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

This paper suggests 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 maximum likelihood method proposed by Feuerverger and Hall in 1999. The paper derives the joint asymptotic limit for the first- and second-order parameters under a weaker assumption. The paper also demonstrates through a simulation study that the suggested estimator for the first-order parameter is better than the estimator proposed by Feuerverger and Hall 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

To estimate large quantiles or extreme tail probabilities of an unknown distribution function, we have to estimate beyond the observations, so we need extra assumptions on the underlying distribution function. One approach is to assume that the underlying distribution has a heavy tail; see Danielsson & de Vries (1997), Hall & Weissman (1997), Danielsson, Hartman & de Vries (1998) and Embrechts, Resnick & Samorodnitsky (1998). Thus, estimating the 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 (1982a), Csörgő, Deheuvels & Mason (1985), Csörgő & Viharos (1997) and de Haan & Peng (1998). Since we make inference about the tail quantity, we can use only 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 introduces a big bias in the estimation, so the choice of k plays an important role. Recently, several procedures have been proposed for choosing the optimal k in the sense of asymptotic minimal mean squared error; see Hall (1990), Dekkers & de Haan (1993), Beirlant, Teugels & Vynckier (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 independent and identically distributed random variables with common cumulative distribution function F which satisfies

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

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

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

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

$$(Y_1, \ldots, Y_k) \xrightarrow{d} (W_1, \ldots, W_k) \quad \text{as } n \to \infty,$$

where the W_i are independent $\text{Exp}(\alpha)$ random variables, i.e. exponentially distributed with mean $1/\alpha$ (see e.g. Weissman, 1978). Therefore, asymptotically the Hill estimator can be viewed as the sample mean of $W_1, \ldots, W_k$. For the consistency of $\hat{\alpha}_H(k)$ we refer to Mason (1982). 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 $c > 0$, $d \neq 0$ 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). For simplicity define $\theta = (\beta/\alpha) - 1$. Under condition (2) it can be shown that, if $\sqrt{k}(k/n)^\theta \to \lambda \in [0, \infty)$,

$$\frac{\hat{\alpha}_H(k) - \alpha}{\alpha/\sqrt{k}} \xrightarrow{d} \text{N}\left(\frac{\lambda \theta d}{(1 + \theta)c^{1+\theta}}, 1 \right)$$

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

$$k^* = \left(\frac{(1 + \theta)^2 c^{2+2\theta}}{2\theta^3 d^2} \right)^\xi n^{1-\xi}, \quad \text{where } \xi = \frac{1}{1 + 2\theta} \tag{3}$$

in the sense of minimal asymptotic mean squared error of the Hill estimator. Using the result

$$Y_i \overset{d}{\sim} \text{Exp}(g_i \alpha) \quad \text{where } g_i = g_i(\theta, \omega) = \exp\left(-\omega \left(\frac{i}{n} \right)^\theta \right), \quad \text{with } \omega = \frac{\theta d}{c^{1+\theta}},$$

Feuerverger & Hall (1999) estimated α, ω and θ by the maximum likelihood method. This results in the estimator

$$\hat{\alpha}_{\text{FH}}(k) = \left(\frac{1}{k} \sum_{i=1}^{k} \frac{Y_i}{g_i(\hat{\theta}, \hat{\omega})} \right)^{-1}, \tag{4}$$

where $(\hat{\theta}, \hat{\omega})$ are chosen to minimize

$$L_1(\theta, \omega) = \frac{1}{k} \sum_{i=1}^{k} \log g_i + \log \left(\frac{1}{k} \sum_{i=1}^{k} \frac{Y_i}{g_i} \right).$$

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

The determination of the optimal sample fraction k^* in (3) 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 (MLE) for left-censored data, rather than basing the estimator on an asymptotic exponential distribution. 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 simultaneous estimation of both α and β, and permits a larger range of sample fractions for the new estimator of the first-order parameter α without introducing any bias. Section 2 gives the detailed method and main results, showing that our new estimator for α has the same asymptotic variance as $\hat{\alpha}_{FH}(k)$ defined in (4). 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. Section 3 presents a simulation study and a real application, 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 omitted in the paper. Interested readers can find them in the technical report by Peng & Qi (2003).

2. Method

Let $T = T(n) \to \infty$ as $n \to \infty$, and define $\delta_i = \mathrm{I}\,(X_i > T)$ for $i = 1, \ldots, n$. Since we can only use a part of upper order statistics to make inference, we view our observations as $((X_i \vee T, \delta_i),\ 1 \le i \le n)$ instead of $(X_i, 1 \le i \le n)$. If we suppose that $1 - F(x) = cx^{-\alpha}$ for $x > T$, then the likelihood for $((X_i \vee T, \delta_i), 1 \le i \le n)$ is

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

Hence we have

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

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

If $1 - F(x) = cx^{-\alpha} + dx^{-\beta}$ for $x > T$, then the likelihood for $((X_i \vee T, \delta_i), 1 \le i \le n)$ is

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

Therefore our new estimators can be obtained as

$$(\bar{\alpha}, \bar{c}, \bar{\beta}, \bar{d}) = \operatorname*{argmax}_{\alpha>0,c>0,\beta>\alpha,d\neq0} L(\alpha, c, \beta, d)\,,$$

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

$$\frac{\partial \log L}{\partial \alpha} = \sum_{i=1}^{n}\frac{\delta_i(cX_i^{-\alpha-1} - c\alpha X_i^{-\alpha-1}\log X_i)}{c\alpha X_i^{-\alpha-1} + d\beta X_i^{-\beta-1}} + \frac{mcT^{-\alpha}\log T}{1 - cT^{-\alpha} - dT^{-\beta}} = 0\,, \qquad (5)$$

$$\frac{\partial \log L}{\partial c} = \sum_{i=1}^{n}\frac{\delta_i\alpha X_i^{-\alpha-1}}{c\alpha X_i^{-\alpha-1} + d\beta X_i^{-\beta-1}} - \frac{mT^{-\alpha}}{1 - cT^{-\alpha} - dT^{-\beta}} = 0\,, \qquad (6)$$

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

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

where $m = \sum_{i=1}^{n}(1 - \delta_i)$, under the constraints

$$\beta > \alpha > 0, \qquad c > 0, \qquad d \neq 0.$$

It follows from (5)–(8) that

$$c = \frac{\alpha \beta T^{\alpha}}{\alpha - \beta} \left(\frac{\sum_{i=1}^{n} \delta_i}{n\beta} - \frac{\sum_{i=1}^{n} \delta_i (\log X_i - \log T)}{n} \right), \qquad (9)$$

$$d = \frac{\alpha \beta T^{\beta}}{\beta - \alpha} \left(\frac{\sum_{i=1}^{n} \delta_i}{n\alpha} - \frac{\sum_{i=1}^{n} \delta_i (\log X_i - \log T)}{n} \right); \qquad (10)$$

see Peng & Qi (2003) for detail. Write

$$\begin{cases} H^*(\alpha) = \dfrac{\sum_{i=1}^{n} \delta_i}{n\alpha} - \dfrac{\sum_{i=1}^{n} \delta_i \log(X_i/T)}{n}, \\[3mm] Q_i^*(\alpha, \beta) = \alpha \left(\dfrac{\sum_{i=1}^{n} \delta_i}{n} + \dfrac{\alpha\beta}{\alpha - \beta} H^*(\alpha) \right) \left(\dfrac{X_i}{T} \right)^{\beta-\alpha} - \dfrac{\alpha\beta^2}{\alpha - \beta} H^*(\alpha). \end{cases}$$

Substituting (9) and (10) into (8), we obtain

$$\frac{1}{n} \sum_{i=1}^{n} \delta_i \frac{1}{Q_i^*(\alpha, \beta)} = \frac{1}{\beta}. \qquad (11)$$

Substituting (9) and (10) into (7) and using (11) we have

$$\frac{1}{n} \sum_{i=1}^{n} \delta_i \frac{1}{Q_i^*(\alpha, \beta)} \log \frac{X_i}{T} = \frac{1}{\beta^2}. \qquad (12)$$

If we take $T = X_{n,n-k}$ and define

$$\begin{cases} H(\alpha) = \dfrac{1}{\alpha} - \dfrac{1}{k} \sum_{i=1}^{k} \log \dfrac{X_{n,n-i+1}}{X_{n,n-k}}, \\[3mm] Q_i(\alpha, \beta) = \dfrac{\alpha}{\beta} \left(1 + \dfrac{\alpha\beta}{\alpha - \beta} H(\alpha) \right) \left(\dfrac{X_{n,n-i+1}}{X_{n,n-k}} \right)^{\beta-\alpha} - \dfrac{\alpha\beta}{\alpha - \beta} H(\alpha), \end{cases}$$

then (11) and (12) become

$$\frac{1}{k} \sum_{i=1}^{k} \frac{1}{Q_i(\alpha, \beta)} = 1 \qquad (13)$$

and

$$\frac{1}{k} \sum_{i=1}^{k} \frac{1}{Q_i(\alpha, \beta)} \log \frac{X_{n,n-i+1}}{X_{n,n-k}} = \frac{1}{\beta}. \qquad (14)$$

Put

$$\beta > \alpha > 0 \quad \text{and} \quad \beta > \hat{\alpha}_H(k). \tag{15}$$

Our reason for imposing the constraint $\beta > \hat{\alpha}_H(k)$ is that $\beta = \hat{\alpha}_H(k)$ is a solution to (13) for any fixed $\alpha < \hat{\alpha}_H(k)$. For any fixed $\beta > \alpha_0$, where α_0 denotes the true parameter value,

$$
\begin{cases}
\dfrac{1}{k} \displaystyle\sum_{i=1}^{k} \dfrac{1}{Q_i(\hat{\alpha}_H(k), \beta)} \xrightarrow{\text{p}} 1, \\[3ex]
\dfrac{1}{k} \displaystyle\sum_{i=1}^{k} \dfrac{1}{Q_i(\hat{\alpha}_H(k), \beta)} \log \dfrac{X_{n,n-i+1}}{X_{n,n-k}} \xrightarrow{\text{p}} \dfrac{1}{\beta},
\end{cases}
$$

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

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

$$\lim_{t \to \infty} \frac{U(tx)/U(t) - x^{1/\alpha}}{A(t)} = x^{1/\alpha} \frac{1 - x^{-\theta}}{\theta}, \quad \text{where } \theta = \frac{\beta}{\alpha} - 1 > 0,$$

i.e. $\displaystyle\lim_{t \to \infty} \frac{1}{A(t)} \left(\log U(tx) - \log U(t) - \frac{1}{\alpha} \log x \right) = \frac{1 - x^{-\theta}}{\theta}$, where $A(t) = -\dfrac{\theta d}{\alpha c^{1+\theta}} \dfrac{1}{t^\theta}$.

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

$$\lim_{t \to \infty} \frac{1}{B(t)} \left(\frac{\log U(tx) - \log U(t) - \alpha^{-1} \log x}{A(t)} - \frac{1 - x^{-\theta}}{\theta} \right)$$
$$= -\frac{1}{\rho} \left(\frac{1 - x^{-\theta-\rho}}{\theta + \rho} - \frac{1 - x^{-\theta}}{\theta} \right) = h(x), \tag{16}$$

where $\rho \geq 0$ can be called the third-order regular variation parameter.

Our main result is as follows.

Theorem 1. *Suppose (16) holds with true parameters $\alpha_0 > 0$ and $\beta_0 > \alpha_0$, and suppose $k = k(n)$ satisfies*

$$k \to \infty, \quad \sqrt{k} \left| A\left(\frac{n}{k}\right) \right| \to \infty, \quad \sqrt{k} A^2\left(\frac{n}{k}\right) \to 0, \quad \sqrt{k} \left| A\left(\frac{n}{k}\right) B\left(\frac{n}{k}\right) \right| \to 0, \quad \text{as } n \to \infty.$$

Assume there exists a solution to (13)–(15), say $(\hat{\alpha}(k), \hat{\beta}(k))$. Then

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

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

The condition $\sqrt{k} |A(n/k)| \to \infty$ ensures that there exists a consistent solution to (13)–(15); see Peng & Qi (2003) for the proof of Theorem 1. We suspect that the theorem in Feuerverger & Hall (1999) requires the consistency of the estimator of β_0 since the expansion $D(i/n)^\theta$ in Feuerverger & Hall (1999 p.776) requires that $(\theta - \theta_0) \log(i/n) \to 0$ uniformly for $i = 1, \ldots, k$, where $\theta = \beta/\alpha - 1$ and $\theta_0 = \beta_0/\alpha_0 - 1$.

TABLE 1

Comparison by using a practical choice of sample fraction. Estimators $\hat{\alpha}(k)$ and $\hat{\alpha}_{FH}(k)$ are computed with $k = \min([\frac{1}{2}k^* \log k^*], \frac{1}{2}n)$, where k^* is given in (3).

	k^*	k	$\hat{\alpha}(k)$	se	$\hat{\alpha}_{FH}(k)$	se
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 of the effect of using different choices of sample fraction. Estimators $\hat{\alpha}(k)$ and $\hat{\alpha}_{FH}(k)$ are computed for $k = k^* + 10i$, $i = 1, \ldots, 10$, where k^* is given in (3).

	Burr(0.5, 0.7)				Burr(2.0, 3.0)			
k	$\hat{\alpha}(k)$	se	$\hat{\alpha}_{FH}(k)$	se	$\hat{\alpha}(k)$	se	$\hat{\alpha}_{FH}(k)$	se
$k^* + 10$	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	0.0784	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	0.3596	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

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 $\hat{\alpha}_{FH}(k)$. Condition (4.1) in Feuerverger & Hall (1999) is slightly stronger than our condition (16). On the other hand, we expect that $\hat{\alpha}(k)$ would behave better than $\hat{\alpha}_{FH}(k)$ since $\hat{\alpha}(k)$ is based on a censored likelihood function rather than an approximate exponential distribution like $\hat{\alpha}_{FH}(k)$. This is confirmed in Section 3.

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. Our estimator for β is a sort of MLE, so it can be considered to be efficient.

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 are able to show, by a refinement of the proof of Theorem 1, that the limit in Theorem 1 has a bias term.

Comparing our estimator with the Hill estimator, we can draw the same conclusions as Feuerverger & Hall (1999), i.e. our new estimator $\hat{\alpha}(k)$ allows the use of a larger sample fraction k.

3. Simulation study and real application

3.1. Simulation study

We report on 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).
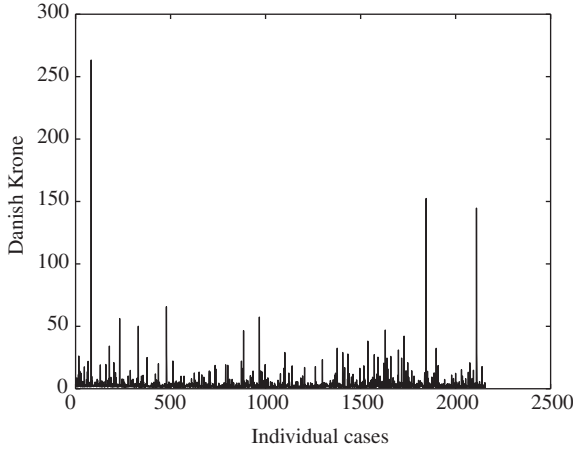
Figure 1. Danish fire loss data. This consists of 2156 losses over
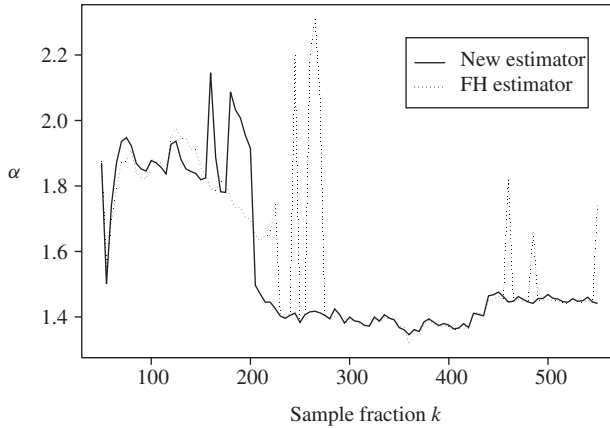one million Danish Krone from the years 1980 to 1990, inclusive.

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.

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 $k = \min([\frac{1}{2}k^* \log k^*], \frac{1}{2}n)$ with the theoretical optimal value of k^* given in (3) 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 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 using various choices of sample fraction for distributions Burr$(0.5, 0.7)$ and Burr$(2.0, 3.0)$; see Table 2. Although both estimators have the same asymptotic variance, we can conclude from Tables 1 and 2 that our new estimator $\hat{\alpha}(k)$ is better than $\hat{\alpha}_{FH}(k)$, because it is based on a censored likelihood function whereas $\hat{\alpha}_{FH}(k)$ is based on an approximate exponential distribution.

3.2. Real application

The dataset we analyse consists of 2156 Danish fire losses totalling over one million Danish Krone 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 furnishings and personal property, as well as loss of profits. This Danish fire dataset was analysed by McNeil (1997). We compute $\hat{\alpha}(k)$ and $\hat{\alpha}_{\mathrm{FH}}(k)$ for $k = 50 + 5i$, $i = 1, \ldots, 100$; see Figure 2. We observe from Figure 2 that our new estimator $\hat{\alpha}(k)$ is much more robust than $\hat{\alpha}_{\mathrm{FH}}(k)$ as the sample fraction k becomes large.

References

BEIRLANT, J., TEUGELS, J.L. & VYNCKIER, P. (1996). Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.* **91**, 1695–1667.

BEIRLANT, J., DIERCKX, G., GOEGEBEUR, Y. & MATTHYS, G. (1999). Tail index estimation and an exponential regression model. *Extremes* **2**, 177–200.

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

CSÖRGŐ, S., DEHEUVELS, P. & MASON, D. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.* **13**, 1050–1077.

DANIELSSON, J. & DE VRIES, C.G. (1997). Tail index and quantile estimation with very high frequency data. *J. Empirical Finance* **4**, 241–257.

DANIELSSON, J., HARTMANN, P. & DE VRIES, C.G. (1998). The cost of conservatism. *Risk* **11**(10), 101–103.

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. *J. Multivariate Anal.* **76**, 226–248.

DE HAAN, L. & PENG, L. (1998). Comparison of tail index estimators. *Statist. Neerlandica* **52**, 60–70.

DE HAAN, L. & STADTMÜLLER, U. (1996). Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A* **61**, 381–395.

DEKKERS, A.L.M. & DE HAAN, L. (1993). Optimal choice of sample fraction in extreme-value estimation. *J. Multivariate Anal.* **47**, 173–195.

DREES, H. & KAUFMANN, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Process. Appl.* **75**, 149–172.

EMBRECHTS, P., RESNICK, S. & SAMORODNITSKY, G. (1998). Living on the edge. *Risk* **11**, 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**, 293–305.

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

HALL, P. (1982b). On estimating the endpoint of a distribution. *Ann. Statist.* **10**, 556–568.

HALL, P. (1990). Using the bootstrap to estimate mean squared error and selecting parameter in nonparametric problems. *J. Multivariate Anal.* **32**, 177–203.

HALL, P. & WEISSMAN, I. (1997). On the estimation of extreme tail probabilities. *Ann. Statist.* **25**, 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.

PENG, L. & QI, Y. (2003). *Estimating the First and Second Order Parameters of a Heavy Tailed Distribution.* Technical Report TR 2003-1. Duluth: University of Minnesota Duluth.

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