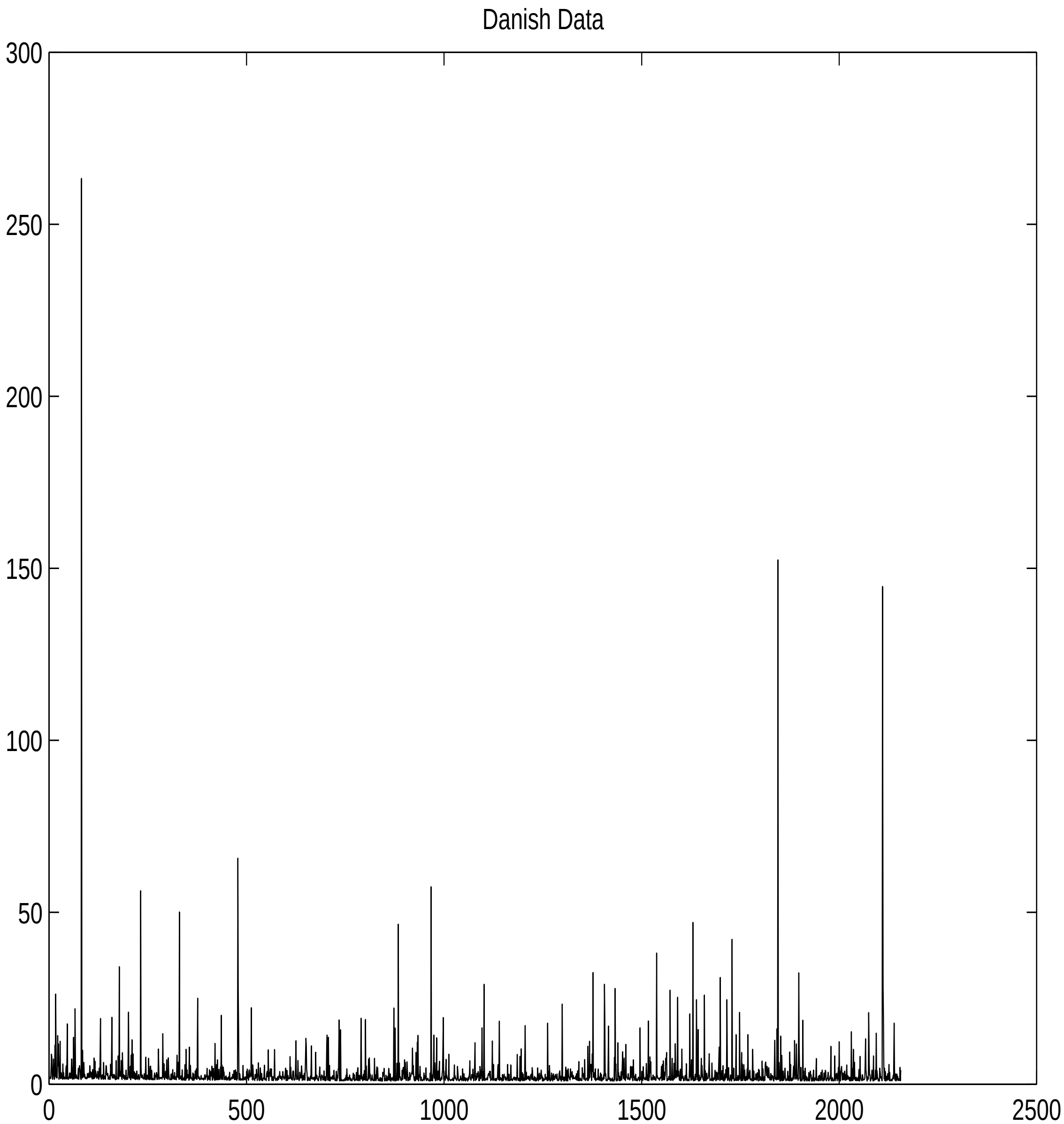

Figure 1: *Danish fire loss data.* This consists of 2156 losses over one million Danish Krone (DKK) from the years 1980 to 1990, inclusive.

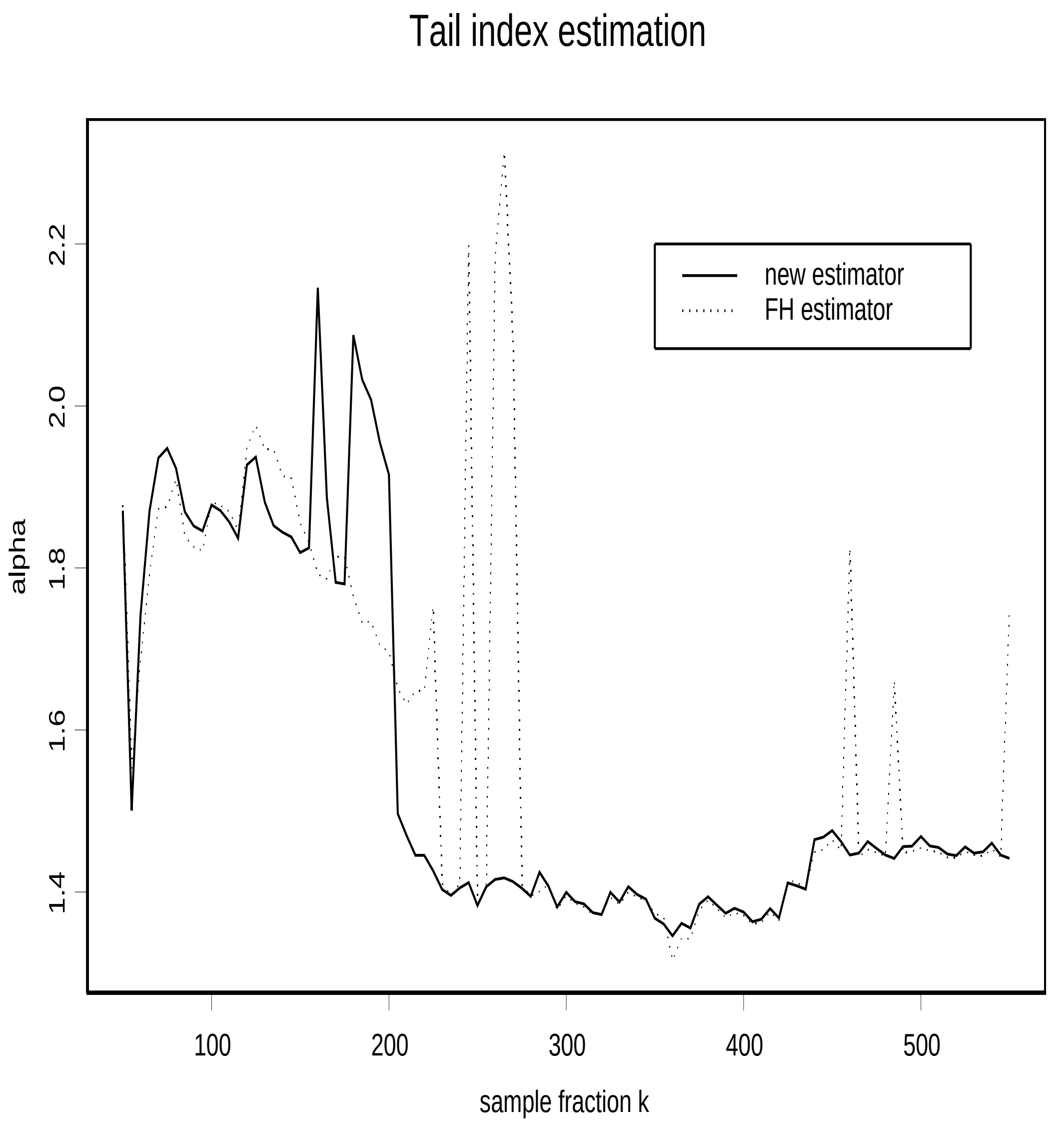

Figure 2: *Tail index estimation of Danish fire loss data.* Here "new estimator" and "FH estimator" denote $\hat{\alpha}(k)$ and $\hat{\alpha}_{FH}(k)$, respectively.