

Uma classe de estimadores do parâmetro de escala de segunda ordem

Frederico Caeiro

Dep. de Matemática e C.M.A., Faculdade de Ciências e Tecnologia - Universidade Nova de Lisboa

M. Ivette Gomes

D.E.I.O. e C.E.A.U.L., Faculdade de Ciências, Universidade de Lisboa

Resumo: Para modelos F pertencentes à classe de Hall, a função quantil $U(t)$ é de variação regular com índice igual a γ , onde $\gamma > 0$ é o índice de cauda. Para $x > 0$, a velocidade de convergência de $U(tx)/U(t) - x^\gamma$ para zero, quando $t \rightarrow \infty$ pode ser medida através da função $A(t) = \gamma \beta t^\rho$, $\rho < 0$, $\beta \in \mathbb{R}$. A estimação de ρ , o parâmetro de “forma” de segunda ordem, é um tema bastante abordado na literatura, mas pouco existe relativamente à estimação do parâmetro de “escala” de segunda ordem β . Num contexto semi-paramétrico, introduzimos uma classe de estimadores de β e provamos a sua consistência. Serão também indicadas condições que garantem a sua normalidade assintótica. Usando o método de Monte Carlo estudamos, para amostras de dimensão finita de alguns modelos de cauda pesada, as propriedades desta classe de estimadores de β .

Palavras-chave: caudas pesadas, estimação semi-paramétrica, parâmetros de segunda ordem, método de Monte Carlo

Abstract: In Hall’s class of heavy-tailed models, the quantile function $U(t)$ is of regular variation with an index equal to γ , where $\gamma > 0$ is the tail index. For every $x > 0$, the rate of convergence of $U(tx)/U(t) - x^\gamma$ towards zero, as $t \rightarrow \infty$ may be measured through a function $A(t) = \gamma \beta t^\rho$, $\rho < 0$, $\beta \in \mathbb{R}$. The estimation of ρ , the “shape” second order parameter, has been extensively addressed in the literature, but practically nothing has been done related to the estimation of the “scale” second order parameter β . Under a semi-parametric framework, we shall introduce a class of β -estimators and study their consistency. We shall also deal with the conditions enabling us to get the asymptotic normality of this class of estimators, and we shall illustrate the behaviour of the estimators, through Monte Carlo simulation techniques, for a wide variety of heavy-tailed models.

Keywords: heavy tails, semi-parametric estimation, second order parameters, Monte Carlo methodology

Classificação MSC2000: 62G20, 62G32, 65C05.

1 Introdução

Os modelos de cauda pesada são muito úteis em áreas tais como o tráfego de telecomunicações ou a análise de dados financeiros. Um modelo F é de cauda pesada se a função cauda $\bar{F} = 1 - F$ é de variação regular com índice negativo $-1/\gamma$, $\gamma > 0$, ou equivalentemente a função quantil, $U(t) = F^{\leftarrow}(1 - 1/t)$, $t \geq 1$, com $F^{\leftarrow}(x) = \inf\{y : F(y) \geq x\}$, é de variação regular com índice γ . O parâmetro $\gamma (> 0)$ é o índice de cauda, um dos parâmetros mais importantes em teoria de valores extremos.

É usual admitir-se uma condição de segunda ordem que mede a velocidade de convergência de $U(tx)/U(t) - x^\gamma$ para zero, para qualquer $x > 0$, onde $|A(t)|$ é uma função de variação regular de índice $\rho \leq 0$ (Geluk e de Haan, 1987):

$$\lim_{t \rightarrow \infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}. \quad (1)$$

Neste trabalho vamos supor que F é um modelo pertencente à classe de Hall (Hall, 1982; Hall e Welsh, 1985), com

$$A(t) = \gamma \beta t^\rho, \quad \rho < 0, \quad \beta \in \mathbb{R}, \quad (2)$$

onde β é o parâmetro de “escala” de segunda ordem que interessa estimar.

Para obter algumas propriedades assintóticas é ainda necessário considerar uma condição de terceira ordem,

$$\lim_{t \rightarrow \infty} \frac{\frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} - \frac{x^\rho - 1}{\rho}}{B(t)} = \frac{x^{\rho+\eta} - 1}{\rho + \eta},$$

para qualquer $x > 0$, onde $\eta < 0$ é o parâmetro de terceira ordem e $|B(t)| \in RV_\eta$. Para muitos dos modelos de cauda pesada, tais como por exemplo o *Fréchet*, *Burr*, *Cauchy* e o *t-Student*, o parâmetro de terceira ordem, η , é igual ao de segunda. Assim, de forma a evitar a necessidade de estimar um parâmetro adicional, vamos considerar a condição anterior com $\eta = \rho < 0$, ou seja vamos supor que:

$$\ln \frac{U(tx)}{U(t)} = \gamma \ln x + A(t) \left(\frac{x^\rho - 1}{\rho} \right) + A(t)B(t) \left(\frac{x^{2\rho} - 1}{2\rho} \right) + o(t^{2\rho}) \quad (3)$$

2 A classe de estimadores de β

De modo a estimar o parâmetro β , presente em (2), vamos utilizar o mesmo tipo de estatísticas consideradas na estimação de ρ em Fraga Alves *et al.* (2003).

Proposição 2.1. *Seja E uma variável aleatória com distribuição exponencial com valor médio igual a 1, e $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$, $t > 0$ a função Gama*

completa. Então defina-se, para $\alpha > 0$,

$$\begin{aligned} \mu_\alpha^{(1)} &:= \mathbb{E}[E^\alpha] = \Gamma(\alpha + 1), \\ \bar{\sigma}_\alpha^{(1)} &:= \frac{\sqrt{V[E^\alpha]}}{\mu_\alpha^{(1)}} = \frac{\sqrt{\Gamma(2\alpha + 1) - \Gamma^2(\alpha + 1)}}{\alpha\Gamma(\alpha)} = \sqrt{\frac{2\Gamma(2\alpha)}{\alpha\Gamma^2(\alpha)} - 1}, \\ \bar{\mu}_\alpha^{(2)}(\rho) &:= \frac{1}{\mu_\alpha^{(1)}} \mathbb{E}\left[E^{\alpha-1} \left(\frac{e^{\rho E} - 1}{\rho}\right)\right] = \frac{1 - (1 - \rho)^\alpha}{\alpha \rho (1 - \rho)^\alpha}, \\ \bar{\sigma}_\alpha^{(2)}(\rho) &:= \frac{\sqrt{V\left[E^{\alpha-1} \left(\frac{e^{\rho E} - 1}{\rho}\right)\right]}}{\mu_\alpha^{(1)}} = \frac{\sqrt{\mu_{2\alpha}^{(3)}(\rho) - \left(\mu_\alpha^{(2)}(\rho)\right)^2}}{\mu_\alpha^{(1)}}, \\ \bar{\mu}_\alpha^{(3)}(\rho) &:= \frac{1}{\mu_\alpha^{(1)}} \mathbb{E}\left[E^{\alpha-2} \left(\frac{e^{\rho E} - 1}{\rho}\right)^2\right] = \frac{2\{\mu_{\alpha-1}^{(2)}(2\rho) - \mu_{\alpha-1}^{(2)}(\rho)\}}{\rho \mu_\alpha^{(1)}} \\ &= \begin{cases} \frac{1}{\rho^2} \ln \frac{(1-\rho)^2}{1-2\rho} & \text{se } \alpha = 1 \\ \frac{1}{\rho^2 \alpha (\alpha-1)} \left\{ \frac{1}{(1-2\rho)^{\alpha-1}} - \frac{2}{(1-\rho)^{\alpha-1}} + 1 \right\} & \text{se } \alpha \neq 1 \end{cases}. \end{aligned}$$

Considere-se as seguintes estatísticas

$$M_n^{(\alpha)}(k) := \frac{1}{k} \sum_{i=1}^k \{\ln X_{n-i+1:n} - \ln X_{n-k:n}\}^\alpha, \quad \alpha > 0, \quad (4)$$

$$D_n^{(\alpha,\tau)}(k) := \left(\frac{M_n^{(\alpha)}(k)}{\Gamma(\alpha + 1)}\right)^\tau - \left(\frac{M_n^{(2\alpha)}(k)}{\Gamma(2\alpha + 1)}\right)^{\tau/2}, \quad \alpha > 0, \quad \tau \in \mathbb{R} \quad (5)$$

onde $M_n^{(1)}(k)$ é o clássico estimador de Hill (Hill, 1975) e $X_{i:n}$ representa a i -ésima estatística ordinal ascendente associada à amostra (X_1, X_2, \dots, X_n) . Usando a notação $d_\alpha(\rho) := \bar{\mu}_\alpha^{(2)}(\rho) - \bar{\mu}_{2\alpha}^{(2)}(\rho)$ vai-se considerar, para qualquer estimador consistente de ρ , a seguinte classe de estatísticas:

$$\hat{\beta}_{\hat{\rho}}^{(\alpha,\tau)}(k) := \frac{2d_{2\alpha}(\hat{\rho})}{\alpha\tau (d_\alpha(\hat{\rho}))^2} \left(\frac{k}{n}\right)^{\hat{\rho}} \frac{\left(D_n^{(\alpha,\tau)}(k)\right)^2}{D_n^{(2\alpha,\tau)}(k)} \quad (6)$$

onde $\tau \in \mathbb{R}$, e $\alpha > 0$ são parâmetros de controlo e $\hat{\rho}$ um estimador consistente de ρ . Após o estudo das propriedades assintóticas, o parâmetro α será fixado e utiliza-se apenas τ como parâmetro de controlo.

3 Propriedades assintóticas

No que se segue vai-se admitir que $k = k(n)$ é uma sequência intermédia, isto é, uma sequência de valores inteiros tal que

$$k = k(n) \rightarrow \infty, \quad k(n) = o(n), \quad n \rightarrow \infty, \quad (7)$$

Proposição 3.1 (Consistência). *Supondo que é válida a condição de segunda ordem, (1), com $A(t) = \gamma \beta t^\rho$, $\rho < 0$, que $\lim_{n \rightarrow \infty} \sqrt{k} A(n/k) = \infty$, e que k é uma sequência intermédia, a classe de estimadores $\widehat{\beta}_\rho^{(\alpha, \tau)}(k)$, $\tau \in \mathbb{R}$ e $\alpha > 0$ é consistente.*

Demonstração. Pode ser visto em Gomes *et al.* (2002) ou em Fraga Alves *et al.* (2003), que

$$\begin{aligned} \left(\frac{M_n^{(\alpha\theta)}(k)}{\Gamma(\alpha\theta + 1)} \right)^{\tau/\theta} &\stackrel{d}{=} \gamma^{\alpha\tau} \left(1 + \frac{\tau \bar{\sigma}_{\alpha\theta}^{(1)} P_n^{(\alpha\theta)}}{\theta \sqrt{k}} + \frac{\tau\alpha}{\gamma} \bar{\mu}_{\alpha\theta}^{(2)}(\rho) A(n/k) \right. \\ &\quad + \frac{\tau\alpha}{\gamma} \bar{\sigma}_{\alpha\theta}^{(2)}(\rho) \frac{A(n/k)}{\sqrt{k}} \bar{P}_n^{(\alpha\theta)} \\ &\quad + \frac{\tau\alpha}{2\gamma^2} A^2(n/k) \left((\alpha\theta - 1) \bar{\mu}_{\alpha\theta}^{(3)}(\rho) + \alpha(\tau - \theta) (\bar{\mu}_{\alpha\theta}^{(2)}(\rho))^2 \right) (1 + o_p(1)) \\ &\quad \left. + \frac{\tau\alpha}{\gamma} \bar{\mu}_{\alpha\theta}^{(2)}(2\rho) A(n/k) B(n/k) (1 + o_p(1)) \right), \end{aligned}$$

onde $P_n^{(\alpha\theta)}$ e $\bar{P}_n^{(\alpha\theta)}$ são assintoticamente normais padrão. Assim,

$$\begin{aligned} D_n^{(\alpha, \tau)}(k) &:= \left(\frac{M_n^{(\alpha)}(k)}{\Gamma(\alpha + 1)} \right)^\tau - \left(\frac{M_n^{(2\alpha)}(k)}{\Gamma(2\alpha + 1)} \right)^{\tau/2} \\ &\stackrel{d}{=} \alpha\tau\gamma^{\tau\alpha-1} \left(\frac{\gamma W_\alpha}{\alpha\sqrt{k}} + A(n/k) d_\alpha(\rho) \right. \\ &\quad \left. + A(n/k) \left(\frac{1}{2\gamma} A(n/k) a_\alpha^{(\tau)} + B(n/k) d_\alpha(2\rho) \right) (1 + o_p(1)) \right. \\ &\quad \left. + \frac{A(n/k)}{\sqrt{k}} \left(\bar{\sigma}_\alpha^{(2)}(\rho) \bar{P}_n^{(\alpha)} - \bar{\sigma}_{2\alpha}^{(2)}(\rho) \bar{P}_n^{(2\alpha)} \right) \right), \end{aligned}$$

com

$$\begin{aligned} a_\alpha^{(\tau)} &:= (\alpha - 1) \bar{\mu}_\alpha^{(3)}(\rho) + \alpha(\tau - 1) (\bar{\mu}_\alpha^{(2)}(\rho))^2 - (2\alpha - 1) \bar{\mu}_{2\alpha}^{(3)}(\rho) \\ &\quad - \alpha(\tau - 2) (\bar{\mu}_{2\alpha}^{(2)}(\rho))^2, \\ W_\alpha &:= \bar{\sigma}_\alpha^{(1)} P_n^{(\alpha)} - \frac{\bar{\sigma}_{2\alpha}^{(1)} P_n^{(2\alpha)}}{2}. \end{aligned}$$

Para k intermédio, e com $\sqrt{k} A(n/k) \xrightarrow[n \rightarrow \infty]{} \infty$, i.e., $1/\sqrt{k} = o(A(n/k))$, o segundo termo da representação anterior, em distribuição, é o dominante. O quarto termo é $o_p(A(n/k))$ e pode ser desprezado. Consequentemente, para modelos na classe de Hall com $A(t) = \gamma \beta t^\rho$, tem-se:

$$\frac{D_n^{(\alpha, \tau)}(k)}{A(n/k)} = \left(\frac{k}{n} \right)^\rho \frac{D_n^{(\alpha, \tau)}(k)}{\gamma \beta} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \alpha \tau \gamma^{\alpha\tau-1} d_\alpha(\rho).$$

resulta que

$$\frac{\left(\left(\frac{k}{n}\right)^\rho D_n^{(\alpha,\tau)}(k)\right)^2}{\left(\frac{k}{n}\right)^\rho D_n^{(2\alpha,\tau)}(k)} = \left(\frac{k}{n}\right)^\rho \frac{\left(D_n^{(\alpha,\tau)}(k)\right)^2}{D_n^{(2\alpha,\tau)}(k)} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \frac{1}{2} \beta \alpha \tau \frac{(d_\alpha(\rho))^2}{d_{2\alpha}(\rho)},$$

ou seja,

$$\frac{2 d_{2\alpha}(\rho)}{\alpha \tau (d_\alpha(\rho))^2} \left(\frac{k}{n}\right)^\rho \frac{\left(D_n^{(\alpha,\tau)}(k)\right)^2}{D_n^{(2\alpha,\tau)}(k)} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \beta.$$

□

O resultado anterior continua a ser válido se ρ for estimado através de um estimador consistente.

Proposição 3.2 (Normalidade). *Se se verificarem as condições da proposição anterior, e a condição de terceira ordem, (3), $\widehat{\beta}_\rho^{(\alpha,\tau)}(k)$ é assintoticamente normal. O valor médio assintótico de $\sqrt{k}A(n/k) \left(\widehat{\beta}_\rho^{(\alpha,\tau)}(k) - \beta\right)$ é nulo sempre que $\sqrt{k} A^2(n/k)$ e $\sqrt{k} A(n/k) B(n/k)$ convergirem para zero.*

Se $\sqrt{k}A^2(n/k) \xrightarrow[n \rightarrow \infty]{} \lambda_A$ e $\sqrt{k}A(n/k)B(n/k) \xrightarrow[n \rightarrow \infty]{} \lambda_{AB}$, eventualmente não nulos, então $\sqrt{k}A(n/k)(\widehat{\beta}_\rho^{(\alpha,\tau)}(k) - \beta)$ tem assintoticamente distribuição normal com valor médio

$$\frac{1}{\gamma} \left(\frac{a_\alpha^{(\tau)}}{d_\alpha(\rho)} - \frac{a_{2\alpha}^{(\tau)}}{2d_{2\alpha}(\rho)} \right) \lambda_A + \left(\frac{2d_\alpha(2\rho)}{d_\alpha(\rho)} - \frac{d_{2\alpha}(2\rho)}{d_{2\alpha}(\rho)} \right) \lambda_{AB}. \quad (8)$$

A variância assintótica de $\sqrt{k}A(n/k) \left(\widehat{\beta}_\rho^{(\alpha,\tau)}(k) - \beta\right)$ é dada por

$$\sigma^2 \left(\widehat{\beta}_\rho^{(\alpha,\tau)}(k) \right) = \left(\frac{\gamma\beta}{\alpha} \right)^2 V \left[\frac{2W_\alpha}{d_\alpha(\rho)} - \frac{W_{2\alpha}}{2d_{2\alpha}(\rho)} \right]. \quad (9)$$

Demonstração. A demonstração é consequência imediata dos seguintes resultados:

$$\begin{aligned} \left(\frac{D_n^{(\alpha,\tau)}}{A(n/k)d_\alpha(\rho)} \right)^{-1} &\stackrel{d}{=} \alpha \tau \gamma^{\tau\alpha-1} \left[1 - \frac{\gamma W_\alpha}{\alpha \sqrt{k} A(n/k) d_\alpha(\rho)} \right. \\ &\quad \left. + \left(\frac{1}{2\gamma} \frac{a_\alpha^{(\tau)}}{d_\alpha(\rho)} A(n/k) + \frac{d_\alpha(2\rho)}{d_\alpha(\rho)} B(n/k) \right) \right] (1 + o_p(1)), \\ \left(\frac{D_n^{(\alpha,\tau)}}{A(n/k)d_\alpha(\rho)} \right)^2 &\stackrel{d}{=} (\alpha \tau \gamma^{\tau\alpha-1})^2 \left[1 + \frac{2\gamma W_\alpha}{\alpha \sqrt{k} A(n/k) d_\alpha(\rho)} \right. \\ &\quad \left. + \left(\frac{1}{\gamma} \frac{a_\alpha^{(\tau)}}{d_\alpha(\rho)} A(n/k) + \frac{2d_\alpha(2\rho)}{d_\alpha(\rho)} B(n/k) \right) \right] (1 + o_p(1)). \end{aligned}$$

Como se admitiu que $A(t) = \gamma \beta t^\rho$, resulta que:

$$\begin{aligned} \widehat{\beta}_\rho^{(\alpha, \tau)}(k) &= \frac{2d_{2\alpha}(\rho)}{\alpha\tau [d_\alpha(\rho)]^2} \left(\frac{k}{n}\right)^\rho \frac{[D_n^{(\alpha, \tau)}(k)]^2}{D_n^{(2\alpha, \tau)}(k)} \stackrel{a}{=} \\ &\stackrel{a}{=} \beta \left[1 + \frac{\gamma}{\alpha\sqrt{k}A(n/k)} \left(\frac{2W_\alpha}{d_\alpha(\rho)} - \frac{W_{2\alpha}}{2d_{2\alpha}(\rho)} \right) + \frac{1}{\gamma} \left(\frac{a_\alpha^{(\tau)}}{d_\alpha(\rho)} - \frac{a_{2\alpha}^{(\tau)}}{2d_{2\alpha}(\rho)} \right) \times \right. \\ &\quad \left. \times A(n/k)(1 + o_p(1)) + \left(\frac{2d_\alpha(2\rho)}{d_\alpha(\rho)} - \frac{d_{2\alpha}(2\rho)}{d_{2\alpha}(\rho)} \right) B(n/k)(1 + o_p(1)) \right]. \end{aligned}$$

Consequentemente,

$$\begin{aligned} \sqrt{k}A(n/k) \left(\widehat{\beta}_\rho^{(\alpha, \tau)}(k) - \beta \right) &\stackrel{a}{=} \frac{\beta\gamma}{\alpha} \left(\frac{2W_\alpha}{d_\alpha(\rho)} - \frac{W_{2\alpha}}{2d_{2\alpha}(\rho)} \right) + \\ &+ \frac{\beta}{\gamma} \left(\frac{a_\alpha^{(\tau)}}{d_\alpha(\rho)} - \frac{a_{2\alpha}^{(\tau)}}{2d_{2\alpha}(\rho)} \right) \sqrt{k}A^2(n/k)(1 + o_p(1)) \\ &+ \beta \left(\frac{2d_\alpha(2\rho)}{d_\alpha(\rho)} - \frac{d_{2\alpha}(2\rho)}{d_{2\alpha}(\rho)} \right) \sqrt{k}A(n/k)B(n/k)(1 + o_p(1)), \end{aligned}$$

ou seja $\sqrt{k}A(n/k) \left(\widehat{\beta}_\rho^{(\alpha, \tau)}(k) - \beta \right)$ tem assintoticamente distribuição normal. O cálculo do valor médio e variância é imediato. \square

Com o objectivo de calcular a variância, (9), apresenta-se a proposição:

Proposição 3.3. *Nas condições da proposição anterior, e usando os resultados presentes em Gomes e Martins (2001) ou em Caeiro (2001), tem-se que:*

$$\begin{aligned} Cov(W_\alpha, W_{2\alpha}) &= \frac{1}{2\alpha} \left[\frac{3\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)} - \frac{3\Gamma(5\alpha)}{4\Gamma(\alpha)\Gamma(4\alpha)} \right. \\ &\quad \left. - \frac{\Gamma(4\alpha)}{\Gamma(2\alpha)^2} + \frac{3\Gamma(6\alpha)}{8\Gamma(2\alpha)\Gamma(4\alpha)} \right] - \frac{1}{4} \\ V[W_\alpha] &= \frac{2}{\alpha} \left[\frac{\Gamma(2\alpha)}{\Gamma^2(\alpha)} + \frac{\Gamma(4\alpha)}{8\Gamma^2(2\alpha)} - \frac{3\Gamma(3\alpha)}{4\Gamma(\alpha)\Gamma(2\alpha)} \right] - \frac{1}{4}. \end{aligned}$$

Das expressões anteriores conclui-se que a variância do estimador não depende do valor do parâmetro de controlo τ . Será por esta razão que se irá fixar o valor de α tendo em conta o efeito deste parâmetro na variância assintótica de $\widehat{\beta}_\rho^{(\alpha, \tau)}(k)$. Há também que considerar o efeito da escolha de α na duração de um estudo de simulação pelo método de Monte Carlo. Sabe-se que se o parâmetro da estatística (4) for um valor inteiro, é possível escrever os algoritmos de modo a tornar a simulação mais rápida.

Embora não se apresente uma expressão analítica da variância assintótica de $\sqrt{k}A(n/k) \left(\widehat{\beta}_\rho^{(\alpha, \tau)}(k) - \beta \right)$ como função de α , na Figura 1 apresenta-se, para dois valores de β , o seu gráfico. A figura indica que perto do valor $\alpha = 1/2$ a variância é mínima. Tendo em conta o trabalho de simulação que é necessário realizar, optou-se por fixar $\alpha = 1$ (valor inteiro onde a variância é mínima).

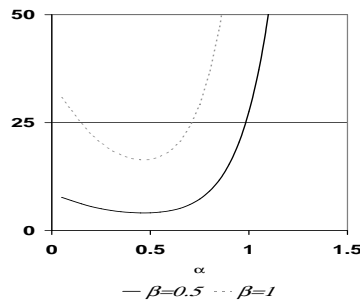


Figura 1: Variância assintótica de $\sqrt{k}A(n/k) \left(\widehat{\beta}_\rho^{(\alpha, \tau)}(k) - \beta \right)$ com $\gamma = 1$ e $\rho = -1$.

Se se fizer $\alpha = 1$ obtemos a classe de estimadores, dependente apenas do parâmetro de controlo $\tau \neq 0$,

$$\widehat{\beta}_{\widehat{\rho}}^{(\tau)}(k) := -\frac{2(2 - \widehat{\rho})^2}{\tau \widehat{\rho}} \left(\frac{k}{n} \right)^{\widehat{\rho}} \frac{\left\{ \left(M_n^{(1)}(k) \right)^\tau - \left(M_n^{(2)}(k)/2 \right)^{\tau/2} \right\}^2}{\left(M_n^{(2)}(k)/2 \right)^\tau - \left(M_n^{(4)}(k)/24 \right)^{\tau/2}},$$

com o usual prolongamento por continuidade para $\tau = 0$, dado por

$$\widehat{\beta}_{\widehat{\rho}}^{(0)}(k) := -\frac{2(2 - \widehat{\rho})^2}{\widehat{\rho}} \left(\frac{k}{n} \right)^{\widehat{\rho}} \frac{\left\{ \ln \left(M_n^{(1)}(k) \right) - \frac{1}{2} \ln \left(M_n^{(2)}(k)/2 \right) \right\}^2}{\ln \left(M_n^{(2)}(k)/2 \right) - \frac{1}{2} \ln \left(M_n^{(4)}(k)/24 \right)}.$$

4 Estimação do parâmetro de segunda ordem ρ

Vamos considerar a classe de estimadores introduzida em Fraga Alves *et al.* (2003) para estimar o parâmetro de segunda ordem ρ . Trata-se de uma classe de estimadores semi-paramétricos que apresentam trajetórias amostrais de k vs $\widehat{\rho}(k)$ bastante estáveis, para uma vasta região de valores elevados de k . Seja

$$\widehat{\rho}_\nu(k) = \widehat{\rho}(k) := - \left| \frac{3(T_n^{(\nu)}(k) - 1)}{T_n^{(\nu)}(k) - 3} \right|, \quad (10)$$

onde

$$T_n^{(\nu)}(k) = \begin{cases} \frac{(M_n^{(1)}(k))^\nu - (M_n^{(2)}(k)/2)^{\nu/2}}{(M_n^{(2)}(k)/2)^{\nu/2} - (M_n^{(3)}(k)/6)^{\nu/3}}, & \text{se } \nu \neq 0, \\ \frac{\ln(M_n^{(1)}(k)) - \frac{1}{2} \ln(M_n^{(2)}(k)/2)}{\frac{1}{2} \ln(M_n^{(2)}(k)/2) - \frac{1}{3} \ln(M_n^{(3)}(k)/6)} & \text{se } \nu = 0. \end{cases} \quad (11)$$

Os resultados teóricos e de simulação apresentados em Fraga Alves *et al.* (2003) e em Gomes e Martins (2002) levam-nos a utilizar o nível

$$k_1 = \min(n - 1, 2n / \ln \ln n) \quad (12)$$

(não é escolhido de forma óptima) e a utilização dos parâmetros de controlo $\nu = 0$ se $\rho \geq -1$ e $\nu = 1$ se $\rho < -1$.

5 Propriedades para amostras de dimensão finita

Os resultados que se seguem obtiveram-se a partir de simulações com 10×1000 amostras de dimensão n . Com as simulações pretendeu-se obter, para $\widehat{\beta}_\rho^{(\tau)}(k)$, estimativas do valor médio, $\mathbb{E}[\cdot]$, e do erro médio quadrático médio, $MSE[\cdot]$. O parâmetro de controlo, $\tau \in \mathbb{R}$ varia consoante o modelo já que foi escolhido de modo a obter-se, para valores elevados de k , estabilidade da trajetória de $(k, \widehat{\beta}_\rho^{(\tau)}(k))$ junto ao verdadeiro valor de β .

5.1 Modelos utilizados

- Modelo *Burr*, com $F(x) = 1 - (1 + x^{-\rho/\gamma})^{-1/\gamma}$. Verifica-se que

$$\ln \frac{U(tx)}{U(t)} = \gamma \ln x + \gamma t^\rho \left(\frac{x^\rho - 1}{\rho} \right) + \gamma t^{2\rho} \left(\frac{x^{2\rho} - 1}{2\rho} \right) + o(t^{2\rho})$$

Trata-se de um modelo que verifica (3) onde ρ é o parâmetro de segunda ordem independente de γ tendo-se $\beta = 1$.

- Modelo *Fréchet*, com $F(x) = \exp(-x^{-1/\gamma})$. Para este modelo,

$$\ln \frac{U(tx)}{U(t)} = \gamma \ln x + \frac{\gamma}{2t} \left(\frac{x^{-1} - 1}{-1} \right) + \frac{5\gamma}{12t^2} \left(\frac{x^{-2} - 1}{-2} \right) + o(t^{-2}).$$

Assim resulta que $\rho = -1$ e $\beta = \frac{1}{2}$.

- Modelo *Cauchy*, $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x)$. Como

$$\ln \frac{U(tx)}{U(t)} = \ln x + \frac{2\pi^2}{3t^2} \left(\frac{x^{-2} - 1}{-2} \right) + \frac{14\pi^4}{45t^4} \left(\frac{x^{-4} - 1}{-4} \right) + o(t^{-4}),$$

este modelo verifica a condição de terceira ordem com $\gamma = 1$, $\rho = -2$ e $\beta = \frac{2}{3}\pi^2 \approx 6.58$.

5.2 Resultados

Apresentamos nas Figuras 2, 3, 4, 5 e 6 e na Tabela 1 os valores simulados do valor médio e do erro médio quadrático. Nas figuras, as estimativas foram baseadas em amostras de dimensão 5000, excepto para o modelo *Cauchy* onde se usou o valor 10000 e não se fez qualquer alteração na localização. Nas figuras apenas não se apresentam as estimativas para pequenos valores de k já que nessa região existe uma enorme volatilidade das trajectórias. Na Tabela 1 consta o valor médio e erro médio quadrático no respectivo nível óptimo (nível onde o erro médio quadrático é mínimo).

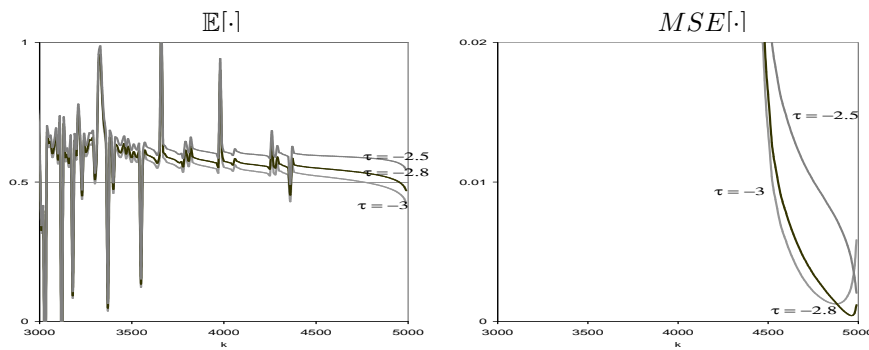


Figura 2: Modelo *Fréchet* com $\gamma = 1$ ($\rho = -0.5, \beta = 0.5$)

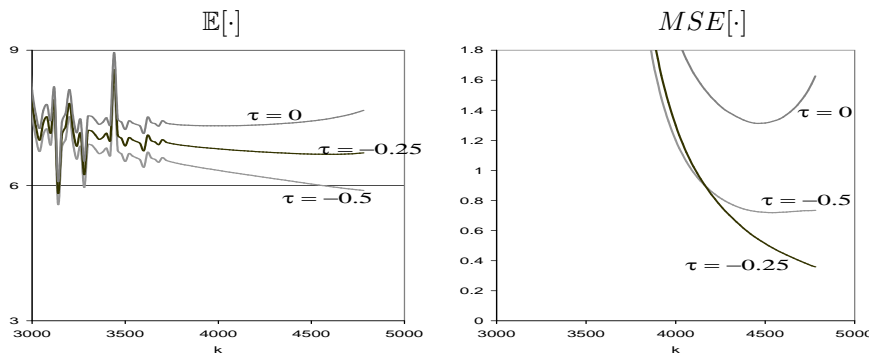


Figura 3: Modelo *Cauchy* ($\gamma = 1, \rho = -2, \beta \approx 6.58$)

Dos resultados apresentados podemos tirar as seguintes conclusões:

- A escolha do parâmetro de controlo τ , para o qual existe estabilidade das trajectórias amostrais junto a β , parece depender do modelo F . Embora

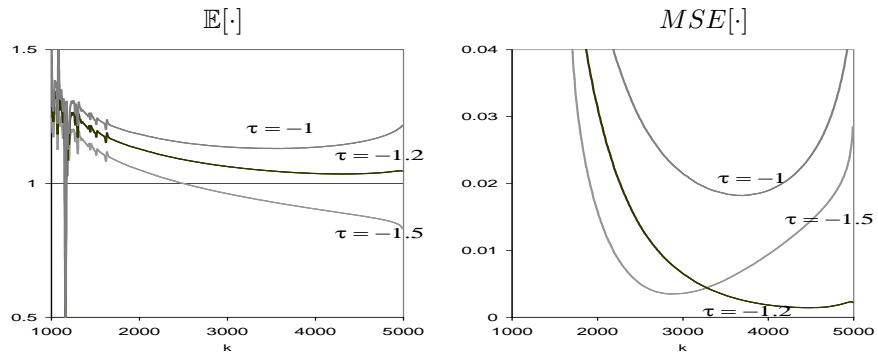
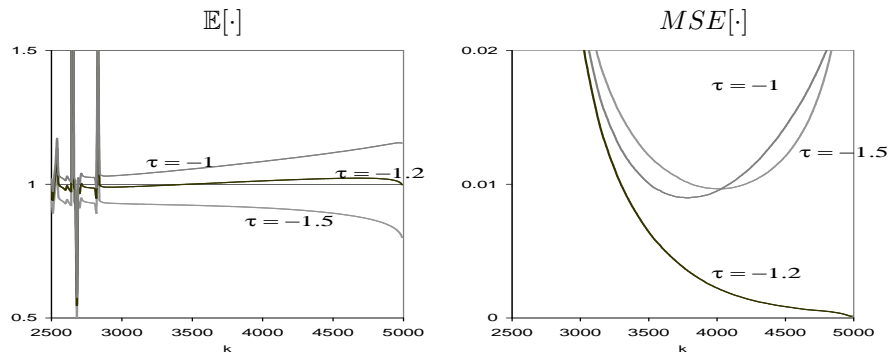
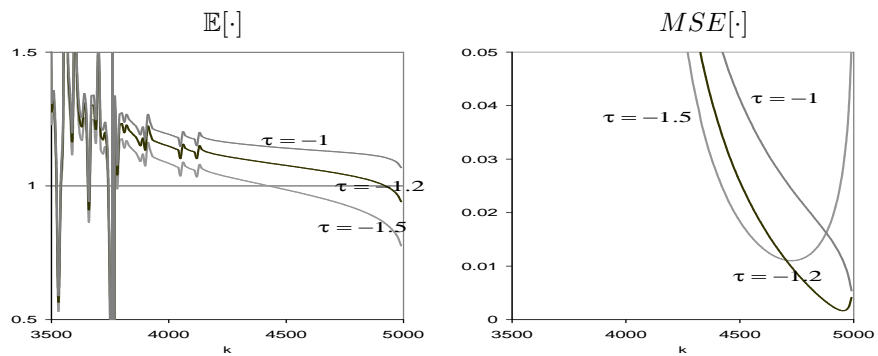
Figura 4: Modelo *Burr* com $\gamma = 1$ e $\rho = -0.5$ ($\beta = 1$)Figura 5: Modelo *Burr* com $\gamma = 1$ e $\rho = -1$ ($\beta = 1$)Figura 6: Modelo *Burr* com $\gamma = 1$ e $\rho = -2$ ($\beta = 1$)

Tabela 1: Valor médio e erro médio quadrático de $\hat{\beta}_{\hat{\rho}}^{(\tau)}(k)$, no nível óptimo.

n	$\mathbb{E}[\cdot]$			$MSE[\cdot]$		
Modelo <i>Fréchet</i> com $\gamma = 1$ ($\rho = -1, \beta = 0.5$)						
	$\tau = -3$	$\tau = -2.8$	$\tau = -2.5$	$\tau = -3$	$\tau = -2.8$	$\tau = -2.5$
100	0.5028	0.5450	0.6155	<u>1.0466</u>	1.1054	1.2082
500	0.4386	0.4653	0.5397	0.0137	<u>0.0101</u>	0.0112
1000	0.4542	0.4753	0.5295	0.0067	0.0041	<u>0.0040</u>
2000	0.4681	0.4853	0.5155	0.0033	0.0015	<u>0.0010</u>
5000	0.4832	0.4946	0.5061	0.0012	0.0004	<u>0.0003</u>
10000	0.4917	0.4982	0.5012	0.0004	<u>0.0001</u>	0.0002
25000	0.4967	0.4996	0.5002	0.0001	<u>0.0000</u>	0.0001
50000	0.4984	0.4999	0.5021	0.0001	<u>0.0000</u>	0.0001
Modelo <i>Cauchy</i> com $\gamma = 1, \rho = -2$ ($\beta \approx 6.58$)						
	$\tau = -0.5$	$\tau = -0.25$	$\tau = 0$	$\tau = -0.5$	$\tau = -0.25$	$\tau = 0$
500	6.6170	7.3656	8.2473	<u>4.8652</u>	6.5211	10.1113
1000	6.9465	7.7840	8.7259	<u>1.3770</u>	3.0141	6.6227
2000	6.9964	7.8996	8.8559	<u>0.5429</u>	2.2406	5.9259
5000	6.1972	7.0426	7.8278	<u>0.4104</u>	0.5604	2.1676
10000	6.0143	6.7547	7.4123	0.7135	<u>0.3467</u>	1.3081
25000	6.0151	6.6321	7.1762	0.7519	<u>0.2338</u>	0.8521
50000	6.0539	6.6012	7.0799	0.6615	<u>0.1644</u>	0.6091
Modelo <i>Burr</i> com $\gamma = 1$ e $\rho = -0.5$ ($\beta = 1$)						
	$\tau = -1.5$	$\tau = -1.2$	$\tau = -1$	$\tau = -1.5$	$\tau = -1.2$	$\tau = -1$
100	0.8635	1.0348	1.1458	0.0226	<u>0.0013</u>	0.0300
500	0.9102	1.0350	1.1285	0.0135	<u>0.0012</u>	0.0210
1000	0.9280	1.0353	1.1288	0.0096	<u>0.0012</u>	0.0193
2000	0.9506	1.0287	1.1252	0.0066	<u>0.0009</u>	0.0173
5000	0.9715	1.0362	1.1309	0.0035	<u>0.0014</u>	0.0181
10000	0.9816	1.0417	1.1346	0.0019	<u>0.0018</u>	0.0187
25000	0.9904	1.0475	1.1383	<u>0.0009</u>	0.0023	0.0194
50000	0.9951	1.0510	1.1405	<u>0.0005</u>	0.0026	0.0199
Modelo <i>Burr</i> com $\gamma = 1$ e $\rho = -1$ ($\beta = 1$)						
	$\tau = -1.5$	$\tau = -1.2$	$\tau = -1$	$\tau = -1.5$	$\tau = -1.2$	$\tau = -1$
100	0.8707	1.0355	1.1689	0.0249	<u>0.0065</u>	0.0391
500	0.9083	1.0377	1.1482	0.0116	<u>0.0014</u>	0.0313
1000	0.9206	1.0380	1.1300	0.0088	<u>0.0014</u>	0.0265
2000	0.9111	1.0033	1.0985	0.0101	<u>0.0001</u>	0.0158
5000	0.9117	1.0020	1.0712	0.0096	<u>0.0001</u>	0.0089
10000	0.9153	1.0017	1.0574	0.0087	<u>0.0001</u>	0.0063
25000	0.9197	1.0014	1.0405	0.0076	<u>0.0001</u>	0.0038
50000	0.9221	1.0009	1.0294	0.0069	<u>0.0000</u>	0.0024
Modelo <i>Burr</i> com $\gamma = 1$ e $\rho = -2$ ($\beta = 1$)						
	$\tau = -1.5$	$\tau = -1.2$	$\tau = -1$	$\tau = -1.5$	$\tau = -1.2$	$\tau = -1$
100	0.8415	0.9955	1.1132	<u>0.5151</u>	0.5424	0.5939
500	0.8630	0.9584	1.0598	0.0312	<u>0.0061</u>	0.0063
1000	0.9003	0.9704	1.0494	0.0203	<u>0.0034</u>	0.0035
2000	0.9199	0.9776	1.0407	0.0150	0.0024	<u>0.0023</u>
5000	0.9359	0.9843	1.0282	0.0110	0.0017	<u>0.0013</u>
10000	0.9469	0.9888	1.0193	0.0081	0.0011	<u>0.0007</u>
25000	0.9626	0.9936	1.0088	0.0049	0.0005	<u>0.0003</u>
50000	0.9724	0.9962	1.0019	0.0033	0.0003	<u>0.0001</u>

os valores positivos sejam admissíveis, a escolha parece recair sempre em valores negativos.

- No modelo *Fréchet* ($\rho = -1$) a região de valores onde existe estabilidade das trajectórias amostrais, $\widehat{\beta}_{\rho}^{(\tau)}(k)$, parece ser muito inferior à de outros modelos. Na figura respeitante a este modelo, só para valores de $k > 4500$ é que existe alguma estabilidade das trajectórias amostrais, o que em outros modelos começa antes.
- No modelo *Cauchy*, como o verdadeiro valor de β é elevado, o estimador $\widehat{\beta}_{\rho}^{(\tau)}(k)$ tem uma variância muito grande, já que esta é proporcional a $\beta^2 = \left(\frac{2}{3}\pi^2\right)^2$. É por esta razão que para amostras de pequena dimensão o erro médio quadrático no nível óptimo é elevado.
- De um modo geral, no nível óptimo obtiveram-se bons resultados para a estimação do valor médio e do erro médio quadrático.

Referências

- [1] Caeiro, F. (2001). *Generalizações de estimadores clássicos do índice de variação regular*. Tese de mestrado, D.E.I.O., Faculdade de Ciências da Universidade de Lisboa.
- [2] Geluk, J. e L. de Haan (1987). *Regular Variation, Extensions and Tauberian Theorems*. Tech. Report CWI Tract 40, Centre for Mathematics and Computer Science, Amsterdam, Netherlands.
- [3] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423-453.
- [4] Fraga Alves, M.I., Gomes, M. I., e de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* **60**:1, 193-213.
- [5] Gomes, M. I., e Martins, M. J. (2001). Generalizations of the Hill estimator - asymptotic versus finite sample behaviour. *J. Statist. Planning and Inference* **93**, 161-180.
- [6] Gomes, M. I., e Martins, M. J. (2002). "Asymptotically Unbiased" Estimators of the Tail Index Based on External Estimation of the Second Order Parameter. *Extremes* **5**:1, 5-31.
- [7] Gomes, M. I., de Haan, L. e L. Peng (2002). Semi-parametric estimation of the second order parameter - asymptotic and finite sample behavior. *Extremes* **5**(4), 387-414.
- [8] Hall, P. (1982). On some Simple Estimates of an Exponent of Regular Variation. *J. R. Statist. Soc.* **44**, no. 1, 37-42.
- [9] Hall, P. e Welsh A.H. (1985). Adaptative estimates of parameters of regular variation. *Ann. Statist.* **13**, 331-341.
- [10] Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *Ann. Statist.* **3**, no. 5, 1163-1174.