A semi-parametric estimator of a shape second order parameter[*]

Frederico Caeiro
Universidade Nova de Lisboa, FCT and CMA

M. Ivette Gomes
Universidade de Lisboa, DEIO, CEAUL and FCUL

November 8, 2012

Abstract

In extreme value theory, any second order parameter is an important parameter related to the speed of convergence of maximum values, linearly normalized, towards its limit law. In this paper we study a new estimator of a shape second order parameter under a third order framework. Applications to a real data set in the field of insurance as well as to simulated data are also provided.

Keywords and phrases. Heavy tails, second order parameter, semi-parametric estimation.

1 Introduction

Let us assume that $X_1, X_2, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables (r.v.'s), with a Pareto-type distribution function (d.f.) F satisfying

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = \lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma} \quad \Leftrightarrow \quad \lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma}, \quad \forall\, x>0, \tag{1.1}$$

where $\gamma(>0)$ and $U(t) := \inf\{x : F(x) \geq 1-1/t\}$. Then we are in the max-domain of attraction of the Extreme Value distribution

$$EV_\gamma(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\}, \quad 1+\gamma x>0,$$

where γ is the extreme value index (EVI). This index measures the heaviness of the right tail $\overline{F}$, and the heavier the right tail, the larger the EVI is. Although we deal with the right tail $\overline{F}$, the results here presented are applicable to the left tail F, after the change of variable $Y = -X$.

[*]Research partially supported by National Funds through **FCT** — Fundação para a Ciência e a Tecnologia, projects PEst-OE/MAT/UI0006/2011 and PEst-OE/MAT/UI0297/2011, and PTDC/FEDER.

The estimation of the EVI is an important subject in Extreme Value Theory. Many classical EVI-estimators, based on the k largest order statistics have a strong asymptotic bias for moderate up to large values of k. To improve the estimation of γ through the adaptive selection of k or through the reduction of bias of the classical EVI estimators, we usually need to know the non positive second-order parameter, ρ, ruling the rate of convergence of the normalized sequence of maximum values towards the limiting law EV, in Equation (1), through the limiting relation

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}, \quad \forall\, x > 0, \tag{1.2}$$

where $|A|$ must then be of regular variation with index ρ (Geluk and de Haan, 1987).

In this paper, we are interested in the estimation of the second order parameter ρ in (1.2). For technical reasons, we shall consider $\rho < 0$. In Section 2, after a brief review of some estimators in the literature, we will introduce a new estimator for the second order parameter ρ. In Section 3, we derive the asymptotic behavior of the ρ-estimators. Finally, in Section 4 we provide some applications to real and simulated data.

2 Estimation of the second order parameter ρ

2.1 A review of some estimators in the literature

Many estimators of the second order parameter ρ, in equation (1.2), are based on the scaled log-spacings U_i or on the log-excesses V_{ik} defined by

$$U_i := i\left\{\ln \frac{X_{n-i+1:n}}{X_{n-i:n}}\right\} \quad \text{and} \quad V_{ik} := \ln \frac{X_{n-i+1:n}}{X_{n-k:n}}, \quad 1 \le i \le k < n, \tag{2.1}$$

where $X_{i:n}$ denotes the $i-th$ ascending order statistic from a sample of size n.

The first estimator of ρ appears in Hall and Welsh (1985). Under the second order condition in (1.2), with $\rho < 0$ and $A(t) = \gamma\,\beta\, t^\rho$, the log-spacings U_i, $1 \le i \le k$, in Equation (2.1), are approximately exponential with mean value $\gamma \exp(\beta(i/n)^{-\rho}$, $1 \le i \le k$. Feuerverger and Hall (1999) considered the joint maximization, in order to γ, β and ρ, of the approximate log-likelihood of the scaled log-spacings. Such a maximization led Feuerverger and Hall to an explicit expression for $\hat\gamma$, as a function of $\hat\beta$ and $\hat\rho$, and to implicit estimators of $\hat\beta = \hat\beta_n^{FH}(k)$ and $\hat\rho = \hat\rho_n^{FH}(k)$. More precisely,

$$(\hat\beta, \hat\rho) := \underset{(\beta,\rho)}{\arg\min} \Big\{ \log\Big(\frac{1}{k}\sum_{i=1}^{k} e^{-\beta(i/n)^{-\rho}}\, U_i\Big) + \beta\Big(\frac{1}{k}\sum_{i=1}^{k}(i/n)^{-\rho}\Big)\Big\}. \tag{2.2}$$

Gomes *et al.* (2002) and Fraga Alves *et al.* (2003) worked with the log-excesses V_{ik}, in (2.1), to obtain new estimators of the second order parameter ρ. As mentioned by Goegebeur *et al.* (2010), the estimator generally considered to be the best working one in practice is a particular member of the class of estimators proposed by Fraga Alves *et al.* (2003). Such a class of estimators has been first parameterised in a tuning parameter $\tau \geq 0$, but more generally, τ can be considered as a real number (Caeiro and Gomes, 2006). It is defined as

$$\hat{\rho}_n^{FAGH(\tau)}(k) = \frac{3(T_{n,k}^{(\tau)} - 1)}{T_{n,k}^{(\tau)} - 3}, \qquad T_{n,k}^{(\tau)} := \frac{\left(M_{n,k}^{(1)}\right)^{\tau} - \left(M_{n,k}^{(2)}/2\right)^{\tau/2}}{\left(M_{n,k}^{(2)}/2\right)^{\tau/2} - \left(M_{n,k}^{(3)}/6\right)^{\tau/3}}, \quad \tau \in \mathcal{R}, \tag{2.3}$$

with $M_{n,k}^{(\alpha)} := \frac{1}{k}\sum_{i=1}^{k}(V_{ik})^{\alpha}$, $\alpha > 0$, and the notation $a^{b\tau} = b\ln a$ whenever $\tau = 0$.

Remark 2.1 (A few comments on the choice of τ). *The use of the estimator $\hat{\rho}_n^{FAGH(\tau)}(k)$ in several articles on reduced-bias tail index estimation, has led several authors to choose $\tau = 0$, if $\rho \geq -1$ and $\tau = 1$ if $\rho < -1$. However, practitioners should not choose blindly the value of τ. It is sensible to draw a few sample paths of k vs. $\hat{\rho}_n^{FAGH(\tau)}(k)$, for several values of τ, electing the one which provides the highest stability for large k.*

More recently, Ciuperca and Mercadier (2010) extended the estimators in Gomes *et al.* (2002) and Fraga Alves *et al.* (2003) and Goegebeur *et al.* (2010) introduced a new class of estimators based on the scaled log-excesses U_i. Further details on this topic can be found in Beirlant *et al.* (2012) and references within.

2.2 A new estimator for the second order parameter ρ

We will now propose a new estimator for the shape second order parameter ρ. First we will consider the ratio of a difference of estimators of the same parameter,

$$R_{n,k}^{(\tau)} = \frac{\left(N_{n,k}^{(1)}\right)^{\tau} - \left(N_{n,k}^{(3/2)}\right)^{\tau}}{\left(N_{n,k}^{(3/2)}\right)^{\tau} - \left(N_{n,k}^{(2)}\right)^{\tau}}, \quad \tau \in \mathcal{R}, \tag{2.4}$$

with $N_{n,k}^{(\alpha)} := \frac{\alpha}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1}U_i$, $\alpha \geq 1$, consistent estimators of $\gamma > 0$, and using again the notation $a^{b\tau} = b\ln a$ whenever $\tau = 0$. It will be shown that, under the second order condition in (1.2), $R_{n,k} = R_{n,k}^{(\tau)}$ will converge to $\frac{2-\rho}{1-\rho}$. Then, by inversion, we get the new estimator for the shape second order parameter ρ with functional expression,

$$\hat{\rho}_n^{CG(\tau)}(k) := 1 + \left(1 - R_{n,k}^{(\tau)}\right)^{-1}, \quad \tau \in \mathcal{R}. \tag{2.5}$$

3 Main asymptotic results

We shall next proceed with the study of the new ρ-estimator in (2.5). To compare the new estimator with others in the literature, we also present asymptotic results for the estimators in (2.2) and (2.3).

We will need to work with intermediate values of k, i.e., a sequence of integers $k = k_n$, $1 \leq k < n$, such that

$$k = k_n \to \infty \quad \text{and} \quad k_n = o(n), \quad \text{as} \quad n \to \infty. \tag{3.1}$$

In order to establish the asymptotic normality of the second-order estimators, it is necessary to further assume a third-order condition, ruling the rate of convergence in (1.2), and which guarantees that, for all $x > 0$,

$$\lim_{t\to\infty} \frac{\frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} - \frac{x^\rho - 1}{\rho}}{B(t)} = \frac{x^{\rho+\rho'} - 1}{\rho + \rho'}, \tag{3.2}$$

where $|B(t)|$ must then be of regular variation with index $\rho' \leq 0$. There appears then this extra non-positive third-order parameter $\rho' \leq 0$. Although ρ and ρ' can also be zero, we shall consider ρ, $\rho' < 0$.

Remark 3.1. *The third order condition in (3.2) holds for models with a tail quantile function*

$$U(t) = Cx^\gamma\big(1 + D_1 x^\rho + D_2 x^{\rho+\rho'} + o\big(x^{\rho+\rho'}\big)\big), \tag{3.3}$$

as $x \to \infty$, *with* $C > 0$, $D_1,\ D_2 \neq 0$, $\rho < 0$.

Remark 3.2. *Note that for most of the common heavy-tailed models ($\gamma > 0$), we have $\rho' = \rho$. Among those models we mention: the Fréchet model, with d.f. $F(x) = \exp(-x^{-1/\gamma})$, $x \geq 0$, for which $\rho' = \rho = -1$ and the generalized Pareto (GP) model, with d.f. $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, $x \geq 0$, for which $\rho' = \rho = -\gamma$.*

Theorem 3.1. *Let us assume that $\widehat{\rho}_n^\bullet(k)$ denotes any of the semi-parametric ρ-estimators defined in (2.2), (2.3) or (2.5). If the second order condition (1.2) holds, with $\rho < 0$, k is intermediate and such that $\sqrt{k}\, A(n/k) \to \infty$, as $n \to \infty$, then $\widehat{\rho}_n^\bullet(k)$ converges in probability to ρ. Under the third order framework in (3.2),*

$$\widehat{\rho}_n^\bullet(k) \ \stackrel{d}{=}\ \rho + \left(\frac{\sigma_\rho^\bullet W_k^\bullet}{\sqrt{k}A(n/k)} + b_1^\bullet A(n/k) + b_2^\bullet B(n/k)\right)(1 + o_p(1)), \tag{3.4}$$

with $W_k^\bullet$ an asymptotically standard normal r.v.,

$$\sigma_\rho^{FH} = \frac{\gamma\ (1-\rho)(1-2\rho)\sqrt{1-2\rho}}{|\rho|},$$

$$\sigma_\rho^{FAGH} = \frac{\gamma(1-\rho)^3\sqrt{2\rho^2 - 2\rho + 1}}{|\rho|},$$

$$\sigma_\rho^{CG} = \frac{\gamma\ (1-\rho)(2-\rho)(3-2\rho)\sqrt{4\rho^2 - 4\rho + 7}}{\sqrt{120}|\rho|},$$

and

$$b_1^{FH} = -\frac{(1-\rho)(1-2\rho)^2}{\gamma\rho(1-3\rho)^2}, \qquad b_2^{FH} = \frac{(1-\rho)(1-2\rho)(\rho+\rho')\rho'}{\rho(1-\rho-\rho')(1-2\rho-\rho')}$$

$$b_1^{FAGH} = \frac{\rho\left[\tau(1-2\rho)^2(3-\rho)(3-2\rho)+6\rho\left(4(2-\rho)(1-\rho)^2-1\right)\right]}{12\gamma(1-\rho)^2(1-2\rho)^2},$$

$$b_2^{FAGH} = \frac{\rho'(\rho+\rho')(1-\rho)^3}{\rho(1-\rho-\rho')^3},$$

$$b_1^{CG} = \frac{(1-\tau)\rho}{2\gamma}, \qquad b_2^{CG} = -\frac{(1-\rho)(2-\rho)(3-2\rho)\rho'(\rho+\rho')}{\rho(1-\rho-\rho')(2-\rho-\rho')(3-2\rho-2\rho')}.$$

Moreover, if $\sqrt{k}\,A^2(n/k) \to \lambda_A$ *and* $\sqrt{k}\,A(n/k)B(n/k) \to \lambda_B$ *(both finite), as* $n \to \infty$*, then* $\sqrt{k}A(n/k)(\hat\rho_n^\bullet(k)-\rho)$ *is asymptotically normal with mean value* $\lambda_A b_1^\bullet + \lambda_B b_2^\bullet$ *and variance* $(\sigma_\rho^\bullet)^2$*.*

Proof. For the estimators $\widehat{\rho}_n^{FAGH}(k)$ and $\widehat{\rho}_n^{FH}(k)$ the proof can be found in Fraga Alves *et al.* (2003) and Caeiro and Gomes (2011), respectively. Regarding the new estimator, $\widehat{\rho}_n^{CG}(k)$, we only need to prove the asymptotic representation in (3.4). Then, consistency and asymptotic normality follows straightforward. Notice that under the third order condition, in (3.2), and for intermediate k we have (Caeiro *et al.* 2009)

$$N_{n,k}^{(\alpha)} \overset{d}{=} \gamma + \frac{\gamma\,\alpha\,Z_k^{(\alpha)}}{\sqrt{(2\alpha-1)k}} + \frac{\alpha A(n/k)}{\alpha-\rho}\left(1+O_p\left(\frac{1}{\sqrt{k}}\right)\right) + \frac{\alpha A(n/k)B(n/k)}{\alpha-\rho-\rho'}(1+o_p(1)),$$

where $Z_k^{(\alpha)}$ is asymptotically standard normal. Then, the use of Taylor expansion $(1+x)^{-1} = 1-x+o(x^2)$, $x \to 0$ and after some cumbersome calculations we get

$$R_{n,k}^{(\tau)} \overset{d}{=} \frac{2-\rho}{1-\rho}\Big\{1 + \frac{\sigma_\rho^R Z_k^R}{\sqrt{k}A(n/n)} + \frac{(\tau-1)\rho}{2\gamma\,(1-\rho)(2-\rho)}A(n/k)\left(1+O_p\left(\frac{1}{\sqrt{k}}\right)\right) + \frac{(3-2\rho)\rho'(\rho+\rho')}{\rho(1-\rho-\rho')(2-\rho-\rho')(3-2\rho-2\rho')}B(n/k)(1+o_p(1)),$$

with Z_k^R an asymptotically standard normal r.v. and

$$\sigma_\rho^R = \frac{\gamma\,(3-2\rho)\sqrt{4\rho^2-4\rho+7}}{\sqrt{120}|\rho|}.$$

Using again Taylor expansion for $(1+x)^{-1}$, (3.4) follows. □

Remark 3.3. *From Theorem 3.1, we conclude that the tuning parameter* τ *affects the asymptotic bias of* $\hat\rho_n^{FAGH(\tau)}$ *and* $\hat\rho_n^{CG(\tau)}$*, but not the asymptotic variance. Consequentially if* $\rho = \rho'$ $(B(n/k) = O(A(n/k)))$*, we can always choose* $\tau = \tau_0$ *so that the asymptotic bias is null, even when* $\sqrt{k}\,A(n/k) \to \infty$ *and* $\sqrt{k}\,A^2(n/k) \to \lambda_A$*. Figure 1 show us the values of the quotients of the asymptotic standard deviations* $\sigma_\rho^{FAGH}/\sigma_\rho^{FH}$ *and* $\sigma_\rho^{CH}/\sigma_\rho^{FH}$*, for* $-4 \le \rho < 0$*. The patterns allow us to conclude that* $\hat\rho_n^{FH}$ *has the smallest asymptotic variance. Also* $\sigma_\rho^{CG} < \sigma_\rho^{FAGH}$ *if* $\rho < -0.2821$*.*

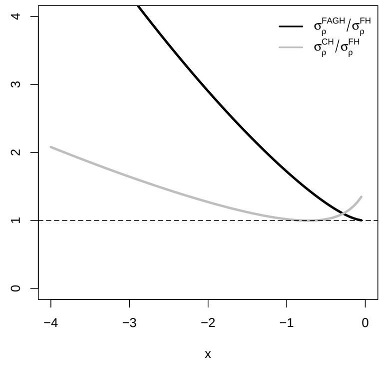

Figure 1: Pattern of the quotients of the asymptotic standard deviations $\sigma_\rho^{FAGH}/\sigma_\rho^{FH}$ and $\sigma_\rho^{CH}/\sigma_\rho^{FH}$.

4 Applications to simulated and real data

4.1 A case study in the field of insurance

We shall next consider an illustration through the analysis of automobile claim amounts exceeding 1,200,000 Euro over the period 1988-2001, gathered from several European insurance companies co-operating with the same re-insurer (Secura Belgian Re). This data set is available in Beirlant *et al.* (2004).

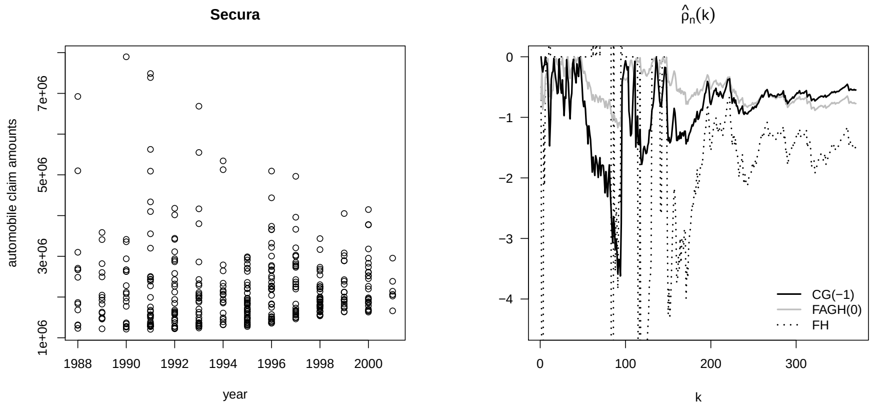

Figure 2: Secura data (*left*) and estimates of the second-order parameter ρ as function of k (*right*).

In Figure 2, we present the value of the automobile claim amounts by year (left) and the sample path of the estimates $\hat\rho_n^{FH}(k)$, $\hat\rho_n^{FAGH(\tau)}(k)$ and $\hat\rho_n^{CG(\tau)}(k)$ in (2.2), (2.3) and (2.5), respectively. We have chosen $\tau = 0$ and $\tau = -1$ for $\hat\rho_n^{FAGH(\tau)}$ and $\hat\rho_n^{CG(\tau)}$, respectively, based on the stability of the sample paths, as function of k. We conclude that, for large values of k, the estimates given by $\hat\rho_n^{FAGH(0)}$ and $\hat\rho_n^{CG(-1)}$ are very close and are both much stable than the estimates given by $\hat\rho_n^{FH}$.

4.2 Simulated data

We have implemented a multi-sample Monte Carlo simulation experiment, with 200 samples of size 2000 and 5000, to obtain the distributional behaviour of the estimators $\hat{\rho}_n^{FH}$, $\hat{\rho}_n^{FAGH(\tau)}$ and $\hat{\rho}_n^{CG(\tau)}$ in (2.2), (2.3) and (2.5), respectively, for the Fréchet model with $\gamma = 0.5$ and d.f. given in Remark 3.2.

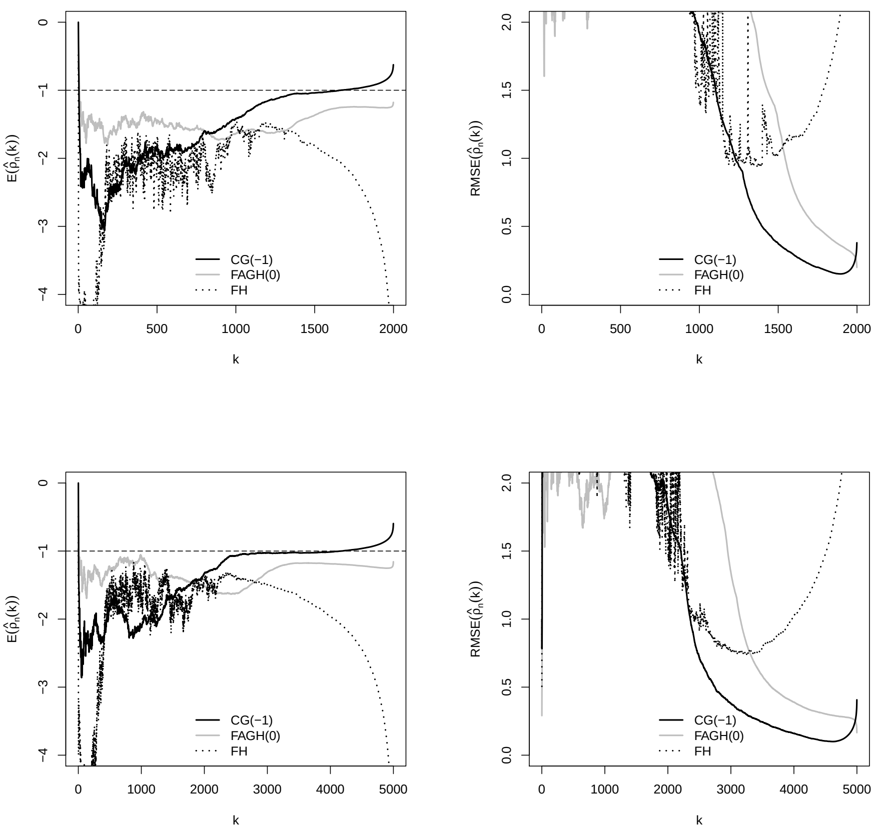

Figure 3: Simulated mean values (*left*) and root mean squared errors (*right*) for the Fréchet model, with $\gamma = 0.5$, $n = 2000$ (above) and $n = 5000$ (below).

In Figure 3, we present the simulated mean values (E) and root mean square errors (RMSE) patterns of the above mentioned ρ-estimators, as functions of k, for $n = 2000$ and $n = 5000$. Since we have $\rho = -1$, we have used $\tau = 0$ in $\hat{\rho}_n^{FAGH(\tau)}(k)$ (see Remark 2.1). Using again the stability of the sample paths of the mean values and the RMSE, for large k, we elect $\tau = -1$ in $\hat{\rho}_n^{CG(\tau)}(k)$.

Figure 3 evidences that, with the proper choice of τ, $\hat{\rho}_n^{CG(-1)}(k)$ has the best performance (not only in terms of bias but also in terms of RMSE), $\hat{\rho}_n^{FAGH(0)}(k)$ is the second best and finally $\hat{\rho}_n^{FH}(k)$ has the worst performance (although it is a maximum likelihood estimator). The adaptive choice of τ is outside the scope of this paper.

References

[1] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J.: Statistics of Extremes. Theory and Applications. Wiley (2004) Beirlant, J., Caeiro, F. and Gomes, M.I: . An Overview and Open Research Topics in the Field of Statistics of Univariate Extremes. Revstat **10**, 1–31 (2012)

[2] Caeiro, F., Gomes, M. I: A new class of estimators of the "scale" second order parameter. Extremes **9**, 193-211 (2006)

[3] Caeiro, F. and Gomes, M.I.: Asymptotic comparison at optimal levels of reduced-bias extreme value index estimators. Statistica Neerlandica **65**, 462–488 (2011)

[4] Caeiro, F., Gomes, M.I. and Henriques Rodrigues, L.: Reduced-bias tail index estimators under a third order framework. Communications in Statistics - Theory and Methods **38**(7), 1019–1040 (2009)

[5] Ciuperca, G., Mercadier, C.: Semi-parametric estimation for heavy tailed distributions. *Extremes* **13**, 55–87 (2010)

[6] Feuerverger, A., Hall, P.: Estimating a tail exponent by modelling departure from a Pareto distribution. Annals Statistics **27**, 760–781 (1999)

[7] Fraga Alves, M.I., Gomes, M.I. and Haan, L. de: A new class of semi-parametric estimators of the second order parameter, Portugaliae Mathematica, **60**(2), 193–213 (2003)

[8] Geluk, J. and de Haan, L.: Regular Variation, Extensions and Tauberian Theorems. CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam, Netherlands (1987)

[9] Goegebeur, Y., Beirlant, J. and de Wet, T.: Kernel estimators for the second order parameter in extreme value statistics. Journal of Statistical Planning and Inference, **140**, 2632–2652 (2010)

[10] Gomes, M.I., de Haan, L. and Peng, L.: Semi-parametric estimation of the second order parameter in statistics of extremes. Extremes **5**, 387–414 (2002)

[11] Hall, P. and Welsh, A.H.: Adaptative estimates of parameters of regular variation, Ann. Statist. **13**, 331–341 (1985)