ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 5, No.2
June 2007

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística = ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística (INE)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: +351 218 426 100
    Fax: +351 218 426 364
    Web site: http://www.ine.pt
    Customer Support Service
    (National network): 808 201 808
    (Other networks): +351 226 050 748
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística*
- **EDITION**
  - *4 000 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT 5% included]

- Single issue .... € 10
- Annual subscription (No. 1 Special Issue, No. 2 and No.3) ... € 24
- Annual subscription (No. 2, No. 3) .... € 16

INDEX

REVSTAT – Statistical Journal
Volume 5, Number 2, June 2007, 137–161

ON THE EXCESS DISTRIBUTION OF SUMS OF RANDOM VARIABLES IN BIVARIATE EV MODELS

Authors: Michael Falk
– Institute of Mathematics, University of Würzburg, Germany
falk@mathematik.uni-wuerzburg.de

Received: July 2006 Accepted: January 2007

Abstract:

- Let (U, V) be a random vector following a bivariate extreme value distribution (EVD) with reverse exponential margins. It is known that the excess distribution $F_c(t) = P\big(U+V > c\,t \mid U+V > c\big)$ of $U+V$ converges to $F(t) = t^2$ as the threshold c increases if U, V are independent, and to $F(t) = t$, $t \in [0,1]$, elsewhere. We investigate the limit of the excess distribution of $aU + bV$ in case of an EVD with arbitrary margins and with arbitrary scale parameters $a, b > 0$. It turns out that the limiting excess df may have a different behavior. For Fréchet margins, independence of U, V does not affect the limit excess distribution, whereas for Gumbel and reverse Weibull margins it does. Unless for Gumbel margins, the limit excess distribution is independent of a, b.
Interpreting a, b as weights and U, V as risks, $aU + bV$ can be viewed as a (short) linear portfolio. The fact that the limiting excess distribution of $aU + bV$ does not depend on a, b, unless for Gumbel margins, implies that risk measures such as the expected shortfall $E\big(aU + bV \mid aU + bV < c\big)$ might fail for multivariate extreme value models.

Key-Words:

- *univariate extreme value distribution; multivariate extreme value distribution; sums of random variables; excess distribution; Pickands dependence function; linear portfolio; risk measure; expected shortfall.*

AMS Subject Classification:

- 60G70.

1. INTRODUCTION

Let (X, Y) be a random vector (rv), whose distribution function (df) is a bivariate extreme value df (EVD) G with reverse exponential margins, i.e., G is max-stable

$$G^n\Big(\frac{x}{n}, \frac{y}{n}\Big) = G(x,y) , \qquad x,y \le 0, \quad n \in \mathbb{N} ,$$

and satisfies

$$G(x,0) = G(0,x) = P(X \le x) = P(Y \le x) = \exp(x) , \qquad x \le 0 .$$

It is well-known that G can be represented as

$$G(x,y) = \exp\bigg((x+y)\, D\Big(\frac{x}{x+y}\Big)\bigg) , \qquad x,y \le 0 , \tag{1.1}$$

where $D\colon [0,1] \to [1/2, 1]$ is a *Pickands dependence function*; see, for example, Sections 4.3, 6.1, 6.2 in Falk *et al.* (2004). A Pickands dependence function is characterized by the two properties

$$D \text{ is convex} , \tag{1.2}$$

$$\max(z, 1-z) \le D(z) \le 1 , \qquad z \in [0,1] , \tag{1.3}$$

i.e., $G(x,y) = \exp\big((x+y)\,D(x/(x+y)\big)$, $x,y \le 0$, defines an EVD G with reverse exponential margins if, and only if the function $D\colon [0,1] \to [1/2,1]$ satisfies condition (1.2) and (1.3) (see Falk (2006)).

A popular example is, with $\lambda \in [1, \infty]$,

$$D(z) = \Big(z^\lambda + (1-z)^\lambda\Big)^{1/\lambda} , \qquad z \in [0,1] ,$$

which yields the Gumbel type B df $G(x,y) = \exp\big(-\big(|x|^\lambda + |y|^\lambda\big)^{1/\lambda}\big)$, $x, y \le 0$, with the convention $D(z) = \max(z, 1-z)$ if $\lambda = \infty$.

Note that the case of independence of X, Y is in general characterized by the constant dependence function $D = 1$, in which case $G(x,y) = \exp(x+y)$, $x, y \le 0$. A major problem in the statistical analysis of given data $(x_1, y_1), ...,$ (x_n, y_n), is the decision whether the data were generated by rvs (X_i, Y_i) with independent margins X_i, Y_i, see, for example, Dupuis and Tawn (2001).

It was observed in Falk and Michel (2006) that the sum $X + Y$ over a high threshold has excellent ability to discriminate between independence and dependence, i.e., between the case of the constant dependence function $D = 1$

and a nonconstant D. Precisely, it was observed in Falk and Michel (2006) that for $t \in [0,1]$

$$(1.4)\qquad P\Big(X+Y > ct \mid X+Y > c\Big) \underset{c\uparrow 0}{\longrightarrow} \begin{cases} t^2 & \text{if } D=1\,, \\ t & \text{elsewhere}\,. \end{cases}$$

The excess distribution of the sum $X+Y$ over a high threshold approaches, consequently, either the df $F(t) = t^2$, $t \in [0,1]$, in case of independence of X, Y, or, elsewhere, the uniform distribution on $[0,1]$.

This observation was used in Falk and Michel (2006) to define a test for independence of X, Y, which is derived from the Neyman–Pearson test for the binary testing problem $F(t) = t^2$ against $F(t) = t$, $t \in [0,1]$, based on n independent copies of (X, Y). It was shown that this test has excellent performance and is able to detect deviations from the constant dependence function $D = 1$ which are of order $O(n^{-1/2})$.

The problem suggests itself, whether the characterization of independence and dependence of X, Y via the limiting excess distribution in (1.4) remains valid, if the rv (X, Y) with EVD G with reverse exponential margins is replaced by a rv (U, V), which follows an *arbitrary* EVD. This will be investigated in the present paper, where our investigations include arbitrary scale parameters $a, b > 0$ as well, i.e., we consider the excess distribution of $aU + bV$ over a high threshold with underlying arbitrary EVD. It turns out that the limit df of the excess distribution of the sum depends heavily on the marginal dfs: In some cases independence of U and V affects the limit, in other cases it does not. The main results can be summarized as follows, where it is generally assumed that the joint df of (U, V) is a bivariate EVD.

Reverse Weibull Margins: Suppose that U, V both follow a reverse Weibull df: $P(U \le x) = \exp\big(-(-x)^{\alpha_1}\big)$, $P(V \le x) = \exp\big(-(-x)^{\alpha_2}\big)$, $x \le 0$, $\alpha_1, \alpha_2 > 0$. Then we obtain for $a, b > 0$ and $t \in [0,1]$ (see Theorem 3.1)

$$(1.5)\qquad P\Big(aU + bV > tc \mid aU + bV > c\Big) \underset{c\uparrow 0}{\longrightarrow}$$

$$\underset{c\uparrow 0}{\longrightarrow} \begin{cases} t^{\alpha_1+\alpha_2} & \text{if } U, V \text{ are independent}\,, \\ t^{\max(\alpha_1,\alpha_2)} & \text{elsewhere}\,. \end{cases}$$

The special case $\alpha_1 = \alpha_2 = a = b = 1$ was established in Falk and Michel (2006). The limit excess df of $aU + bV$ is, therefore, determined by independence or dependence of U, V, but it is not affected by the scale parameters $a, b > 0$.

Fréchet Margins: Suppose that U, V both follow a Fréchet df: $P(U \le x) = \exp(-x^{-\alpha_1})$, $P(V \le x) = \exp(-x^{-\alpha_2})$, $x > 0$, $\alpha_1, \alpha_2 > 0$. Then we have for $a, b > 0$ and $t \ge 1$

$$(1.6)\qquad P\Big(aU + bV > tc \mid aU + bV > c\Big) \underset{c\to\infty}{\longrightarrow} t^{-\min(\alpha_1,\alpha_2)}\,.$$

In the case $\alpha_1 = \alpha_2$ and dependence of U, V, the preceding result requires an additional weak condition on the underlying Pickands dependence function, see Theorem 3.3 and 3.2 for details.

In case of Fréchet margins, the limiting excess df of $aU + bV$ is, consequently, invariant under dependence and independence of U, V and it is not affected by the choice of the scale parameters $a, b > 0$.

Gumbel Margins: If U, V both follow the Gumbel df $F(x) = \exp(-e^{-x})$, $x \in \mathbb{R}$, then we obtain for $a, b > 0$ and $t > 0$

$$(1.7) \qquad P\Big(aU + bV > c + t \mid aU + bV > c\Big) \underset{c\to\infty}{\longrightarrow}$$

$$\underset{c\to\infty}{\longrightarrow} \begin{cases} \exp\Big(-\dfrac{t}{\max(a,b)}\Big) & \text{if } U, V \text{ are independent}\,, \\ \exp\Big(-\dfrac{t}{a+b}\Big) & \text{elsewhere}\,, \end{cases}$$

see Theorem 3.4. In case of Gumbel margins, dependence and independence of U, V determine, consequently, the limiting excess df of $aU + bV$. But different to the other two cases above, it depends on the scale factors $a, b > 0$ as well.

The cases of mixed margins is determined by that df among the two dfs involved, which has a heavier tail, see Theorem 3.5, 3.6 and 3.7. Note that additional location parameters of U and V can simply be incorporated in the preceding results by shifting them to the threshold.

The transformation of the univariate margins of a multivariate EVD to arbitrary univariate extreme value distributions yields again a multivariate EVD. A common approach in multivariate extreme value theory is, therefore, the transformation of a given EVD to an EVD with one's favorite univariate margins. This approach might, however, be misleading as the preceding results reveal that the marginal distributions of a multivariate EVD, actually, can matter.

Extreme value theory has become a standard toolkit within quantitative finance useful for describing non normal phenomena, see, e.g., Embrechts (2000, 2004), Klüppelberg (2004), Section 13 in Reiss and Thomas (2001). The above results now reveal surprising facts in particular about the expected shortfall, which is a popular risk measure of a linear portfolio. Interpreting a, b as *weights* and U, V as *risks*, the sum $aU + bV$ can be viewed as a (short) *linear portfolio*. Note that the limit excess df of $aU + bV$ *above* a high threshold can in case of reverse Weibull margins readily be turned into the limiting excess df of a linear portfolio *below* a small threshold approaching zero: A rv (U, V) follows a bivariate max-stable df with reverse (standard) Weibull margins if, and only if, the rv $(\tilde{U}, \tilde{V}) := (-U, -V)$ follows a bivariate min-stable df with Weibull margins.

The standard exponential df on $(0,\infty)$ is a particular example. The limit result (1.5) now becomes with arbitrary $a, b > 0$ and $t \in [0,1]$

$$P\Big(a\tilde{U} + b\tilde{V} < tc \mid a\tilde{U} + b\tilde{V} < c\Big) \underset{c\downarrow 0}{\longrightarrow}$$

$$\underset{c\downarrow 0}{\longrightarrow} \begin{cases} t^{\alpha_1+\alpha_2} & \text{if } U, V \text{ are independent}\,, \\ t^{\max(\alpha_1,\alpha_2)} & \text{elsewhere}\,. \end{cases}$$

We see that in various cases, as the threshold increases or decreases, the limit excess distribution of $aU + bV$ does not depend on the parameters $a, b > 0$. A risk measure of a portfolio such as the *expected shortfall* (Acerbi and Sirtori (2001), Acerbi and Tasche (2001), Acerbi and Tasche (2002)), i.e., the expectation of $aU + bV$ given that the sum exceeds a high or a low threshold, is in this case asymptotically independent of the weights a, b. Such a risk measure of a linear portfolio has, consequently, to be taken with care, if the underlying joint df of the risks is assumed to be a max-stable or a min-stable df. For a linear portfolio $\sum_{i\le d} a_i U_i$ of arbitrary length d this was already observed in Macke (2005) in the case where $(U_1, ..., U_d)$ follows a d-dimensional EVD G with reverse exponential margins.

We remark that corresponding results might be established in higher dimensions as well, see, for a special case, Macke (2005). But the case of a dimension higher than two requires additional conditions such as very smooth dependence functions; it does not, however, provide essential new insight into the limit behavior of the corresponding excess distributions. In the two-dimensional case our mathematical tools are, on the other hand, so refined that we can establish our results under most general conditions. That is why we restrict ourselves in this paper to sums $aU + bV$ of length two.

It would, of course, be desirable to extend the preceding results (1.5)–(1.7) to rvs (U, V), whose distribution lies *in the domain of attraction* of a multivariate EVD. But this is not possible without further assumption. Take, for example, a rv (U, V), which follows a bivariate normal distribution $N(\mathbf{0}, \mathbf{\Sigma})$ with mean vector $\mathbf{0}$, variances 1 and and covariance $\rho \in (-1, 1)$. Then $N(\mathbf{0}, \mathbf{\Sigma})$ is in the domain of attraction of the EVD $G(x, y) = \exp(-e^{-x} - e^{-y})$, $x, y \in \mathbb{R}$, with independent Gumbel margins, i.e., there exist constants $a_n, c_n > 0$, $b_n, d_n \in \mathbb{R}$ such that

$$P\Big(\max_{1\le i\le n} U_i \le b_n + a_n x,\ \max_{1\le i\le n} V_i \le d_n + c_n y\Big) \underset{n\to\infty}{\longrightarrow} G(x, y)\,, \qquad x, y \in \mathbb{R}\,,$$

where $(U_1, V_1), (U_2, V_2), ...$ are independent copies of (U, V), see, e.g., equation (9.7) in Reiss and Thomas (2001). According to equation (1.7) one, therefore, should expect in this case that the limit of $P\big(aU + bV > c + t \mid aU + bV > c\big)$ is $\exp(-t/\max(a, b))$ as c converges to infinity. Standard arguments, however, yield that

$$P\Big(aU + bV > c + t \mid aU + bV > c\Big) \underset{c\to\infty}{\longrightarrow} 0$$

for arbitrary $a, b, t > 0$.

The results in this paper are related to results by Wüthrich (2003), Alink *et al.* (2004, 2005a, 2005b) and Barbe *et al.* (2006), who establish $P\big(\sum_{i\le d} X_i > t\big) \sim \Delta P(X_1 > t)$ as $t\to\infty$ with some diversification constant $\Delta>0$. This is achieved under various conditions on the joint distribution of $(X_1,...,X_d)$, thus extending the well known result with $\Delta = d$ in case of iid regularly varying X_i (Feller (1971, p. 279)) to dependent rvs. The above authors work, however, with identically distributed X_i so that the results stated here are not included in these papers.

This paper is organized as follows. As the derivation of our results is highly technical, we compile in Section 2 in a preparatory step various auxiliary results and tools. The main results are established in Section 3.

2. AUXILIARY RESULTS AND TOOLS

In a preparatory step we provide in this section several auxiliary results and mathematical tools, which might be of interest of their own.

A bivariate and nondegenerate EVD H has the characteristic property of max-stability, i.e., for each $n\in\mathbb{N}$ there are constants $a_{in}>0$, $b_{i,n}\in\mathbb{R}$, $i=1,2$, such that

$$H^n\big(a_{1n}x+b_{1n},\, a_{2n}y+b_{2n}\big) \;=\; H(x,y)\;, \qquad x,y\in\mathbb{R}\;.$$

The margins of H are, consequently, univariate EVDs. The family of nondegenerate univariate EVDs is, with $\alpha>0$, up to a scale and location shift given by

$$\begin{aligned}
F_\alpha(x) \;&:=\; \begin{cases}\exp\big(-(-x)^\alpha\big), & x\le 0\,,\\ 1, & x>0\;,\end{cases}\\
F_{-\alpha}(x) \;&:=\; \begin{cases}0, & x\le 0\,,\\ \exp(-x^{-\alpha}), & x>0\;,\end{cases} \qquad\qquad (2.1)\\
F_0(x) \;&:=\; \exp(-e^{-x})\,, \qquad x\in\mathbb{R}\;,
\end{aligned}$$

being the family of (reverse) Weibull, Fréchet dfs and the Gumbel df; see, e.g., Section 2.2 in Falk *et al.* (2004). Note that F_1 is the standard reverse exponential df.

Let now (U,V) be a rv, which follows a bivariate EVD H with standard univariate extreme value margins as in (2.1). It is well-known that the df H of (U,V) equals that of $\big(H_1^{-1}(\exp(X)), H_2^{-1}(\exp(Y))\big)$, where (X,Y) follows an EVD G with reverse exponential margins F_1. By $F^{-1}(q) := \inf\{t\in\mathbb{R}\colon F(t)\ge q\}$, $q\in(0,1)$, we denote the generalized inverse of a univariate df F; see, for example, Lemma 5.4.7 in Falk *et al.* (2004). In different notation we have, thus,

$$H(x,y) = G\Big(\log\big(H_1(x)\big), \log\big(H_2(y)\big)\Big) = G\big(\psi_1(x), \psi_2(y)\big) ,$$

where $\psi_i(x) = \log(H_i(x))$, $i = 1,2$, is each one of the three functions defined as follows:

$$\psi(x) := \begin{cases} -(-x)^\alpha, & x \le 0 , \\ -x^{-\alpha}, & x > 0 , \\ -e^{-x}, & x \in \mathbb{R} . \end{cases}$$

We have, consequently,

$$(U,V) =_D \big(\psi_1^{-1}(X), \psi_2^{-1}(Y)\big) , \tag{2.2}$$

where $=_D$ denotes equality in distribution, and, we have by equation (1.1) for x,y with $0 < H_1(x), H_2(x) < 1$

$$H(x,y) = \exp\left(\Big(\psi_1(x) + \psi_2(y)\Big)\, D\left(\frac{\psi_1(x)}{\psi_1(x) + \psi_2(y)}\right)\right) , \tag{2.3}$$

where D is a Pickands dependence function as defined by (1.2) and (1.3).

Note that $\big(\psi_1^{-1}(X), \psi_2^{-1}(Y)\big)$ follows for an arbitrary choice of an EVD G with reverse exponential margins an EVD H with margins H_1, H_2 and, thus, representation (2.3) characterizes up to a scale and location parameter the complete class of bivariate EVDs with arbitrary margins.

The following auxiliary result provides a representation of an arbitrary Pickands dependence function D, which will be crucial for the derivation of our subsequent results. It implies in particular that any D is absolutely continuous and provides its derivative D'. For a proof of this result we refer to Lemma 6.2.1 in Falk *et al.* (2004).

Lemma 2.1. *An arbitrary Pickands dependence function D can be represented as*

$$D(z) = 1 + \int_0^z M(x) - 1\, dx = 1 - \int_z^1 M(x) - 1\, dx ,$$

where $M\colon [0,1] \to [0,2]$ is a measure generating function with $M(1) = 2$, $\int_0^1 M(x)\, dx = 1$. The dependence function D is, consequently, absolutely continuous with derivative

$$D'(z) := M(z) - 1 \in [-1,1] .$$

It is easy to see that the converse of the preceding result is also true: any function $D\colon [0,1] \to [0,\infty)$ that can be represented as $D(z) = 1 + \int_0^z M(x) - 1\, dx$, with $M\colon [0,1] \to [0,2]$ as in Lemma 2.1, satisfies condition (1.2) and (1.3) and is, consequently, a Pickands dependence function.

We will make extensive use of the conditional df $P(Y \le v \mid X = u)$, where (X, Y) follows a bivariate EVD with reverse exponential margins. This conditional df is provided in the next lemma. For a proof we refer to Lemma 2.1 in Falk and Michel (2006); the arguments are taken from Ghoudi *et al.* (1998).

Lemma 2.2. *Suppose that the rv (X, Y) follows an EVD G with reverse exponential margins and Pickands dependence function D. Then we have for $u < 0$*

$$P\Big(Y \le v \mid X = u\Big) =$$
$$= \begin{cases} \exp\Big\{u\Big(D\big(\tfrac{u}{u+v}\big) - 1\Big) + v\,D\big(\tfrac{u}{u+v}\big)\Big\}\,\Big(D\big(\tfrac{u}{u+v}\big) + D'\big(\tfrac{u}{u+v}\big)\big(1 - \tfrac{u}{u+v}\big)\Big) & \text{if } v < 0\,, \\ 1 & \text{if } v \ge 0\,. \end{cases}$$

3. MAIN RESULTS

In this section we compute the limiting excess df of the sum $aU + bV$, where (U, V) follows an arbitrary bivariate EVD. Without loss of generality (wlog) we suppose that the marginal univariate dfs have scale parameter 1. We begin with the case of reverse Weibull margins.

Theorem 3.1 (Reverse Weibull Margins)**.** *Suppose that (U,V) follows a bivariate EVD with reverse Weibull margins: $P(U \le x) = \exp\big(-(-x)^{\alpha_1}\big)$, $P(V \le x) = \exp\big(-(-x)^{\alpha_2}\big)$, $x \le 0$, $\alpha_1, \alpha_2 > 0$. If U, V are not independent, then we have for $a, b > 0$ and $0 \le t \le 1$*

$$P\Big(aU + bV > tc \mid aU + bV > c\Big) \underset{c\uparrow 0}{\longrightarrow} t^{\max(\alpha_1,\alpha_2)}\,.$$

If U, V are independent, then we have

$$P\Big(aU + bV > tc \mid aU + bV > c\Big) \underset{c\uparrow 0}{\longrightarrow} t^{\alpha_1+\alpha_2}\,.$$

Proof: Wlog we assume $\alpha_1 \ge \alpha_2$. The assertion is an immediate consequence of

$$(3.1) \qquad (-c)^{-\alpha_1}\,P\big(aU + bV > c\big) \underset{c\uparrow 0}{\longrightarrow} K(a,b) > 0$$

if U, V are not independent, and of

$$(3.2) \qquad (-c)^{-(\alpha_1+\alpha_2)}\,P\big(aU + bV > c\big) \underset{c\uparrow 0}{\longrightarrow} a^{-\alpha_1}\,b^{-\alpha_2}\,\alpha_1 \int_0^1 (1-u)^{\alpha_2}\,u^{\alpha_1 - 1}\,du$$

if U, V are independent. This will be established in the following.

Wlog we can by (2.2) assume that $(U,V) = \big(-(-X)^{1/\alpha_1}, -(-Y)^{1/\alpha_2}\big)$, where (X,Y) follows a bivariate EVD $G(x,y) = \exp\big((x+y)\,D(x/(x+y))\big)$, $x,y \le 0$, with reverse exponential margins and Pickands dependence function D.

By conditioning on $X = u$, we obtain from Lemma 2.2 the representation

$$
\begin{aligned}
P\big(aU+bV>c\big) &= \\
&= \int_{-\infty}^{0} P\Big(-(-Y)^{1/\alpha_2} > \frac{c+a(-u)^{1/\alpha_1}}{b} \;\mid\; X=u\Big)\exp(u)\;du \\
&= \int_{-(\frac{-c}{a})^{\alpha_1}}^{0} \bigg(1 - P\bigg(Y \le -\Big(-\frac{c+a(-u)^{1/\alpha_1}}{b}\Big)^{\alpha_2} \;\mid\; X=u\bigg)\bigg)\exp(u)\;du \qquad (3.3)\\
&= 1 - \exp\Big(-\Big(\frac{-c}{a}\Big)^{\alpha_1}\Big) \qquad (3.4)\\
&\quad - \int_{-(\frac{-c}{a})^{\alpha_1}}^{0} \exp\bigg(u\big(D(\tilde u)-1\big) - \Big(-\frac{c+a(-u)^{1/\alpha_1}}{b}\Big)^{\alpha_2} D(\tilde u)\bigg) \\
&\qquad\qquad \times \Big(D(\tilde u) + D'(\tilde u)\,(1-\tilde u)\Big)\exp(u)\;du\;,
\end{aligned}
$$

where for $u \in \big(-(-c/a)^{\alpha_1}, 0\big]$

$$
\tilde u \;:=\; \frac{u}{u - \Big(-\frac{c+a(-u)^{1/\alpha_1}}{b}\Big)^{\alpha_2}} \;\in\; [0,1]\;.
$$

In case of independence, i.e., $D = 1$, we obtain from equation (3.3) by using Taylor expansion of exp at 0 and substituting $u \mapsto -(-cu/a)^{\alpha_1}$

$$
\begin{aligned}
P\big(aU+bV>c\big) &= \\
&= \int_{-(\frac{-c}{a})^{\alpha_1}}^{0} \bigg(1 - \exp\bigg(-\Big(-\frac{c+a(-u)^{1/\alpha_1}}{b}\Big)^{\alpha_2}\bigg)\bigg)\exp(u)\;du \\
&= -\Big(\frac{-c}{a}\Big)^{\alpha_1} \int_0^1 \bigg(1 - \exp\Big(-\Big(\frac{-c}{b}\Big)^{\alpha_2}(1-u)^{\alpha_2}\Big)\bigg) \\
&\qquad\qquad \times \exp\Big(-\Big(\frac{-c}{a}\Big)^{\alpha_1} u^{\alpha_1}\Big)\,\alpha_1\,u^{\alpha_1-1}\;du \\
&= \frac{(-c)^{\alpha_1+\alpha_2}}{a^{\alpha_1}\,b^{\alpha_2}}\;\alpha_1 \int_0^1 (1-u)^{\alpha_2}\,u^{\alpha_1-1}\big(1+o(1)\big)\;du\;,
\end{aligned}
$$

which implies equation (3.2).

It remains to establish equation (3.1). From equation (3.4) we obtain with the substitution $u \mapsto -(-cu/a)^{\alpha_1}$

$$
\begin{aligned}
&P\big(aU+bV>c\big) \;= \\
&\quad = 1-\exp\Big(-\Big(\frac{-c}{a}\Big)^{\alpha_1}\Big) \\
&\quad -\Big(\frac{-c}{a}\Big)^{\alpha_1}\int_0^1 \exp\Big(-\Big(\frac{-c}{a}\Big)^{\alpha_1} u_1^{\alpha}\big(D(u_c)-1\big)-\Big(\frac{-c}{b}\Big)^{\alpha_2}(1-u)^{\alpha_2}\,D(u_c)\Big) \\
&\qquad\qquad \times\Big(D(u_c)+D'(u_c)\,(1-u_c)\Big)\,\exp\Big(-\Big(\frac{-c}{a}\Big)^{\alpha_1}u^{\alpha_1}\Big)\,\alpha_1\,u^{\alpha_1-1}\,du\;,
\end{aligned}
\tag{3.5}
$$

where for $u\in(0,1)$

$$
u_c \;:=\; \frac{u^{\alpha_1}}{u^{\alpha_1}+(-c)^{\alpha_2-\alpha_1}\,\frac{a^{\alpha_1}}{b^{\alpha_2}}\,(1-u)^{\alpha_2}} \;\underset{c\uparrow 0}{\downarrow}\; 0 \qquad \text{if } \alpha_1>\alpha_2\;.
$$

Hence we obtain in the case $\alpha_1>\alpha_2$

$$
\Big(\frac{-c}{a}\Big)^{-\alpha_1} P\big(aU+bV>c\big) \;\underset{c\uparrow 0}{\longrightarrow}\; -D'(0) \;=\; 1-M(0)\;>\;0\;.
$$

The fact that $M(0)<1$ can be seen as follows: Suppose that $M(0)\ge 1$. Then we obtain from Lemma 2.1 that $D(z)=1+\int_0^z M(x)-1\,dx\ge 1$, $0\le z\le 1$, and, thus, D is the constant function 1. But this case was excluded. Thus we have established equation (3.1) in the case $\alpha_1>\alpha_2$. It remains to prove (3.1) also in the case $\alpha_1=\alpha_2$.

Suppose that $\alpha_1=\alpha_2$. Equation (3.5) implies

$$
\begin{aligned}
&\Big(\frac{-c}{a}\Big)^{-\alpha_1} P\big(aU+bV>c\big) \;\underset{c\uparrow 0}{\longrightarrow} \\
&\qquad \underset{c\uparrow 0}{\longrightarrow}\; \int_0^1 \Big(1-D(u^*)-D'(u^*)\,(1-u^*)\Big)\,\alpha_1\,u^{\alpha_1-1}\,du \;>\; 0
\end{aligned}
\tag{3.6}
$$

where for $u\in[0,1]$

$$
u^* \;:=\; \frac{u^{\alpha_1}}{u^{\alpha_1}+(1-u)^{\alpha_1}\,(\frac{a}{b})^{\alpha_1}} \;\in\; [0,1]\;.
$$

We show in the following that the limit integral in (3.6) is strictly positive. Note that we have by Lemma 2.1 for $u\in[0,1]$

$$
1-D(u)-D'(u)\,(1-u) \;=\; \int_u^1 M(x)-M(u)\;dx \;\ge\; 0\;,
$$

where the integral on the right hand side above is a function in u, which is continuous from the right. Suppose that the integral in equation (3.6) is zero. This implies $\int_u^1 M(x)-M(u)\,dx=0$ for $u\in[0,1)$. Then we have in particular $\int_0^1 M(x)-M(0)\,dx=0$, which implies $M(x)=M(0)$, $x\in[0,1)$, and, thus, $D(z)=1+\int_0^z M(0)-1\,dx=1+z\big(M(0)-1\big)$, $z\in[0,1]$. From the fact that $D(1)=1$ we obtain $M(0)=1$ and, hence, that $D(z)=1$, $z\in[0,1]$. But this case was excluded. The limit integral in (3.6) is, therefore, strictly positive. This completes the proof of equation (3.1) and, thus, the proof of Theorem 3.1. □

The case of Fréchet margins requires completely different proofs for identical and nonidentical margins. The two cases are, therefore, stated separately in Theorem 3.3 and in Theorem 3.2. We begin with the case of different margins, since this case is an immediate consequence of the following result for regularly varying rvs. For a proof of this result we refer to Lemma 2 in Klüppelberg *et al.* (2006) [17].

Lemma 3.1. *Let Y and Z be rvs on a common probability space such that Y has regularly varying right tail with index $-\kappa < 0$. Let $d > \kappa$ and suppose that $E(|Z|^d) < \infty$. Then*

$$\lim_{x\to\infty} \frac{P\big(Y+Z > x\big)}{P\big(Y > x\big)} = 1 .$$

Theorem 3.2 (Different Fréchet Margins)**.** *Suppose that (U, V) follows a bivariate EVD with different standard Fréchet margins: $P(U \le x) = \exp(-x^{-\alpha_1})$, $P(V \le x) = \exp(-x^{-\alpha_2})$, $x > 0$, $\alpha_1 \ne \alpha_2$. Then we have for $a, b > 0$ and $t \ge 1$*

$$P\Big(aU + bV > ct \mid aU + bV > c\Big) \underset{c\to\infty}{\longrightarrow} t^{-\min(\alpha_1,\alpha_2)} .$$

Note that the case of identical Fréchet margins $\alpha_1 = \alpha_2 =: \alpha$ is not covered by Lemma 3.1, as in this case $E(|U|^d) = E(|V|^d) = \infty$ for any $d > \alpha$.

Theorem 3.3 (Identical Fréchet Margins)**.** *Suppose that (U, V) follows a bivariate EVD with identical Fréchet margins: $P(U \le x) = P(V \le x) = \exp(-x^{-\alpha})$, $x > 0$, for some $\alpha > 0$. Then we obtain for $a, b > 0$ and $t \ge 1$*

$$P\Big(aU + bV > ct \mid aU + bV > c\Big) \underset{c\to\infty}{\longrightarrow} t^{-\alpha} \tag{3.7}$$

if U, V are independent. If U, V are not independent, this result remains true if we require in addition that the underlying dependence function D satisfies for some $\delta > 1$ the expansion

$$1 - D(z) - D'(z)\,(1 - z) = O\big((1 - z)^\delta\big) . \tag{3.8}$$

Condition (3.8) is, for example, satisfied by the dependence function $D(z) = \big(z^\lambda + (1 - z)^\lambda\big)^{1/\lambda}$, $1 \le \lambda \le \infty$, which corresponds to the Gumbel type B EVD. It is also obviously satisfied by the dependence function $D(z) = 1-\lambda\min(z, 1-z)$, $\lambda \in [0, 1]$, which corresponds to the Marshall–Olkin EVD. We conjecture that it is satisfied by an arbitrary dependence function, but this is an open question.

Proof: Wlog we can assume $(U,V) = \big((-X)^{-1/\alpha}, (-Y)^{-1/\alpha}\big)$, where (X,Y) follows a bivariate EVD G with reverse exponential margins and dependence function D.

First we consider the case $D(z)=1$, $z\in[0,1]$, i.e., the case of independence of X,Y or, equivalently, of U,V. We claim that in this case

$$c^\alpha P\big(aU+bV>c\big) \underset{c\to\infty}{\longrightarrow} a^\alpha+b^\alpha\ , \tag{3.9}$$

from which equation (3.7) follows immediately. Equation (3.9) can be seen as follows. Note that $P\big(aU+bV>c \mid X=u\big)=1$ if $u>-(a/c)^\alpha$ and, thus,

$$\begin{aligned}
P\big(aU+bV>c\big) \;&=\\
&= \int_{-\infty}^{0} P\Big(aU+bV>c \mid X=u\Big)\exp(u)\,du\\
&= \int_{-\left(\frac{a}{c}\right)^\alpha}^{0} \exp(u)\,du + \int_{-\infty}^{-\left(\frac{a}{c}\right)^\alpha} P\left(Y>-\left(\frac{b}{c-a(-u)^{-1/\alpha}}\right)^\alpha \mid X=u\right)\exp(u)\,du\\
&= 1-\exp\left(-\left(\frac{a}{c}\right)^\alpha\right) + \int_{-\infty}^{-\left(\frac{a}{c}\right)^\alpha}\left(1-\exp\left(-\left(\frac{b}{c-a(-u)^{-1/\alpha}}\right)^\alpha\right)\right)\exp(u)\,du\ .
\end{aligned}$$

Since $1-\exp\big(-(a/c)^\alpha\big)=(a/c)^\alpha\big(1+o(1)\big)$, it suffices to show that the integral on the right hand side above equals $(b/c)^\alpha\big(1+o(1)\big)$. Split the integral into the sum of the subintegrals $\int_{-\infty}^{-2(a/c)^\alpha}\cdots\,du+\int_{-2(a/c)^\alpha}^{-(a/c)^\alpha}\cdots\,du$. By the substitution $u\mapsto-(a/c)^\alpha u$ the second subintegral equals

$$\left(\frac{a}{c}\right)^\alpha\int_1^2\left(1-\exp\left(-\left(\frac{b}{c}\right)^\alpha\big(1-u^{-1/\alpha}\big)^{-\alpha}\right)\right)\exp\left(-\left(\frac{a}{c}\right)^\alpha u\right)du \;=\; o(c^{-\alpha})$$

by the dominated convergence theorem. From the Taylor expansion $\exp(-x)=1-x+\exp(-\vartheta_x\,x)\,x^2/2$ with $0<\vartheta_x<1$ and the fact that $0<\exp(-\vartheta_x\,x)<1$ for $x>0$ we obtain that the first subintegral equals

$$\int_{-\infty}^{-2\left(\frac{a}{c}\right)^\alpha}\left(\frac{b}{c}\right)^\alpha\left(1-\frac{a}{c}(-u)^{-1/\alpha}\right)^{-\alpha}\exp(u)\,du+O(c^{-2\alpha}) \;=\; \left(\frac{b}{c}\right)^\alpha\big(1+o(1)\big)\ .$$

Thus we have shown (3.9).

If D is not the constant function 1, we have

$$\begin{aligned}
&c^\alpha P\big(aU+bV>c\big) \underset{c\to\infty}{\longrightarrow}\\
&\underset{c\to\infty}{\longrightarrow}\; b\int_0^1\Big(1-D(z)-D'(z)(1-z)\Big)\,z^{1/\alpha-1}\,\frac{\Big(b\,z^{1/\alpha}+a(1-z)^{1/\alpha}\Big)^{\alpha-1}}{(1-z)^2}\,dz\\
&\quad +b^\alpha\Big(2-M(1-0)\Big) \;>\; 0\ ,
\end{aligned} \tag{3.10}$$

where M is the measure generating function in the representation $D(z)=1+\int_0^z M(x)-1\,dx$ and $M(1-0):=\lim_{\varepsilon\downarrow 0}M(1-\varepsilon)$ is the limit from the left of M at 1.

This is established in the following. Repeating previous arguments we obtain

$$
\begin{aligned}
P\big(aU+bV>c\big) &= \\
&= 1-\exp\Big(-\Big(\frac{a}{c}\Big)^{\alpha}\Big) \\
&\quad + \int_{-\infty}^{-\left(\frac{a}{c}\right)^{\alpha}} \Bigg(1-P\Big(Y\le -\Big(\frac{b}{c-a(-u)^{-1/\alpha}}\Big)^{\alpha} \mid X=u\Big)\Bigg)\exp(u)\,du\ .
\end{aligned}
$$

The integral equals, by Lemma 2.2,

$$
\begin{aligned}
&\int_{-\infty}^{-\left(\frac{a}{c}\right)^{\alpha}} \Bigg\{1-\exp\Bigg(\Bigg(u-\Big(\frac{b}{c-a(-u)^{-1/\alpha}}\Big)^{\alpha}\Bigg)D(\tilde u)\Bigg)\exp(-u) \\
&\qquad \times\Big(D(\tilde u)+D'(\tilde u)\,(1-\tilde u)\Big)\Bigg\}\exp(u)\,du \;= \\
&\quad = \int_{-\infty}^{-\left(\frac{a}{c}\right)^{\alpha}} \Big(1-D(\tilde u)-D'(\tilde u)\,(1-\tilde u)\Big)\exp(u)\,du \\
&\qquad + \int_{-\infty}^{-\left(\frac{a}{c}\right)^{\alpha}} \Bigg\{1-\exp\Bigg(u\big(D(\tilde u)-1\big)-\Big(\frac{b}{c-a(-u)^{-1/\alpha}}\Big)^{\alpha}D(\tilde u)\Bigg)\Bigg\} \\
&\qquad\qquad \times\Big(D(\tilde u)+D'(\tilde u)(1-\tilde u)\Big)\exp(u)\,du\ ,
\end{aligned}
\tag{3.11}
$$

where for $u<-(a/c)^{\alpha}$

$$
\tilde u := \frac{1}{1+\frac{b^{\alpha}}{\left(c(-u)^{1/\alpha}-a\right)^{\alpha}}} \;\in\; (0,1)
$$

converges to 1 as $c\to\infty$. Putting for $z\in(0,1)$

$$
g(z) := -\frac{1}{c^{\alpha}}\Bigg(b\Big(\frac{z}{1-z}\Big)^{1/\alpha}+a\Bigg)^{\alpha}.
$$

and substituting $u\mapsto g(z)$, we obtain that the first integral in equation (3.11) equals

$$
\begin{aligned}
&-\int_0^1 \Big(1-D(z)-D'(z)\,(1-z)\Big)\exp\big(g(z)\big)\,g'(z)\,dz \;= \\
&\qquad = \frac{b}{c^{\alpha}}\int_0^1 \Big(1-D(z)-D'(z)\,(1-z)\Big)\exp\big(g(z)\big) \\
&\qquad\qquad \times\Big(b\,z^{1/\alpha}+a(1-z)^{1/\alpha}\Big)^{\alpha-1}(1-z)^{-2}\,z^{1/\alpha-1}\,dz\ .
\end{aligned}
$$

From condition (3.8) and the dominated convergence theorem we, therefore, obtain

$$
\begin{aligned}
&c^{\alpha}\int_{-\infty}^{-\left(\frac{a}{c}\right)^{\alpha}} \Big(1-D(\tilde u)-D'(\tilde u)\,(1-\tilde u)\Big)\exp(u)\,du \;\underset{c\to\infty}{\longrightarrow} \\
&\qquad \underset{c\to\infty}{\longrightarrow}\; b\int_0^1\Big(1-D(z)-D'(z)\,(1-z)\Big)\,z^{1/\alpha-1} \\
&\qquad\qquad \times\Big(b\,z^{1/\alpha}+a(1-z)^{1/\alpha}\Big)^{\alpha-1}(1-z)^{-2}\,dz \;\in\; (0,\infty)\ .
\end{aligned}
\tag{3.12}
$$

The second integral in equation (3.11) is split into the sum of the subintervals

$$\int_{-\infty}^{-2\left(\frac{a}{c}\right)^\alpha} \cdots \, du + \int_{-2\left(\frac{a}{c}\right)^\alpha}^{-\left(\frac{a}{c}\right)^\alpha} \cdots \, du \ =: \ I(c) + II(c) \ .$$

Substituting $u \mapsto -(a/c)^\alpha u$ and putting for $u \in (1,2)$

$$\bar{u} := \frac{1}{1 + (\frac{b}{a})^\alpha \, (u^{1/\alpha} - 1)^{-\alpha}} \ \in \ (0,1) \ ,$$

the second subintegral above equals

$$\begin{aligned} II(c) \ &= \ \left(\frac{a}{c}\right)^\alpha \int_1^2 \left\{ 1 - \exp\left(-\left(\frac{a}{c}\right)^\alpha u \big(D(\bar{u}) - 1\big) - \left(\frac{b}{c}\right)^\alpha \big(1 - u^{-1/\alpha}\big)^{-\alpha} D(\bar{u}) \right) \right\} \\ &\qquad \times \Big(D(\bar{u}) + D'(\bar{u}) \, (1 - \bar{u}) \Big) \exp\left(-\left(\frac{a}{c}\right)^\alpha u \right) \, du \\ &= \ o\big(c^{-\alpha}\big) \end{aligned}$$

by the dominated convergence theorem.

Taylor expansion of exp at zero yields that the first subintegral equals

$$\begin{aligned} I(c) \ &= \ \int_{-\infty}^{-2\left(\frac{a}{c}\right)^\alpha} \left(\big(1 - D(\tilde{u})\big) \, u + \left(\frac{b}{c - a(-u)^{-1/\alpha}} \right)^\alpha D(\tilde{u}) \right) \\ &\qquad \times \ \exp\left\{ \vartheta_u \, u \big(D(\tilde{u}) - 1\big) - \vartheta_u \left(\frac{b}{c - a(-u)^{-1/\alpha}} \right)^\alpha D(\tilde{u}) \right\} \\ &\qquad \times \ \Big(D(\tilde{u}) + D'(\tilde{u}) \, (1 - \tilde{u}) \Big) \exp(u) \ du \ , \end{aligned}$$

where $0 < \vartheta_u < 1$. Recall that $1 - D(\tilde{u}) \in [0, 1/2]$, $D'(\tilde{u})(1 - \tilde{u}) \in [-1, 1]$ and note that for $u \le -2(a/c)^\alpha$

$$\left(\frac{b}{c - a(-u)^{-1/\alpha}} \right)^\alpha \ = \ \left(\frac{b}{c}\right)^\alpha \left(\frac{1}{1 - \frac{1}{\left(-\left(\frac{c}{a}\right)^\alpha u\right)^{1/\alpha}}} \right)^\alpha \ \le \ \frac{1}{(1 - 2^{-1/\alpha})^\alpha} \left(\frac{b}{c}\right)^\alpha .$$

We have, further, by Lemma 2.1

$$1 - D(\tilde{u}) \ = \ \int_{\tilde{u}}^1 M(x) - 1 \, dx \ = \ \big(M(1-0) - 1\big) \, (1 - \tilde{u}) \, \big(1 + r(\tilde{u})\big) \ ,$$

where

$$0 \;\geq\; r(\tilde u) \;:=\; \frac{\displaystyle\int_{\tilde u}^1 M(x)-1\,dx}{\big(M(1-0)-1\big)\,(1-\tilde u)} - 1$$

$$= \frac{\displaystyle\int_{\tilde u}^1 M(x)-M(1-0)\,dx}{\big(M(1-0)-1\big)\,(1-\tilde u)}$$

$$\geq \frac{M(\tilde u)-M(1-0)}{M(1-0)-1}$$

$$\geq -\frac{M(1-0)-M(0)}{M(1-0)-1}$$

is bounded and converges to 0 as $c\to\infty$. We have, further, for $u \leq -2(a/c)^\alpha$

$$(1-\tilde u)\,u \;=\; \frac{\frac{b^\alpha}{\left(c(-u)^{1/\alpha}-a\right)^\alpha}}{1+\frac{b^\alpha}{\left(c(-u)^{1/\alpha}-a\right)^\alpha}}\;u$$

$$= \frac{b^\alpha}{c^\alpha}\;\frac{u}{\left((-u)^{1/\alpha}-\frac{a}{c}\right)^\alpha}\;\frac{1}{1+\frac{b^\alpha}{\left(c(-u)^{1/\alpha}-a\right)^\alpha}}$$

$$= -\frac{b^\alpha}{c^\alpha}\big(1+s_c(u)\big)\;,$$

where s_c is bounded and $s_c(u) \underset{c\to\infty}{\longrightarrow} 0$.

We obtain, consequently, from the dominated convergence theorem

$$c^\alpha I(c) \;\underset{c\to\infty}{\longrightarrow}\; \int_{-\infty}^0 b^\alpha\big(2-M(1-0)\big)\exp(u)\,du \;=\; \big(2-M(1-0)\big)\,b^\alpha \;\geq\; 0\,. \tag{3.13}$$

Equation (3.10) now follows from (3.11), (3.12) and (3.13). □

Theorem 3.4 (Gumbel Margins)**.** *Suppose that the rv* (U,V) *follows a bivariate EVD with identical Gumbel margins:* $P(U\leq x) = P(V\leq x) = \exp(-e^{-x})$, $x\in\mathbb{R}$. *Then we obtain for* $a,b>0$ *and* $t\geq 0$

$$P\Big(aU+bV>c+t \;\big|\; aU+bV>c\Big) \;\underset{c\to\infty}{\longrightarrow}$$

$$\underset{c\to\infty}{\longrightarrow}\; \begin{cases} \exp\!\Big(-\dfrac{t}{\max(a,b)}\Big) & \text{if } U,V \text{ are independent}\,,\\[2ex] \exp\!\Big(-\dfrac{t}{a+b}\Big) & \text{elsewhere}\,.\end{cases}$$

Proof: We consider first the case, where U, V are independent. Wlog we assume $a > b$. The case $a = b$ requires a different approach, see below. The assertion is immediate from

$$e^{\frac{c}{a}}\,P\big(aU+bV>c\big) \underset{c\to\infty}{\longrightarrow} \frac{b}{a}\int_{-\infty}^{\infty}\Big(1-\exp(-e^{-u})\Big)\,e^{\frac{b}{a}u}\,du \;\in\; (0,\infty)\;, \tag{3.14}$$

which we establish in the sequel.

Put $F(u) := \exp(-e^{-u})$, $u \in \mathbb{R}$. We have

$$\begin{aligned} P\big(aU+bV>c\big) &= \int_{-\infty}^{\infty} P\Big(aU+bV>c \mid U=u\Big)\,F'(u)\,du \\ &= \int_{-\infty}^{\infty}\bigg(1-F\Big(\frac{c-au}{b}\Big)\bigg)\,F'(u)\,du\;. \end{aligned}$$

With the substitution $u \mapsto (c-bu)/a$, the preceding integral equals

$$\begin{aligned} &\frac{b}{a}\int_{-\infty}^{\infty}\Big(1-F(u)\Big)\,F'\Big(\frac{c-bu}{a}\Big)\,du \;= \\ &\qquad= \frac{b}{a}\int_{-\infty}^{\infty}\Big(1-\exp(-e^{-u})\Big)\,e^{\frac{bu-c}{a}}\,\exp\big(-e^{\frac{bu-c}{a}}\big)\,du \\ &\qquad= e^{-\frac{c}{a}}\,\frac{b}{a}\int_{-\infty}^{\infty}\Big(1-\exp(-e^{-u})\Big)\,e^{\frac{b}{a}u}\,\exp\big(-e^{\frac{bu-c}{a}}\big)\,du \end{aligned}$$

and, thus, equation (3.14) follows from the dominated convergence theorem; recall that $a > b$.

Next we consider the case, where U, V are independent and $a = b$. The assertion is a consequence of

$$\frac{e^{c/a}}{c/a}\,P\big(aU+aV>c\big) \underset{c\to\infty}{\longrightarrow} 1\;, \tag{3.15}$$

which we establish in the following. Wlog we assume $a = 1$. Repeating the arguments in the derivation of (3.14) we obtain

$$\begin{aligned} P\big(aU+bV>c\big) &= e^{-c}\bigg\{\int_{-\infty}^{0}\Big(1-\exp(-e^{-u})\Big)\,e^{u}\,\exp(-e^{u-c})\,du \\ &\qquad + \int_{0}^{\infty}\Big(1-\exp(-e^{-u})\Big)\,e^{u}\,\exp(-e^{u-c})\,du\bigg\} \\ &=: e^{-c}\Big\{I(c)+I\!I(c)\Big\}\;. \end{aligned}$$

The dominated convergence theorem implies that

$$I(c) \underset{c\to\infty}{\longrightarrow} \int_{-\infty}^{0}\Big(1-\exp(-e^{-u})\Big)\,e^{u}\,du \;\in\; (0,1)\;.$$

Taylor expansion of exp at 0 and the substitution $u \mapsto u + c$ yields

$$\begin{aligned}
I\!I(c) &= \int_0^\infty \Big(e^{-u} + O(e^{-2u})\Big)\, e^u \exp(-e^{u-c})\; du \\
&= \int_0^\infty \exp(-e^{u-c})\; du \;+\; O(1) \\
&= \int_{-c}^0 \exp(-e^u)\; du \;+ \int_0^\infty \exp(-e^u)\; du \;+\; O(1) \\
&= \int_{-c}^0 \exp(-e^u)\; du \;+\; O(1)\;.
\end{aligned}$$

In order to establish equation (3.15) it suffices, therefore, to show that

$$c^{-1} \int_{-c}^0 \exp(-e^u)\; du \;\underset{c\to\infty}{\longrightarrow}\; 1\;.$$

But this follows from straightforward computations.

Finally we consider the case, where U, V are not independent. Wlog we assume that $a \geq b$ and that $(U,V) = \big(-\log(-X), -\log(-Y)\big)$, where (X,Y) follows a bivariate EVD with reverse exponential margins and dependence function D, which is not the constant function 1. The assertion is a consequence of the fact

$$\text{(3.16)} \qquad \exp\Big(\frac{c}{a+b}\Big)\, P\big(aU + bV > c\big) \;\underset{c\to\infty}{\longrightarrow}$$
$$\underset{c\to\infty}{\longrightarrow}\; \frac{b}{a+b} \int_0^1 \Big(1 - D(z) - D'(z)\,(1-z)\Big)\, z^{-\frac{a}{a+b}}\,(1-z)^{\frac{a}{a+b}-2}\; dz \;\in\; (0,\infty)\;,$$

which we establish in the following.

Put for $u < 0$

$$\tilde u \;:=\; \frac{1}{1 + \exp(-\frac{c}{b})\,(-u)^{-(a+b)/b}} \;\in\; (0,1)\;.$$

Then we have by Lemma 2.2

$$\begin{aligned}
P\big(aU + bV > c\big) &= \int_{-\infty}^0 \left(1 - P\left(Y \leq -\exp\Big(-\frac{c}{b}\Big)(-u)^{-a/b} \mid X = u\right)\right) \exp(u)\; du \\
&= \int_{-\infty}^0 \Bigg(1 - \exp\Big\{u\big(D(\tilde u) - 1\big) - \exp\Big(-\frac{c}{b}\Big)\,(-u)^{-a/b} D(\tilde u)\Big\} \\
&\qquad\qquad \times \Big(D(\tilde u) + D'(\tilde u)\,(1-\tilde u)\Big)\Bigg) \exp(u)\; du \\
&= \int_{-\infty}^0 \Big(1 - D(\tilde u) - D'(\tilde u)\,(1-\tilde u)\Big) \exp(u)\; du \\
&\quad + \int_{-\infty}^0 \Bigg(1 - \exp\Big\{u\big(D(\tilde u) - 1\big) - \exp\Big(-\frac{c}{b}\Big)\,(-u)^{-a/b} D(\tilde u)\Big\}\Bigg) \\
&\qquad\qquad \times \Big(D(\tilde u) + D'(\tilde u)\,(1-\tilde u)\Big) \exp(u)\; du \\
&=: \; \tilde I(c) + \widetilde{I\!I}(c)\;.
\end{aligned}$$

Put for $z \in (0,1)$

$$g(z) := -\exp\Big(-\frac{c}{a+b}\Big)\Big(\frac{z}{1-z}\Big)^{\frac{b}{a+b}} .$$

Then we have $\widetilde{g(z)} = z$ and, thus, with

$$g'(z) = -\exp\Big(-\frac{c}{a+b}\Big)\frac{b}{a+b}\, z^{-\frac{a}{a+b}}(1-z)^{\frac{a}{a+b}-2} ,$$

the substitution $u \mapsto g(z)$ yields

$$\begin{aligned}\tilde{I}(c) &= -\int_0^1 \Big(1 - D(z) - D'(z)(1-z)\Big)\exp\big(g(z)\big)\, g'(z)\, dz \\ &= \exp\Big(-\frac{c}{a+b}\Big)\frac{b}{a+b} \\ &\quad\times \int_0^1 \Big(1 - D(z) - D'(z)(1-z)\Big)\exp\big(g(z)\big)\, z^{-\frac{a}{a+b}}(1-z)^{\frac{a}{a+b}-2}\, dz .\end{aligned}$$

Note that the function g depends on the threshold c with $g(z) \underset{c\to\infty}{\longrightarrow} 0$ and that $g(z) < 0$, $z \in (0,1)$. The dominated convergence theorem implies, therefore, that

$$\begin{aligned}&\exp\Big(\frac{c}{a+b}\Big)\tilde{I}(c) \underset{c\to\infty}{\longrightarrow} \\ &\underset{c\to\infty}{\longrightarrow} \frac{b}{a+b}\int_0^1 \Big(1 - D(z) - D'(z)(1-z)\Big) z^{-\frac{a}{a+b}}(1-z)^{\frac{a}{a+b}-2}\, dz \in (0,\infty) .\end{aligned}$$

The integral on the right hand side is finite since $1 - D(z) \le 1 - z$ and $D'(z) \in [-1,1]$. It is positive by the arguments at the end of the proof of Theorem 3.1.

In order to establish (3.16) it suffices, therefore, to show that

$$\exp\Big(\frac{c}{a+b}\Big)\widetilde{II}(c) \underset{c\to\infty}{\longrightarrow} 0 . \tag{3.17}$$

This can be seen as follows. Choose $z_c \in (0,1)$ with $g(z_c) = -c/b$, i.e.,

$$z_c = \frac{1}{1 + \big(\frac{b}{c}\big)^{(a+b)/b}\exp\big(-\frac{c}{b}\big)} .$$

Split the integral $\widetilde{II}(c)$ into the sum of the subintervals

$$\widetilde{II}(c) = \int_{-\infty}^{g(z_c)} \cdots\, du + \int_{g(z_c)}^0 \cdots\, du .$$

The first integral is of order $O\big(\exp(-2c/(3b))\big) = o\big(\exp(-c/(a+b))\big)$; recall that we assume $a \ge b$ and that $1 - D(\tilde{u}) < 1/3$ for $u \le -c/b$ if c is large. By using again the substitution $u \mapsto g(z)$ and Taylor expansion of exp at 0, the second integral on the right hand side above equals

$$\exp\Big(-\frac{c}{a+b}\Big)\,\frac{b}{a+b}$$
$$\times\int_0^{z_c}\left(1-\exp\Big\{g(z)\,\big(D(z)-1\big)-\exp\Big(-\frac{c}{a+b}\Big)\Big(\frac{1-z}{z}\Big)^{\frac{a}{a+b}}D(z)\Big\}\right)$$
$$\times\Big(D(z)+D'(z)(1-z)\Big)\,\exp\big(g(z)\big)\,z^{-\frac{a}{a+b}}\,(1-z)^{\frac{a}{a+b}-2}\,dz\;=$$

$$=\exp\Big(-\frac{c}{a+b}\Big)\,\frac{b}{a+b}$$
$$\times\int_{1/2}^{z_c}\left(1-\exp\Big\{g(z)\,\big(D(z)-1\big)-\exp\Big(-\frac{c}{a+b}\Big)\Big(\frac{1-z}{z}\Big)^{\frac{a}{a+b}}D(z)\Big\}\right)$$
$$\times\Big(D(z)+D'(z)(1-z)\Big)\,\exp\big(g(z)\big)\,z^{-\frac{a}{a+b}}\,(1-z)^{\frac{a}{a+b}-2}\,dz$$
$$+\;o\Big(\exp\Big(-\frac{c}{a+b}\Big)\Big)$$

$$=\exp\Big(-\frac{c}{a+b}\Big)\,\frac{b}{a+b}$$
$$\times\int_{1/2}^{z_c}\left(-g(z)\,\big(D(z)-1\big)+\exp\Big(-\frac{c}{a+b}\Big)\Big(\frac{1-z}{z}\Big)^{\frac{a}{a+b}}D(z)\right)(1-z)^{\frac{a}{a+b}-2}\,O(1)\,dz$$
$$+\;o\Big(\exp\Big(-\frac{c}{a+b}\Big)\Big)$$

$$=o\Big(\exp\Big(-\frac{c}{a+b}\Big)\Big)\,,$$

which follows from elementary computations; recall that $g(z)\underset{c\to\infty}{\longrightarrow}0$ and that $1-D(z)\le 1-z$. We have, thus, established (3.17), which completes the proof of Theorem 3.4. □

In the subsequent theorems we compile the limit excess distributions of $aU+bV$ for all combinations of different marginal univariate EVDs. Note that the df of (U,V) is a bivariate EVD if, and only if the df of (V,U) is a bivariate EVD. This implies that the order of the prescribed marginal dfs of (U,V) in the subsequent results does not matter.

Theorem 3.5 (Reverse Weibull and Gumbel Margins)**.** *Suppose that* (U,V) *follows a bivariate EVD and that* $P(U\le x)=\exp(-e^{-x})$, $x\in\mathbb{R}$, $P(V\le y)=\exp\big(-(-y)^{\alpha}\big)$, $y\le 0$, $\alpha>0$. *Then we have for* $a,b>0$ *and* $t\ge 0$

$$P\Big(aU+bV>c+t\mid aU+bV>c\Big)\underset{c\to\infty}{\longrightarrow}\exp(-t/a)\,.$$

The combination of a reverse Weibull and a Gumbel margin is, consequently, dominated by the Gumbel part. The corresponding scale parameter is preserved in the limit.

Proof: Wlog we can assume the representation $U=-\log(-X)$, $V=-(-Y)^{1/\alpha}$, where (X,Y) follows a bivariate EVD with reverse exponential margins and dependence function $D(z)=1+\int_0^z M(x)-1\,dx$, see Lemma 2.1. We will establish in the following

$$(3.18)\qquad e^{c/a}P\big(aU+bV>c\big) \underset{c\to\infty}{\longrightarrow} \int_{-\infty}^{0}\exp\Big(-\frac{b}{a}(-u)^{1/\alpha}\Big)\exp(u)\,du \in (0,1)$$

if U and V are independent and

$$(3.19)\qquad e^{c/a}P\big(aU+bV>c\big) \underset{c\to\infty}{\longrightarrow} 1-M(0)\int_0^1\exp\Big(-\big(-\log(u^{a/b})\big)^{\alpha}\Big)\,du \in (0,1)$$

elsewhere. This implies the assertion.

First we establish (3.18). Conditioning on $Y=u$ we obtain

$$\begin{aligned}
P\big(aU+bV>c\big) &= \int_{-\infty}^{0}P\Big(-a\log(-X)-b(-Y)^{1/\alpha}>c \mid Y=u\Big)\exp(u)\,du\\
&= \int_{-\infty}^{0}\Bigg(1-P\Bigg(X\le-\exp\Big(-\frac{c+b(-u)^{1/\alpha}}{a}\Big)\mid Y=u\Bigg)\Bigg)\exp(u)\,du\\
&= \int_{-\infty}^{0}\Big(1-\exp\big(-e^{-c/a-b(-u)^{1/\alpha}/a}\big)\Big)\exp(u)\,du\\
&= \int_{-\infty}^{0}\Big(e^{-c/a-b(-u)^{1/\alpha}/a}+O\big(e^{-2c/a-2b(-u)^{1/\alpha}/a}\big)\Big)\exp(u)\,du
\end{aligned}$$

and, thus,

$$e^{c/a}P\big(aU+bV>c\big) \underset{c\to\infty}{\longrightarrow} \int_{-\infty}^{0}\exp\Big(-\frac{b}{a}(-u)^{1/\alpha}\Big)\exp(u)\,du\ ,$$

which is (3.18).

Next we establish (3.19). Conditioning on $X=u$ we obtain from Lemma 2.2

$$\begin{aligned}
P\big(aU+bV>c\big) &= \int_{-\infty}^{0}P\Big(-a\log(-u)-b(-Y)^{1/\alpha}>c \mid X=u\Big)\exp(u)\,du\\
&= \int_{-\exp(-c/a)}^{0}\Bigg(1-P\Bigg(Y\le-\Big(\frac{-c-a\log(-u)}{b}\Big)^{\alpha}\mid X=u\Bigg)\Bigg)\exp(u)\,du\\
&= \int_{-\exp(-c/a)}^{0}\Bigg(1-\exp\Bigg\{u\big(D(\tilde u)-1\big)-\Big(\frac{-c-a\log(-u)}{b}\Big)^{\alpha}D(\tilde u)\Bigg\}\\
&\qquad\times\Big(D(\tilde u)+D'(\tilde u)(1-\tilde u)\Big)\Bigg)\exp(u)\,du\ ,
\end{aligned}$$

where

$$\tilde u := \frac{u}{u-\Big(\frac{-c-a\log(-u)}{b}\Big)^{\alpha}} \in [0,1]\ .$$

With the substitution $u \mapsto -\exp(-c/a)\,u$, the above integral equals

$$\exp\Bigl(-\frac{c}{a}\Bigr)\int_0^1 \biggl(1-\exp\Bigl\{-\exp\Bigl(-\frac{c}{a}\Bigr)\,u\bigl(D(\bar u)-1\bigr)-\bigl(-\log(u^{a/b})\bigr)^{\alpha} D(\bar u)\Bigr\} \\ \times \Bigl(D(\bar u)+D'(\bar u)\,(1-\bar u)\Bigr)\biggr)\exp\Bigl(-\exp\Bigl(-\frac{c}{a}\Bigr)\,u\Bigr)\,du\ ,$$

where for $u \in (0,1)$

$$\bar u := \bigl(-\widetilde{\exp(-c/a)}\,u\bigr) = \frac{u}{u+\exp(c/a)\,\bigl(-\log(u^{a/b})\bigr)^{\alpha}} \ \underset{c\to\infty}{\downarrow}\ 0\ .$$

We obtain, consequently,

$$\begin{aligned}
\exp\Bigl(\frac{c}{a}\Bigr)\,P\bigl(aU+bV>c\bigr) &= \\
&= \int_0^1 \Bigl(1-D(\bar u)-D'(\bar u)\,(1-\bar u)\Bigr)\exp\Bigl(-\exp\Bigl(-\frac{c}{a}\Bigr)\,u\Bigr)\,du \\
&\quad + \int_0^1 \biggl\{1-\exp\Bigl(-\exp\Bigl(-\frac{c}{a}\Bigr)\,u\,\bigl(1-D(\bar u)\bigr)-\bigl(-\log(u^{a/b})\bigr)^{\alpha} D(\bar u)\Bigr)\biggr\} \\
&\qquad \times \Bigl(D(\bar u)+D'(\bar u)\,(1-\bar u)\Bigr)\exp\Bigl(-\exp\Bigl(-\frac{c}{a}\Bigr)\,u\Bigr)\,du \\
&\underset{c\to\infty}{\longrightarrow} -D'(0) + \int_0^1 \Bigl\{1-\exp\Bigl(-\bigl(-\log(u^{a/b})\bigr)^{\alpha}\Bigr)\Bigr\}\,\bigl(1+D'(0)\bigr)\,du \\
&= 1-M(0)\int_0^1 \exp\Bigl(-\bigl(-\log(u^{a/b})\bigr)^{\alpha}\Bigr)\,du \ \in\ (0,1)\ .
\end{aligned}$$

Note that necessarily $M(0)<1$. Otherwise we had $D(z) = 1+\int_0^z M(x)-1\,dx \ge 1$ and, thus, D would be the constant function 1. But this case was excluded. Thus we have established (3.19), which completes the proof of Theorem 3.5. □

Theorem 3.6 (Reverse Weibull and Fréchet Margins)**.** *Suppose that* (U,V) *follows a bivariate EVD with* $P(U\le x) = \exp\bigl(-(-x)^{\alpha_1}\bigr)$, $x\le 0$, *and* $P(V\le y) = \exp(-y^{-\alpha_2})$, $y>0$, $\alpha_1,\alpha_2>0$. *Then we have for* $a,b>0$ *and* $t\ge 1$

$$P\Bigl(aU+bV>t\,c \mid aU+bV>c\Bigr) \underset{c\to\infty}{\longrightarrow} t^{-\alpha_2}\ .$$

The combination of a reverse Weibull and a Fréchet margin is, therefore, determined by the Fréchet part. The limit excess df is independent of the scale parameters.

Proof: It is sufficient to show that for $n\in\mathbb{N}$

$$n^{-1/\alpha_2}\max_{1\le i\le n}\bigl(aU_i+bV_i\bigr) \underset{D}{\longrightarrow} \exp\bigl(-(y/b)^{-\alpha_2}\bigr)\ , \qquad y>0\ , \tag{3.20}$$

where (U_1, V_1), (U_2, V_2), ... are independent copies of (U, V). But (3.20) is immediate from the inequalities

$$a \min_{1\le i\le n} U_i + b \max_{1\le i\le n} V_i \;\le\; \max_{1\le i\le n}\big(a U_i + b V_i\big) \;\le\; b \max_{1\le i\le n} V_i$$

and the facts that

$$n^{-1/\alpha_2} \max_{1\le i\le n} V_i \;\underset{D}{=}\; \exp(-y^{-\alpha_2}) \;, \qquad y > 0 \;,$$
$$n^{-1/\alpha_2} \min_{1\le i\le n} U_i \;\underset{n\to\infty}{\longrightarrow}\; 0 \qquad \text{in probability} \;. \qquad \square$$

Theorem 3.7 (Fréchet and Gumbel Margins)**.** *Suppose that the rv* (U,V) *follows a bivariate EVD with* $P(U \le x) = \exp(-x^{-\alpha})$*,* $x > 0$*, where* $\alpha > 0$*, and* $P(V \le y) = \exp(-e^{-y})$*,* $y \in \mathbb{R}$*. Then we have for* $a, b > 0$ *and* $t \ge 1$

$$P\Big(aU + bV > tc \mid aU + bV > c\Big) \;\underset{c\to\infty}{\longrightarrow}\; t^{-\alpha} \;.$$

The combination of a Fréchet and a Gumbel margin is, consequently, determined by the Fréchet part. The limit excess df is independent of the scale parameters.

Proof: It is sufficient to show that for $n \in \mathbb{N}$

$$n^{-1/\alpha} \max_{1\le i\le n}\big(a U_i + b V_i\big) \;\underset{D}{\longrightarrow}\; \exp\big(-(x/b)^{-\alpha}\big) \;, \qquad x > 0 \;, \tag{3.21}$$

where (U_1, V_1), (U_2, V_2), ... are independent copies of (U, V). But (3.21) is immediate from the inequalities

$$a \max_{1\le i\le n} U_i + b \min_{1\le i\le n} V_i \;\le\; \max_{1\le i\le n}\big(a U_i + b V_i\big) \;\le\; a \max_{1\le i\le n} U_i + b \max_{1\le i\le n} V_i$$

and the facts that

$$n^{-1/\alpha} \max_{1\le i\le n} U_i \;\underset{D}{=}\; \exp(-x^{-\alpha}) \;, \qquad x > 0 \;,$$
$$n^{-1/\alpha} \min_{1\le i\le n} V_i \underset{n\to\infty}{\longrightarrow} 0 \,, \quad n^{-1/\alpha} \max_{1\le i\le n} V_i \;\underset{n\to\infty}{\longrightarrow}\; 0 \qquad \text{in probability} \;. \qquad \square$$

ACKNOWLEDGMENTS

The author is indebted to the referee for bringing to his attention the papers by Alink *et al.* (2004, 2005a, 2005b) and Barbe *et al.* (2006).

REFERENCES

[1] Acerbi, C. and Sirtori, C. (2001). *Expected shortfall as a tool for financial risk management.* http://www.gloriamundi.org/picsresources/ncs.pdf.

[2] Acerbi, C. and Tasche, D. (2001). *Expected shortfall: a natural coherent alternative to value at risk.* http://www.gloriamundi.org/picsresources/expshortfall.pdf.

[3] Acerbi, C. and Tasche, D. (2002). *On the coherence of expected shortfall.* http://www.gloriamundi.org/picsresources/cadt.pdf.

[4] Alink, S.; Löwe, M. and Wüthrich, M.V. (2004). Diversification of aggregate dependent risks, *Insurance: Mathematics and Economics*, **35**, 77–95.

[5] Alink, S.; Löwe, M. and Wüthrich, M.V. (2005a). Analysis of the expected shortfall of aggregate dependent risks, *Astin Bulletin*, **35**, 25–43.

[6] Alink, S.; Löwe, M. and Wüthrich, M.V. (2005b). *Analysis of the diversification effect of aggregate dependent risks.* http://www.math.ethz.ch/~wueth/papers2.html.

[7] Barbe, Ph.; Fougères, A.-L. and Genest, C. (2006). On the tail behaviour of dependent risks, *Astin Bulletin*, **36**, 361–373.

[8] Dupuis, D.J. and Tawn, J.A. (2001). Effects of mis-specification in bivariate extreme value problems, *Extremes*, **2**, 339–365.

[9] Embrechts, P. (2000). *Extremes and Integrated Risk Management*, Risk Books, London.

[10] Embrechts, P. (2004). *Extremes in economics and the economics of extremes.* In "Extreme Values in Finance, Telecommunications, and the Environment" (B. Finkenstädt and H. Rootzén, Eds.), Chapman & Hall/CRC, Boca Raton, 169–183.

[11] Falk, M. (2006). A representation of bivariate extreme value distributions via norms on $\mathbb{R}^2$, *Extremes*, **9**, 63–68.

[12] Falk, M. and Michel, R. (2006). Testing for tail independence in extreme value models, *Ann. Inst. Statist. Math.*, **58**, 261–290.

[13] Falk, M.; Hüsler, J. and Reiss, R.-D. (2004). *Laws of Small Numbers: Extremes and Rare Events*, 2nd ed., Birkhäuser, Basel.

[14] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Volume 2, 2nd ed., Wiley, New York.

[15] Ghoudi, K.; Khoudraji, A. and Rivest, L.P. (1998). Statistical properties of couples of bivariate extreme–value copulas, *Canad. J. Statist.*, **26**, 187–197.

[16] Klüppelberg, C. (2004). *Risk management with extreme value theory.* In "Extreme Values in Finance, Telecommunications, and the Environment" (B. Finkenstädt and H. Rootzén, Eds.), Chapman & Hall/CRC, Boca Raton, 101–168.

[17] Klüppelberg, C.; Lindner, A. and Maller, R. (2006). *Continuous time volatility modelling: COGARCH versus Ornstein-Uhlenbeck models.* In "From Stochastic Calculus to Mathematical Finance. The Shiryaev Festschrift" (Yu. Kabanov, R. Lipster and J. Stoyanov, Eds.), Springer, Berlin, 393–419.

[18] Macke, M. (2005). *On the distribution of linear combinations of random variables with underlying multivariate extreme value and generalized Pareto distributions and an application to the expected shortfall of portfolios*, Diploma thesis (in German), University of Würzburg.

[19] Reiss, R.-D. and Thomas, M. (2001). *Statistical Analysis of Extreme Values*, 2nd ed., Birkhäuser, Basel.

[20] Wüthrich, M.V. (2003). Asymptotic value-at-risk estimates for sums of dependent random variables, *Astin Bulletin*, **33**, 75–92.

REVSTAT – Statistical Journal
Volume 5, Number 2, June 2007, 163–176

DECOMPOSITIONS OF MARGINAL HOMOGENEITY MODEL USING CUMULATIVE LOGISTIC MODELS FOR MULTI-WAY CONTINGENCY TABLES

Authors: Kouji Tahata
– Graduate School of Sciences and Technology,
Tokyo University of Science, Noda City, Chiba, 278-8510, Japan
tahata@jul.rikadai.jp

Shingo Katakura
– Graduate School of Sciences and Technology,
Tokyo University of Science, Noda City, Chiba, 278-8510, Japan

Sadao Tomizawa
– Department of Information Sciences, Faculty of Science and Technology,
Tokyo University of Science, Noda City, Chiba, 278-8510, Japan
tomizawa@is.noda.tus.ac.jp

Received: November 2006 Accepted: January 2007

Abstract:

- For square contingency tables with ordered categories, Agresti (1984, 2002) considered the marginal cumulative logistic (ML) model, which is an extension of the marginal homogeneity (MH) model. Miyamoto, Niibe and Tomizawa (2005) proposed the conditional marginal cumulative logistic (CML) model which is defined off the main diagonal cells, and gave the decompositions of the MH model using the ML (CML) model. This paper (1) considers the ML and CML models for multi-way tables, and (2) gives the decompositions of the MH model into the ML (CML) model and the model of the equality of marginal means for multi-way tables. An example is given.

Key-Words:

- *decomposition; marginal cumulative logistic model; marginal homogeneity; marginal mean; multi-way contingency table.*

AMS Subject Classification:

- 62H17.

1. INTRODUCTION

For an $R\times R$ square contingency table with ordered categories, let p_{ij} denote the probability that an observation will fall in the cell in row i and column j ($i=1,...,R;\ j=1,...,R$), and let X_1 and X_2 denote the row and column variables, respectively. The marginal homogeneity (MH) model is defined by

$$\Pr(X_1=i) = \Pr(X_2=i) \qquad \text{for} \quad i=1,...,R \ ;$$

that is

$$p_{i\cdot} = p_{\cdot i} \qquad \text{for} \quad i=1,...,R \ ,$$

where $p_{i\cdot} = \sum_{k=1}^{R} p_{ik}$ and $p_{\cdot i} = \sum_{k=1}^{R} p_{ki}$. This model indicates that the row marginal distribution is identical to the column marginal distribution (Stuart, 1955; Bhapkar, 1966; Bishop, Fienberg and Holland, 1975, p. 294; Tomizawa, 1991, 1993, 1998). Let $F_i^{(1)}$ and $F_i^{(2)}$ denote the marginal cumulative probabilities of X_1 and X_2, respectively. These are $F_i^{(1)} = \Pr(X_1 \le i) = \sum_{k=1}^{i} p_{k\cdot}$ and $F_i^{(2)} = \Pr(X_2 \le i) = \sum_{k=1}^{i} p_{\cdot k}$ for $i=1,...,R-1$. Then the MH model may also be expressed as

$$F_i^{(1)} = F_i^{(2)} \qquad \text{for} \quad i=1,...,R-1 \ .$$

Let $L_i^{(1)}$ and $L_i^{(2)}$ denote the marginal cumulative logit of X_1 and X_2, respectively. These are given as

$$L_i^{(1)} = \operatorname{logit}\bigl[\Pr(X_1 \le i)\bigr] = \log\left[\frac{\Pr(X_1 \le i)}{1-\Pr(X_1 \le i)}\right] ,$$

and

$$L_i^{(2)} = \operatorname{logit}\bigl[\Pr(X_2 \le i)\bigr] = \log\left[\frac{\Pr(X_2 \le i)}{1-\Pr(X_2 \le i)}\right] ,$$

for $i=1,...,R-1$. Then the MH model may be further expressed as

$$L_i^{(1)} = L_i^{(2)} \qquad \text{for} \quad i=1,...,R-1 \ .$$

As an extension of the MH model, Agresti (1984, p. 205; 2002, p. 420) considered the marginal cumulative logistic (ML) model defined by

$$L_i^{(1)} = L_i^{(2)} + \Delta \qquad \text{for} \quad i=1,...,R-1 \ .$$

This model states that the odds that X_1 is i or below instead of $i+1$ or above, is $\exp(\Delta)$ times higher than the odds that X_2 is i or below instead of $i+1$ or above, for every $i=1,...,R-1$. Note that the MH model implies the ML model. Consider the marginal mean equivalence (ME) model defined by

$$\sum_{i=1}^{R} i\, p_{i\cdot} = \sum_{i=1}^{R} i\, p_{\cdot i} \qquad (\text{i.e.,}\ \ E(X_1) = E(X_2)) \ .$$

Miyamoto, Niibe and Tomizawa (2005) gave the following theorem.

Theorem 1.1. *The MH model holds if and only if both the ML and ME models hold.*

Using the conditional probabilities, the MH model may also be expressed as

$$\Pr\Big(X_1 = i \mid X_1 \neq X_2\Big) = \Pr\Big(X_2 = i \mid X_1 \neq X_2\Big) \qquad \text{for} \quad i = 1, ..., R \ ;$$

that is

$$p^c_{i\cdot} = p^c_{\cdot i} \qquad \text{for} \quad i = 1, ..., R \ ,$$

where

$$\begin{aligned} p^c_{i\cdot} &= \frac{p_{i\cdot} - p_{ii}}{\delta} = \Pr\Big(X_1 = i \mid X_1 \neq X_2\Big) \ , \\ p^c_{\cdot i} &= \frac{p_{\cdot i} - p_{ii}}{\delta} = \Pr\Big(X_2 = i \mid X_1 \neq X_2\Big) \ , \\ \delta &= \sum\sum_{s \neq t} p_{st} = \Pr\big(X_1 \neq X_2\big) \ . \end{aligned}$$

Let $F^{c(1)}_i$ and $F^{c(2)}_i$ denote the conditional marginal cumulative probabilities of X_1 and X_2 given that $X_1 \neq X_2$, i.e.,

$$\begin{aligned} F^{c(1)}_i &= \Pr\Big(X_1 \leq i \mid X_1 \neq X_2\Big) = \sum_{k=1}^{i} p^c_{k\cdot} \ , \\ F^{c(2)}_i &= \Pr\Big(X_2 \leq i \mid X_1 \neq X_2\Big) = \sum_{k=1}^{i} p^c_{\cdot k} \ , \end{aligned}$$

for $i = 1, ..., R-1$. Then the MH model may be further expressed as $F^{c(1)}_i = F^{c(2)}_i$ for $i = 1, ..., R-1$. Miyamoto *et al.* (2005) also considered the conditional marginal cumulative logistic (CML) model defined by

$$L^{c(1)}_i = L^{c(2)}_i + \Delta^* \qquad \text{for} \quad i = 1, ..., R-1 \ ,$$

where

$$\begin{aligned} L^{c(1)}_i &= \operatorname{logit}\Big[\Pr\Big(X_1 \leq i \mid X_1 \neq X_2\Big)\Big] \ , \\ L^{c(2)}_i &= \operatorname{logit}\Big[\Pr\Big(X_2 \leq i \mid X_1 \neq X_2\Big)\Big] \ . \end{aligned}$$

Miyamoto *et al.* (2005) also gave the following theorem.

Theorem 1.2. *The MH model holds if and only if both the CML and ME models hold.*

For analyzing the data of multi-way tables of the same classifications with ordered categories, the some models of symmetry, e.g., the symmetry model, the MH model (e.g., Bishop *et al.* 1975, pp. 300–307), and the ML model (Agresti, 2002, pp. 439–440) are applied. The symmetry and the MH models do not depend on the main diagonal cell probabilities, however, the ML model depends on them. So, we are now interested in the another ML model which does not depend on the main diagonal cell probabilities, namely, in the conditional ML model on condition that an observation will fall in one of off-diagonal cells of the table.

The purpose of this paper is (1) to extend the CML model into the multi-way tables (Section 2.4) and (2) to extend Theorems 1.1 and 1.2 into the multi-way tables (Section 3).

2. EXTENSION TO MULTI-WAY TABLES

2.1. The MH model

Consider an R^T table ($T \geq 3$) having ordered categories. Let X_t denote the t-th random variable for $t = 1, ..., T$ and let $\Pr(X_1 = i_1, ..., X_T = i_T) = p_{i_1 \ldots i_T}$ for $i_t = 1, ..., R$. The marginal homogeneity (MH) model is defined by

$$\Pr(X_1 = i) = \cdots = \Pr(X_T = i) \qquad \text{for} \quad i = 1, ..., R \;;$$

that is

$$p_i^{(1)} = \cdots = p_i^{(T)} \qquad \text{for} \quad i = 1, ..., R \;,$$

where

$$p_i^{(t)} = \Pr(X_t = i) \qquad \text{for} \quad t = 1, ..., T \;.$$

Let $F_i^{(t)}$ denote the marginal cumulative probabilities and let $L_i^{(t)}$ denote the marginal cumulative logit of X_t for $i = 1, ..., R-1$; $t = 1, ..., T$. Namely, $F_i^{(t)} = \sum_{s=1}^{i} p_s^{(t)}$, and $L_i^{(t)} = \mathrm{logit}\big[\Pr(X_t \leq i)\big]$. Then the MH model may also be expressed as

$$F_i^{(k)} = F_i^{(1)} \qquad \text{for} \quad i = 1, ..., R-1; \quad k = 2, ..., T \;,$$

or

$$L_i^{(k)} = L_i^{(1)} \qquad \text{for} \quad i = 1, ..., R-1; \quad k = 2, ..., T \;.$$

2.2. The ML model

Agresti (2002, p. 442) considered the marginal cumulative logistic (ML) model, defined by

$$L_i^{(k)} = L_i^{(1)} - \Delta_{k-1} \qquad \text{for} \quad i = 1, ..., R-1; \quad k = 2, ..., T\ .$$

By putting $L_i^{(1)} = \theta_i$, this model may be expressed as

$$F_i^{(k)} = \frac{\exp(\theta_i - \Delta_{k-1})}{1 + \exp(\theta_i - \Delta_{k-1})} \qquad \text{for} \quad i = 1, ..., R-1; \quad k = 1, ..., T\ ,$$

where $\Delta_0 = 0$. A special case of this model obtained by putting $\Delta_1 = \cdots = \Delta_{T-1} = 0$ is the MH model.

2.3. Other expressions of MH model

The MH model may also be expressed as

$$\Pr\Big(X_k = i \mid (X_1, ..., X_T) \neq (s, ..., s),\ s = 1, ..., R\Big) = $$
$$= \Pr\Big(X_1 = i \mid (X_1, ..., X_T) \neq (s, ..., s),\ s = 1, ..., R\Big)\ ,$$

for $i = 1, ..., R$; $k = 2, ..., T$; that is

$$p_i^{c(k)} = p_i^{c(1)} \qquad \text{for} \quad i = 1, ..., R; \quad k = 2, ..., T\ ,$$

where, for $m = 1, ..., T$,

$$p_i^{c(m)} = \frac{p_i^{(m)} - p_{ii\cdots i}}{\delta} = \Pr\Big(X_m = i \mid (X_1, ..., X_T) \neq (s, ..., s),\ s = 1, ..., R\Big)\ ,$$

$$\delta = 1 - \sum_{i=1}^{R} p_{ii\cdots i} = \Pr\Big((X_1, ..., X_T) \neq (s, ..., s),\ s = 1, ..., R\Big)\ .$$

Let $F_i^{c(k)}$ denote the conditional marginal cumulative probabilities of X_k given that $(X_1, ..., X_T) \neq (s, ..., s)$, $s = 1, ..., R$, i.e.,

$$F_i^{c(k)} = \Pr\Big(X_k \leq i \mid (X_1, ..., X_T) \neq (s, ..., s),\ s = 1, ..., R\Big) = \sum_{t=1}^{i} p_t^{c(k)}$$

for $i = 1, ..., R-1$; $k = 1, ..., T$. Then the MH model may be further expressed as

$$F_i^{c(k)} = F_i^{c(1)} \qquad \text{for} \quad i = 1, ..., R-1; \ k = 2, ..., T\ .$$

2.4. The CML model

Consider now a model defined by

$$L_i^{c(k)} = L_i^{c(1)} - \Delta_{k-1}^* \qquad \text{for} \quad i=1,...,R-1;\; k=2,...,T\;,$$

where, for $m=1,...,T$,

$$\begin{aligned} L_i^{c(m)} &= \operatorname{logit}\Big[\Pr\Big(X_m \le i \mid (X_1,...,X_T) \ne (s,...,s),\; s=1,...,R\Big)\Big] \\ &= \log\left[\frac{\Pr\Big(X_m \le i \mid (X_1,...,X_T) \ne (s,...,s),\; s=1,...,R\Big)}{1-\Pr\Big(X_m \le i \mid (X_1,...,X_T) \ne (s,...,s),\; s=1,...,R\Big)}\right]. \end{aligned}$$

We shall refer to this model as the conditional marginal cumulative logistic (CML) model. By putting $L_i^{c(1)} = \theta_i^*$, this model may be expressed as

$$F_i^{c(k)} = \frac{\exp(\theta_i^* - \Delta_{k-1}^*)}{1+\exp(\theta_i^* - \Delta_{k-1}^*)} \qquad \text{for} \quad i=1,...,R-1;\; k=1,...,T\;,$$

where $\Delta_0^* = 0$. A special case of the CML model obtained by putting $\Delta_1^* = \cdots = \Delta_{T-1}^* = 0$ is the MH model.

The CML model states that for $k=2,...,T$, on condition that the values of random variables are not all same, the odds that X_1 is i or below instead of $i+1$ or above, is $\exp(\Delta_{k-1}^*)$ times higher than the odds that X_k is i or below instead of $i+1$ or above, for every $i=1,...,R-1$. Thus, if $\Delta_{k-1}^* > 0$, on the same condition, X_1 rather than X_k tends to be i or below instead of $i+1$ or above for every $i=1,...,R-1$.

3. DECOMPOSITIONS OF THE MARGINAL HOMOGENEITY MODEL

We shall consider two kinds of decompositions of the MH model.

3.1. A decomposition of the MH model using the ML model

Consider a model defined as

$$\sum_{i=1}^{R} i\, p_i^{(1)} = \cdots = \sum_{i=1}^{R} i\, p_i^{(T)} \qquad (\text{i.e.,}\;\; E(X_1) = \cdots = E(X_T))\;. \tag{3.1}$$

Namely, the means of variables X_k $(k=1,...,T)$ are equal. Note that the MH model implies model (3.1).

Consider a specified monotonic function $g(k)$ satisfying $g(1) \leq \cdots \leq g(R)$ or $g(1) \geq \cdots \geq g(R)$, where at least one strict inequality holds. Using the function $g(k)$, model (3.1) is generalized as

$$(3.2) \qquad \sum_{i=1}^{R} g(i)\, p_i^{(1)} = \cdots = \sum_{i=1}^{R} g(i)\, p_i^{(T)} \qquad (\text{i.e., } \; E(g(X_1)) = \cdots = E(g(X_T)))\ .$$

We shall refer to (3.2) as the marginal mean equivalence (ME) model.

The $\{g(k)\}$ may be considered as the ordered scores $\{u_k\}$ assigned to the categories if it is possible to assign the scores; namely, $g(k) = u_k$ satisfying $u_1 \leq \cdots \leq u_R$ or $u_1 \geq \cdots \geq u_R$. In particular, when the scores are equal-interval; that is, when $u_2 - u_1 = u_3 - u_2 = \cdots = u_R - u_{R-1}$, then the ME model with $g(k) = u_k$ is equivalent to the model (3.1). We now obtain the following theorem.

Theorem 3.1. *For multi-way tables, the MH model holds if and only if both the ML and ME models hold.*

Proof: If the MH model holds, then both the ML and ME models hold. Therefore, assuming that both the ML and ME models hold, we shall show that the MH model holds. We have

$$\begin{aligned} E(g(X_1)) &= \sum_{k=1}^{R} g(k)\, p_k^{(1)} \\ &= g(1) + \sum_{k=2}^{R} \left(d_k \sum_{i=k}^{R} p_i^{(1)} \right) \\ &= g(1) + \sum_{k=2}^{R} d_k \big(1 - F_{k-1}^{(1)}\big) \\ &= g(R) - \sum_{k=2}^{R} d_k\, F_{k-1}^{(1)}\ , \end{aligned}$$

where

$$d_k = g(k) - g(k-1)\ .$$

Similarly, we have

$$E(g(X_2)) = g(R) - \sum_{k=2}^{R} d_k\, F_{k-1}^{(2)}\ .$$

This yields

$$E(g(X_2)) - E(g(X_1)) = \sum_{k=2}^{R} d_k \big(F_{k-1}^{(1)} - F_{k-1}^{(2)}\big)\ .$$

Since the ML and ME models hold, we obtain

$$\sum_{k=2}^{R} d_k \left(\frac{\exp(\theta_{k-1})}{1+\exp(\theta_{k-1})} - \frac{\exp(\theta_{k-1}-\Delta_1)}{1+\exp(\theta_{k-1}-\Delta_1)} \right) = 0 \ .$$

Thus

$$\Big(1-\exp(-\Delta_1)\Big) \sum_{k=2}^{R} d_k \frac{\exp(\theta_{k-1})}{\big(1+\exp(\theta_{k-1})\big)\big(1+\exp(\theta_{k-1}-\Delta_1)\big)} = 0 \ .$$

Then

$$\sum_{k=2}^{R} d_k \frac{\exp(\theta_{k-1})}{\big(1+\exp(\theta_{k-1})\big)\big(1+\exp(\theta_{k-1}-\Delta_1)\big)} \neq 0 \ ,$$

because $d_k \geq 0$ for all $k = 2, ..., R$ (or $d_k \leq 0$ for all $k = 2, ..., R$), with at least one of the d_k's being not equal to zero. Therefore we obtain $\Delta_1 = 0$. In the similar way, we obtain $\Delta_k = 0$ for $k = 2, ..., T-1$. Thus, the MH model holds. The proof is completed. □

3.2. A decomposition of the MH model using the CML model

We now obtain the following theorem.

Theorem 3.2. *For multi-way tables, the MH model holds if and only if both the CML and ME models hold.*

We omit the proof because it can be obtained in a similar way as the proof of Theorem 3.1.

Generally, consider a decomposition of model such that model M_1 holds if and only if both models M_2 and M_3 hold. When models M_1 and M_2 fit the data poorly but model M_3 fits the data well, we can then understand that the poor fit of model M_1 is caused by the lack of structure of model M_2 rather than the structure of model M_3. Thus, the decomposition of model M_1 may be useful to see the reason for the poor fit of model M_1.

Let $n_{i_1 \cdots i_T}$ denote the observed frequency in the $(i_1, ..., i_T)$ cell of the R^T table with $n = \sum \cdots \sum n_{i_1 \cdots i_T}$, and let $m_{i_1 \cdots i_T}$ denote the corresponding expected frequency. We assume that $\{n_{i_1 \cdots i_T}\}$ have a multinomial distribution. The maximum likelihood estimates (MLEs) of the expected frequencies under each model can be obtained using a Newton–Raphson method to solve the likelihood equation (see Appendix for the CML model). Denote the likelihood ratio chi-squared statistic for testing the goodness-of-fit of model M by $G^2(M)$. For testing that model M_1 holds assuming that model M_2 holds true, the likelihood ratio statistic is given as $G^2(M_1 \,|\, M_2) = G^2(M_1) - G^2(M_2)$ (≥ 0). The numbers of degrees of freedom (df) for testing the goodness-of-fit of the MH, ML (CML), and ME models are $(T-1)(R-1)$, $(T-1)(R-2)$, and $T-1$, respectively.

Table 1: Opinions about government spending; from Lang and Agresti (1994). The upper and lower paranthesized values are the MLEs of expected frequencies under the ML and CML models, respectively.

Cities	Health	Law Enforcement		
		(1)	(2)	(3)
(1)	(1)	76 (71.31) (76.00)	20 (17.03) (21.00)	5 (5.92) (5.66)
(1)	(2)	13 (12.29) (15.22)	11 (9.43) (12.59)	0 (0.00) (0.00)
(1)	(3)	4 (3.68) (3.31)	3 (2.51) (2.44)	2 (2.31) (1.72)
(2)	(1)	113 (122.83) (108.92)	56 (54.44) (52.96)	5 (7.16) (5.06)
(2)	(2)	30 (32.89) (31.29)	28 (27.43) (28.00)	1 (1.45) (1.10)
(2)	(3)	4 (4.25) (3.04)	1 (0.95) (0.75)	2 (2.78) (1.58)
(3)	(1)	103 (100.86) (103.88)	41 (36.28) (40.54)	15 (18.76) (15.92)
(3)	(2)	29 (28.61) (31.77)	21 (18.71) (22.51)	5 (6.32) (5.79)
(3)	(3)	6 (5.76) (4.73)	8 (6.95) (6.21)	5 (6.09) (5.00)

Note: (1) – too little; (2) – about right; (3) – too much.

4. EXAMPLE

The data in Table 1, taken directly from Lang and Agresti (1994), is the 1989 General Social Survey conducted by the National Opinion Research Center at the University of Chicago. Subjects in the sample were asked their opinion regarding government spending on the health (X_1), the law enforcement (X_2), and the assistance to big cities (X_3). The common response scale is (1) too little, (2) about right, and (3) too much. Table 2 presents the values of likelihood ratio statistic G^2 for each model.

The MH model fits these data very poorly. However the CML model fits these data well although the ML model does not fit so well. Also, the ME model with $g(k) = k$, $k = 1, 2, 3$, fits these data very poorly.

Consider the hypothesis that the MH model holds under the assumption that the ML (CML) model holds; namely, the hypothesis that $\Delta_1 = \Delta_2 = 0$ ($\Delta_1^* = \Delta_2^* = 0$) under the assumption. Because $G^2(\text{MH}|\text{ML}) = G^2(\text{MH}) - G^2(\text{ML}) = 328.57$ and $G^2(\text{MH}|\text{CML}) = G^2(\text{MH}) - G^2(\text{CML}) = 331.49$ with 2 df, we reject these hypotheses at the 0.05 level. These show the rejection of $\Delta_1 = \Delta_2 = 0$ ($\Delta_1^* = \Delta_2^* = 0$) in the ML (CML) model.

Under the CML model the MLEs of $\exp(\Delta_k^*)$ are $\exp(\hat{\Delta}_1^*) = 1.59$ and $\exp(\hat{\Delta}_2^*) = 17.3$ (i.e., $\hat{\Delta}_1^* = 0.46$ and $\hat{\Delta}_2^* = 2.85$). Thus, the CML model provides that (1) under the condition that the opinions are not all same, the odds that the opinion is 'too little' instead of not 'too little' are estimated to be 1.59 times higher in health than in law, and (2) the odds that the opinion is not 'too much' instead of 'too much' are estimated to be 1.59 times higher in health than in law, and similarly, (3) the odds that the opinion is 'too little' instead of not 'too little' are estimated to be 17.3 times higher in health than in cities, and (4) the odds that the opinion is not 'too much' instead of 'too much' are estimated to be 17.3 times higher in health than in cities.

Table 2: Likelihood ratio statistic G^2 for models applied to the data in Table 1.

Models	Table 1	
	df	G^2
MH	4	334.62*
ML	2	6.05*
CML	2	3.13
ME	2	316.01*

* means significant at 0.05 level.

Note: $g(k)$ for the ME model are the equal-interval scores.

5. CONCLUDING REMARKS

When the MH model fits the data poorly, the decompositions of the MH model may be useful for seeing the reason for its poor fit. Indeed, for the data in Table 1, the poor fit of the MH model is caused by the poor fit of the ME model rather than the ML (or CML) model.

Each of the MH, CML and ME models does not depend on the probabilities $\{p_{ii\cdots i}\}$ on the main diagonal of the table, but the ML model depends on them. Notice that the estimated expected frequencies on the main diagonal cells under the ML model are different from the observed frequencies on the main diagonal (see Table 1).

When the MH model does not hold, if we want to see the reason why the equalities of the conditional marginal cumulative probabilities $\{F_i^{c(k)}\}$ do not hold, the analyst would be interested in inferring the structure of only off-diagonal probabilities. In this case, the decomposition of the MH model into the CML and ME models may be preferable to that into the ML and ME models.

Also, the MH model indicates the equalities of marginal cumulative probabilities $\{F_i^{(k)}\}$, which include the probabilities $\{p_{ii\cdots i}\}$ on the main diagonal. Therefore, when the MH model does not hold, if we want to see the reason why the equalities of $\{F_i^{(k)}\}$ do not hold, the analyst would be interested in inferring the structure of $\{F_i^{(k)}\}$. Then, the decomposition of the MH model into the ML and ME models may be preferable to that into the CML and ME models.

The decompositions of the MH model described here should be considered for ordinal categorical data, because each of the decomposed models is not invariant under the same arbitrary permutations of all categories.

APPENDIX

We consider the MLEs of the expected frequencies $\{m_{ijt}\}$ under the CML model. We give the case of three-way table below and omit the case of more multi-way table because those are obtained in the similar way.

To obtain the MLEs under the CML model, we must maximize the Lagrangian

$$\begin{aligned} L \;=\; & \sum_{i=1}^{R}\sum_{j=1}^{R}\sum_{t=1}^{R} n_{ijt}\log p_{ijt} \;-\; \mu\left(\sum_{i=1}^{R}\sum_{j=1}^{R}\sum_{t=1}^{R} p_{ijt} - 1\right) \\ & -\sum_{i=1}^{R-1}\lambda_{1i}\Big(F_i^{c(1)}\big(1-F_i^{c(2)}\big) - \exp(\Delta_1^*)\big(1-F_i^{c(1)}\big)F_i^{c(2)}\Big) \\ & -\sum_{i=1}^{R-1}\lambda_{2i}\Big(F_i^{c(1)}\big(1-F_i^{c(3)}\big) - \exp(\Delta_2^*)\big(1-F_i^{c(1)}\big)F_i^{c(3)}\Big) \end{aligned}$$

with respect to $\{p_{ijt}\}$, μ, $\{\lambda_{1i}\}$, $\{\lambda_{2i}\}$, Δ_1^*, and Δ_2^*. Setting the partial derivatives of L equal to zeros, we obtain the equations

$$\begin{aligned} m_{ijt} &= \frac{n_{ijt}}{1+\frac{1}{n}(T_{1ij}+T_{2it})} \qquad \text{for} \quad i,j,t=1,...,R; \quad (i,j,t)\neq(i,i,i)\;, \\ m_{iii} &= n_{iii} \qquad \text{for} \quad i=1,...,R\;, \end{aligned}$$

where

$$\begin{aligned} T_{1ij} \;=\; \delta^{-1}\sum_{u=1}^{R-1}\Bigg[& I_{(u\ge i)}\,I_{(u\ge j)}\,\lambda_{1u}\Big\{\big(1-F_u^{c(2)}\big) - \exp(\Delta_1^*)\big(1-F_u^{c(1)}\big)\Big\} \\ & + I_{(u<i)}\,I_{(u\ge j)}\,\lambda_{1u}\Big\{-\exp(\Delta_1^*)\Big(F_u^{c(2)} + \big(1-F_u^{c(1)}\big)\Big)\Big\} \\ & + I_{(u\ge i)}\,I_{(u<j)}\,\lambda_{1u}\Big\{\big(1-F_u^{c(2)}\big) + F_u^{c(1)}\Big\} \\ & + I_{(u<i)}\,I_{(u<j)}\,\lambda_{1u}\Big\{F_u^{c(1)} - \exp(\Delta_1^*)\,F_u^{c(2)}\Big\}\Bigg]\,, \end{aligned}$$

$$\begin{aligned} T_{2it} \;=\; \delta^{-1}\sum_{u=1}^{R-1}\Bigg[& I_{(u\ge i)}\,I_{(u\ge t)}\,\lambda_{2u}\Big\{\big(1-F_u^{c(3)}\big) - \exp(\Delta_2^*)\big(1-F_u^{c(1)}\big)\Big\} \\ & + I_{(u<i)}\,I_{(u\ge t)}\,\lambda_{2u}\Big\{-\exp(\Delta_2^*)\Big(F_u^{c(3)} + \big(1-F_u^{c(1)}\big)\Big)\Big\} \\ & + I_{(u\ge i)}\,I_{(u<t)}\,\lambda_{2u}\Big\{\big(1-F_u^{c(3)}\big) + F_u^{c(1)}\Big\} \\ & + I_{(u<i)}\,I_{(u<t)}\,\lambda_{2u}\Big\{F_u^{c(1)} - \exp(\Delta_2^*)\,F_u^{c(3)}\Big\}\Bigg]\,, \end{aligned}$$

and

$$\sum_{i=1}^{R-1} \lambda_{k-1,i} \left(1-F_i^{c(1)}\right) F_i^{c(k)} = 0 \qquad \text{for} \quad k=2,3 \ ,$$

$$F_i^{c(1)}\left(1-F_i^{c(k)}\right) = \exp(\Delta^*_{k-1}) \left(1-F_i^{c(1)}\right) F_i^{c(k)} \qquad \text{for} \quad i=1,...,R-1; \quad k=2,3 \ ,$$

where $m_{ijt} = n p_{ijt}$ and $I_{(\cdot)}$ is the indicator function. Using the Newton–Raphson method, we can solve the equations with respect to $\{p_{ijt}\}$, $\{\lambda_{1i}\}$, $\{\lambda_{2i}\}$, Δ^*_1 and Δ^*_2. Therefore, we can obtain the MLEs of $\{m_{ijt}\}$, Δ^*_1 and Δ^*_2 under the CML model.

ACKNOWLEDGMENTS

The authors would like to express their sincere thanks to the referee for the helpful comments.

REFERENCES

[1] Agresti, A. (1984). *Analysis of Ordinal Categorical Data*, Wiley, New York.

[2] Agresti, A. (2002). *Categorical Data Analysis*, 2nd edition, Wiley, New York.

[3] Bhapkar, V.P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data, *Journal of the American Statistical Association*, **61**, 228–235.

[4] Bishop, Y.M.M.; Fienberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press, Cambridge, Massachusetts.

[5] Lang, J.B. and Agresti, A. (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses, *Journal of the American Statistical Association*, **89**, 625–632.

[6] Miyamoto, N.; Niibe, K. and Tomizawa, S. (2005). Decompositions of marginal homogeneity model using cumulative logistic models for square contingency tables with ordered categories, *Austrian Journal of Statistics*, **34**, 361–373.

[7] Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification, *Biometrika*, **42**, 412–416.

[8] Tomizawa, S. (1991). Decomposing the marginal homogeneity model into two models for square contingency tables with ordered categories, *Calcutta Statistical Association Bulletin*, **41**, 201–207.

[9] Tomizawa, S. (1993). A decomposition of the marginal homogeneity model for square contingency tables with ordered categories, *Calcutta Statistical Association Bulletin*, **43**, 123–125.

[10] Tomizawa, S. (1998). A decomposition of the marginal homogeneity model into three models for square contingency tables with ordered categories, *Sankhyā, Series B*, **60**, 293–300.

REVSTAT – Statistical Journal
Volume 5, Number 2, June 2007, 177–207

IMPROVING SECOND ORDER REDUCED BIAS EXTREME VALUE INDEX ESTIMATION

Authors: M. Ivette Gomes
– Universidade de Lisboa, D.E.I.O. and C.E.A.U.L., Portugal
ivette.gomes@fc.ul.pt

M. João Martins
– D.M., I.S.A., Universidade Técnica de Lisboa, Portugal
mjmartins@isa.utl.pt

Manuela Neves
– D.M., I.S.A., Universidade Técnica de Lisboa, Portugal
manela@isa.utl.pt

Received: November 2006 Accepted: February 2007

Abstract:

- Classical extreme value index estimators are known to be quite sensitive to the number k of top order statistics used in the estimation. The recently developed second order reduced-bias estimators show much less sensitivity to changes in k. Here, we are interested in the improvement of the performance of reduced-bias extreme value index estimators based on an exponential second order regression model applied to the scaled log-spacings of the top k order statistics. In order to achieve that improvement, the estimation of a "scale" and a "shape" second order parameters in the bias is performed at a level k_1 of a larger order than that of the level k at which we compute the extreme value index estimators. This enables us to keep the asymptotic variance of the new estimators of a positive extreme value index γ equal to the asymptotic variance of the Hill estimator, the maximum likelihood estimator of γ, under a strict Pareto model. These new estimators are then alternatives to the classical estimators, not only around optimal and/or large levels k, but for other levels too. To enhance the interesting performance of this type of estimators, we also consider the estimation of the "scale" second order parameter only, at the same level k used for the extreme value index estimation. The asymptotic distributional properties of the proposed class of γ-estimators are derived and the estimators are compared with other similar alternative estimators of γ recently introduced in the literature, not only asymptotically, but also for finite samples through Monte Carlo techniques. Case-studies in the fields of finance and insurance will illustrate the performance of the new second order reduced-bias extreme value index estimators.

Key-Words:

- *statistics of extremes; semi-parametric estimation; bias estimation; heavy tails; maximum likelihood.*

AMS Subject Classification:

- 62G32, 62H12; 65C05.

1. INTRODUCTION AND MOTIVATION FOR THE NEW CLASS OF EXTREME VALUE INDEX ESTIMATORS

Examples of heavy-tailed models are quite common in the most diversified fields. We may find them in computer science, telecommunication networks, insurance, economics and finance, among other areas of application. In the area of *extreme value theory*, a model F is said to be *heavy-tailed* whenever the *tail function*, $\overline{F} := 1 - F$, is a regularly varying function with a negative index of regular variation equal to $-1/\gamma$, $\gamma > 0$, denoted $\overline{F} \in RV_{-1/\gamma}$, where the notation RV_α stands for the class of *regularly varying* functions at infinity with an *index of regular variation* equal to α, i.e., positive measurable functions g such that $\lim_{t\to\infty} g(tx)/g(t) = x^\alpha$, for all $x > 0$. Equivalently, the quantile function $U(t) = F^{\leftarrow}(1-1/t),\ t \geq 1$, with $F^{\leftarrow}(x) = \inf\{y\colon F(y) \geq x\}$, is of regular variation with index γ, i.e.,

$$(1.1) \qquad F \text{ is heavy-tailed} \quad \Longleftrightarrow \quad \overline{F} \in RV_{-1/\gamma} \quad \Longleftrightarrow \quad U \in RV_\gamma ,$$

for some $\gamma > 0$. Then, we are in the domain of attraction for maxima of an *extreme value* distribution function (d.f.),

$$EV_\gamma(x) \;=\; \begin{cases} \exp\bigl(-(1+\gamma x)^{-1/\gamma}\bigr), & 1+\gamma x \geq 0 \qquad \text{if } \gamma \neq 0\,, \\ \exp\bigl(-\exp(-x)\bigr), & x \in \mathbb{R} \qquad\qquad\ \text{if } \gamma = 0\,, \end{cases}$$

but with $\gamma > 0$, and we write $F \in \mathcal{D}_{\mathcal{M}}(EV_{\gamma>0})$. The parameter γ is the *extreme value index*, one of the primary parameters of extreme or even rare events.

The *second order parameter* ρ rules the rate of convergence in the first order condition (1.1), let us say the rate of convergence towards zero of $\ln U(tx) - \ln U(t) - \gamma \ln x$, and is the non-positive parameter appearing in the limiting relation

$$(1.2) \qquad \lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} \;=\; \frac{x^\rho - 1}{\rho}\ ,$$

which we assume to hold for all $x > 0$, and where $|A(t)|$ must then be of regular variation with index ρ (Geluk and de Haan, 1987). We shall assume everywhere that $\rho < 0$. The second order condition (1.2) has been widely accepted as an appropriate condition to specify the tail of a Pareto-type distribution in a semi-parametric way, and it holds for most common Pareto-type models.

Remark 1.1. For Hall–Welsh class of Pareto-type models (Hall and Welsh, 1985), i.e., models such that, with $C > 0$, $D_1 \neq 0$ and $\rho < 0$,

$$(1.3) \qquad U(t) \;=\; C\,t^\gamma \bigl(1 + D_1 t^\rho + o(t^\rho)\bigr), \qquad \text{as } \ t \to \infty\ ,$$

condition (1.2) holds and we may choose $A(t) = \rho\, D_1 t^\rho$.

Here, although not going into a general third order framework, as the one found in Gomes et al. (2002) and Fraga Alves et al. (2003), in papers on the estimation of ρ, as well as in Gomes et al. (2004a), in a paper on the estimation of a positive extreme value index γ, we shall further specify the term $o(t^\rho)$ in Hall–Welsh class of models, and, for some particular details in the paper, we shall assume to be working with a Pareto-type class of models with a quantile function

$$U(t) \;=\; C\,t^\gamma \left(1 + D_1 t^\rho + D_2\, t^{2\rho} + o(t^{2\rho})\right) , \tag{1.4}$$

as $t \to \infty$, with $C > 0$, $D_1, D_2 \neq 0$, $\rho < 0$. Consequently, we may obviously choose, in (1.2),

$$A(t) \;=\; \rho\, D_1\, t^\rho \;=:\; \gamma\, \beta\, t^\rho , \qquad \beta \neq 0, \quad \rho < 0 , \tag{1.5}$$

and, with

$$B(t) \;=\; (2\,D_2/D_1 - D_1)\, t^\rho \;=:\; \beta'\, t^\rho \;=\; \frac{\beta' A(t)}{\beta\, \gamma} , \tag{1.6}$$

we may write

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

The consideration of models in (1.4) enables us to get full information on the asymptotic bias of the so-called second-order reduced-bias extreme value index estimators, the type of estimators under consideration in this paper.

Remark 1.2. Most common heavy-tailed d.f.'s, like the Fréchet, the Generalized Pareto (GP), the Burr and the Student's t belong to the class of models in (1.4), and consequently, to the class of models in (1.3) or, to the more general class of parents satisfying (1.2).

For intermediate k, i.e., a sequence of integers $k = k_n$, $1 \le k < n$, such that

$$k = k_n \to \infty , \qquad k_n = o(n), \quad \text{as } n \to \infty , \tag{1.7}$$

and with $X_{i:n}$ denoting the i-th ascending order statistic (o.s.), $1 \le i \le n$, associated to an independent, identically distributed (i.i.d.) random sample $(X_1, X_2, ..., X_n)$, we shall consider, as basic statistics, both the log-excesses over the random high level $\ln X_{n-k:n}$, i.e.,

$$V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n} , \qquad 1 \le i \le k < n , \tag{1.8}$$

and the scaled log-spacings,

$$U_i := i\left\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\right\} , \qquad 1 \le i \le k < n . \tag{1.9}$$

We have a strong obvious link between the log-excesses and the scaled log-spacings provided by the equation, $\sum_{i=1}^k V_{ik} = \sum_{i=1}^k U_i$.

It is well known that for intermediate k, and whenever we are working with models in (1.1), the log-excesses V_{ik}, $1 \le i \le k$, are approximately the k o.s.'s from an exponential sample of size k and mean value γ. Also, under the same conditions, the scaled log-spacings U_i, $1 \le i \le k$, are approximately i.i.d. and exponential with mean value γ. Consequently, the Hill estimator of γ (Hill, 1975),

$$H(k) \equiv H_n(k) = \frac{1}{k}\sum_{i=1}^k V_{ik} = \frac{1}{k}\sum_{i=1}^k U_i \,, \tag{1.10}$$

is consistent for the estimation of γ whenever (1.1) holds and k is intermediate, i.e., (1.7) holds. Under the second order framework in (1.2) the asymptotic distributional representation

$$H_n(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + \frac{A(n/k)}{1-\rho}\big(1 + o_p(1)\big) \tag{1.11}$$

holds, where $Z_k^{(1)} = \sqrt{k}\,\big(\sum_{i=1}^k E_i/k - 1\big)$, with $\{E_i\}$ i.i.d. standard exponential random variables (r.v.'s), is an asymptotically standard normal random variable. Consequently, $\sqrt{k}\,(H_n(k) - \gamma)$ converges weakly towards a normal r.v. with variance γ^2 and a non-null mean value equal to $\lambda/(1-\rho)$, whenever $\sqrt{k}\,A(n/k) \to \lambda \ne 0$, finite.

The adequate accommodation of the bias of Hill's estimator has been extensively addressed in recent years by several authors. Beirlant et al. (1999) and Feuerverger and Hall (1999) consider exponential regression techniques, based on the exponential approximations $U_i \approx \gamma\big(1 + b(n/k)\,(k/i)^\rho\big)\,E_i$ and $U_i \approx \gamma\,\exp\big(\beta\,(n/i)^\rho\big)\,E_i$, respectively, $1 \le i \le k$. They then proceed to the joint maximum likelihood (*ML*) estimation of the three unknown parameters or functionals at the same level k. Considering also the scaled log-spacings U_i in (1.9) to be approximately exponential with mean value $\mu_i = \gamma\,\exp\big(\beta(n/i)^\rho\big)$, $1 \le i \le k$, $\beta \ne 0$, Gomes and Martins (2002) advance with the so-called "external" estimation of the second order parameter ρ, i.e., an adequate estimation of ρ at a level k_1 higher than the level k used for the extreme value index estimation, together with a first order approximation for the *ML* estimator of β. They then obtain "quasi-*ML*" explicit estimators of γ and β, both computed at the same level k, and through that "external" estimation of ρ, are then able to reduce the asymptotic variance of the extreme value index estimator proposed, comparatively to the asymptotic variance of the extreme value index estimator in Feuerverger and Hall (1999), where the three parameters γ, β and ρ are estimated at the same level k. With the notation

$$d_k(\alpha) = \frac{1}{k}\sum_{i=1}^k \Big(\frac{i}{k}\Big)^{\alpha-1}, \qquad D_k(\alpha) = \frac{1}{k}\sum_{i=1}^k \Big(\frac{i}{k}\Big)^{\alpha-1} U_i \,, \tag{1.12}$$

for any real $\alpha \geq 1$ [$D_k(1) \equiv H(k)$ in (1.10)], and with $\hat\rho$ any consistent estimator of ρ, such estimators are

$$\widehat{\gamma}_n^{ML}(k) \;=\; H(k) - \hat\beta(k;\hat\rho)\,\Big(\frac{n}{k}\Big)^{\hat\rho} D_k(1-\hat\rho) \tag{1.13}$$

and

$$\hat\beta(k;\hat\rho) \;:=\; \Big(\frac{k}{n}\Big)^{\hat\rho}\,\frac{d_k(1-\hat\rho)\times D_k(1)-D_k(1-\hat\rho)}{d_k(1-\hat\rho)\times D_k(1-\hat\rho)-D_k(1-2\,\hat\rho)}\;, \tag{1.14}$$

for γ and β, respectively. This means that β, in (1.5), which is also a second order parameter, is estimated at the same level k at which the γ-estimation is performed, being $\hat\beta(k;\hat\rho)$ — not consistent for the estimation of β whenever $\sqrt{k}\,A(n/k)\to\lambda$, finite, but consistent for models in (1.2) and intermediate k such that $\sqrt{k}\,A(n/k)\to\infty$ (Gomes and Martins, 2002), — plugged in the extreme value index estimator in (1.13). In all the above mentioned papers, authors have been led to the now called "classical" second order reduced-bias extreme value index estimators with an asymptotic variance larger or equal to $\gamma^2\big((1-\rho)/\rho\big)^2$, the minimal asymptotic variance of an asymptotically unbiased estimator in Drees class of functionals (Drees, 1998).

We here propose an "external" estimation of both β and ρ, through $\hat\beta$ and $\hat\rho$, respectively, both using a number of top o.s.'s, k_1, larger than the number of top o.s.'s, k, used for the extreme value index estimation. We shall thus consider the estimator

$$ML_{\hat\beta,\hat\rho}(k) \;:=\; H(k) - \hat\beta\,\Big(\frac{n}{k}\Big)^{\hat\rho} D_k(1-\hat\rho)\;, \tag{1.15}$$

for adequate consistent estimators $\hat\beta$ and $\hat\rho$ of the second order parameters β and ρ, respectively, to be specified in subsection 3.3 of this paper. Additionally, we shall also deal with the estimator

$$\overline{ML}_{\hat\beta,\hat\rho}(k) \;=\; \frac{1}{k}\sum_{i=1}^{k} U_i\,\exp\!\big(-\hat\beta(n/i)^{\hat\rho}\big)\;, \tag{1.16}$$

the estimator directly derived from the likelihood equation for γ with β and ρ fixed and based upon the exponential approximation, $U_i \approx \gamma\exp\!\big(\beta(n/i)^{\rho}\big)\,E_i$, $1\leq i\leq k$. Doing this, we are able to reduce the bias without increasing the asymptotic variance, which is kept at the value γ^2, the asymptotic variance of Hill's estimator. The estimators are thus better than the Hill estimator for all k.

Remark 1.3. If, in (1.15), we estimate β at the same level k used for the estimation of γ, we may be led to $\widehat{\gamma}_n^{ML}(k)$ in (1.13). Indeed, $\widehat{\gamma}_n^{ML}(k) = ML_{\hat\beta(k;\hat\rho),\hat\rho}(k)$, with $\hat\beta(k;\hat\rho)$ defined in (1.14).

Remark 1.4. The ML estimator in (1.15) may be obtained from the estimator in (1.16) through the use of the first order approximation, $\big\{1-\hat\beta(n/i)^{\hat\rho}\big\}$, for the exponential weight, $e^{-\hat\beta(n/i)^{\hat\rho}}$, of the scaled log-spacing U_i, $1\leq i\leq k$.

Remark 1.5. The estimators in (1.15) and (1.16) have been inspired in the recent papers of Gomes et al. (2004*b*) and Caeiro et al. (2005). These authors consider, in different ways, the joint external estimation of both the "scale" and the "shape" parameters in the A function in (1.2), parameterized as in (1.5), being able to reduce the bias without increasing the asymptotic variance, which is kept at the value γ^2, the asymptotic variance of Hill's estimator. Those estimators are also going to be considered here for comparison with the new estimators in (1.15) and (1.16). The reduced-bias extreme value index estimator in Gomes et al. (2004*b*) is based on a linear combination of the log-excesses V_{ik} in (1.8), and is given by

$$(1.17)\qquad WH_{\hat\beta,\hat\rho}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{-\hat\beta (n/k)^{\hat\rho}\,\psi_{\hat\rho}(i/k)}\, V_{ik}\;, \qquad \psi_\rho(x) = -\frac{x^{-\rho}-1}{\rho \ln x}\;,$$

with the notation *WH* standing for *Weighted Hill* estimator. Caeiro et al. (2005) consider the estimator

$$(1.18)\qquad \overline{H}_{\hat\beta,\hat\rho}(k) := H(k)\left(1-\frac{\hat\beta}{1-\hat\rho}\left(\frac{n}{k}\right)^{\hat\rho}\right)\;,$$

where the dominant component of the bias of Hill's estimator $H(k)$ in (1.10), given by $A(n/k)/(1-\rho) = \beta\,\gamma(n/k)^\rho/(1-\rho)$, is thus estimated through $H(k)\,\hat\beta(n/k)^{\hat\rho}/(1-\hat\rho)$, and directly removed from Hill's classical extreme value index estimator. As before, both in (1.17) and (1.18), $\hat\beta$ and $\hat\rho$ need to be adequate consistent estimators of the second order parameters β and ρ, respectively, so that the new estimators are better than the Hill estimator for all k.

In section 2 of this paper, and assuming first that only γ is unknown, we shall state a theorem that provides an obvious technical motivation for the estimators in (1.15) and (1.16). Next, in section 3, we consider the derivation of the asymptotic behavior of the classes of estimators in (1.15) and (1.16), for an appropriate estimation of β and ρ at a level k_1 larger than the value k used for the extreme value index estimation. We also do that only with the estimation of ρ, estimating β at the same level k used for the extreme value index estimation. In this same section, we finally briefly review the estimation of the two second order parameters β and ρ. In section 4, using simulation techniques, we exhibit the performance of the *ML* estimator in (1.15) and the $\overline{ML}$ estimator in (1.16), comparatively to the other *"Unbiased Hill"* (*UH*) estimators, *WH* and $\overline{H}$, in (1.17) and (1.18), respectively, to the classical Hill estimator H in (1.10) and to the "asymptotically unbiased" estimator $\widehat\gamma_n^{ML}(k)$ in (1.13), studied in Gomes and Martins (2002), or equivalently, $ML_{\hat\beta(k;\hat\rho),\hat\rho}$, with $ML_{\hat\beta,\hat\rho}$ the estimator in (1.15). Section 5 is devoted to the illustration of the behavior of these estimators for the Daily Log-Returns of the *Euro* against the *UK* Pound and automobile claims gathered from several European insurance companies co-operating with the same re-insurer (Secura Belgian Re).

2. ASYMPTOTIC BEHAVIOR OF THE ESTIMATORS (ONLY γ IS UNKNOWN)

For real values $\alpha \geq 1$, and denoting again $\{E_i\}$ a sequence of i.i.d. standard exponential r.v.'s, let us introduce the following notation:

$$Z_k^{(\alpha)} = \sqrt{(2\alpha-1)\,k}\,\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1} E_i - \frac{1}{\alpha}\right). \tag{2.1}$$

With the same kind of reasoning as in Gomes et al. (2005a), we state:

Lemma 2.1. *Under the second order framework in* (1.2)*, for intermediate k-sequences, i.e., whenever* (1.7) *holds, and with U_i given in* (1.9)*, we may guarantee that, for any real $\alpha \geq 1$, and $D_k(\alpha)$ given in* (1.12)*,*

$$D_k(\alpha) \overset{d}{=} \frac{\gamma}{\alpha} + \frac{\gamma\, Z_k^{(\alpha)}}{\sqrt{(2\alpha-1)\,k}} + \frac{A(n/k)}{\alpha-\rho}\,\big(1+o_p(1)\big)\ ,$$

where $Z_k^{(\alpha)}$, given in (2.1)*, is an asymptotically standard normal random variable. If we further assume to be working with models in* (1.4)*, and with the same notation as before, we may write for any $\alpha, \beta \geq 1$, $\alpha \neq \beta$, the joint distribution*

$$\begin{aligned}\big(D_k(\alpha), D_k(\beta)\big) \overset{d}{=}\ & \left(\frac{\gamma}{\alpha}, \frac{\gamma}{\beta}\right) + \frac{\gamma}{\sqrt{k}}\left(\frac{Z_k^{(\alpha)}}{\sqrt{(2\alpha-1)}}, \frac{Z_k^{(\beta)}}{\sqrt{(2\beta-1)}}\right) \\ & + A(n/k)\left(\frac{1}{\alpha-\rho}, \frac{1}{\beta-\rho}\right) + \frac{\beta'\,A^2(n/k)}{\beta\,\gamma}\left(\frac{1}{\alpha-2\rho}, \frac{1}{\beta-2\rho}\right) \\ & + O_p\left(\frac{A(n/k)}{\sqrt{k}}\right) + o_p\big(A^2(n/k)\big)\ ,\end{aligned} \tag{2.2}$$

with β and β' given in (1.5) *and* (1.6)*, respectively.*

Let us assume that only the extreme value index parameter γ is unknown, and generally denote $\widetilde{ML}$ either ML or $\overline{ML}$. This case obviously refers to a situation which is rarely encountered in practice, but reveals the potential of the classes of estimators in (1.15) and (1.16).

2.1. Known β and ρ

We may state:

Theorem 2.1. *Under the second order framework in* (1.2)*, further assuming that* $A(t)$ *may be chosen as in* (1.5)*, and for levels* k *such that* (1.7) *holds, we get asymptotic distributional representations of the type*

$$\widetilde{ML}_{\beta,\rho}(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + o_p\big(A(n/k)\big) \;, \tag{2.3}$$

where $Z_k^{(1)}$ *is the asymptotically standard normal r.v. in* (2.1) *for* $\alpha = 1$. *Consequently,* $\sqrt{k}\,\big(\widetilde{ML}_{\beta,\rho}(k) - \gamma\big)$ *is asymptotically normal with variance equal to* γ^2, *and with a null mean value not only when* $\sqrt{k}\,A(n/k) \longrightarrow 0$, *but also when* $\sqrt{k}\,A(n/k) \longrightarrow \lambda \neq 0$, *finite, as* $n \to \infty$.

For models in (1.4)*, we may further specify the term* $o_p\big(A(n/k)\big)$*, writing*

$$ML_{\beta,\rho}(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + \frac{(\beta'-\beta)\,A^2(n/k)}{\beta\,\gamma\,(1-2\rho)}\,\big(1+o_p(1)\big) \;, \tag{2.4}$$

$$\overline{ML}_{\beta,\rho}(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + \frac{(2\beta'-\beta)\,A^2(n/k)}{2\,\beta\,\gamma\,(1-2\rho)}\,\big(1+o_p(1)\big) \;, \tag{2.5}$$

with β *and* β' *given in* (1.5) *and* (1.6)*, respectively. Consequently, even if* $\sqrt{k}\,A(n/k) \to \infty$, *with* $\sqrt{k}\,A^2(n/k) \to \lambda_A$, *finite,* $\sqrt{k}\,(ML_{\beta,\rho}(k) - \gamma)$ *and* $\sqrt{k}\,(\overline{ML}_{\beta,\rho}(k) - \gamma)$ *are asymptotically normal with variance equal to* γ^2 *and asymptotic bias equal to*

$$b_{ML} = \frac{(\beta'-\beta)\,\lambda_A}{\beta\,\gamma\,(1-2\rho)} \qquad \text{and} \qquad b_{\overline{ML}} = \frac{(2\beta'-\beta)\,\lambda_A}{2\,\beta\,\gamma\,(1-2\rho)} \;, \tag{2.6}$$

respectively.

Proof: If all parameters are known, apart from the extreme value index γ, we get directly from Lemma 2.1,

$$\begin{aligned} ML_{\beta,\rho}(k) \;&:=\; D_k(1) - \beta\Big(\frac{n}{k}\Big)^{\rho} D_k(1-\rho) \\ &\stackrel{d}{=}\; \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + \frac{A(n/k)}{1-\rho} \\ &\quad\; - \frac{A(n/k)}{\gamma}\left(\frac{\gamma}{1-\rho} + \frac{\gamma}{\sqrt{(1-2\rho)k}}\, Z_k^{(1-\rho)} + \frac{A(n/k)}{1-2\rho}\,\big(1+o_p(1)\big)\right) \\ &\stackrel{d}{=}\; \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + o_p\big(A(n/k)\big) \;. \end{aligned}$$

Similarly, since we may write

$$\begin{aligned} \overline{ML}_{\beta,\rho}(k) \;&=\; ML_{\beta,\rho}(k) + \frac{A^2(n/k)}{2\,\gamma^2}\, D_k(1-2\rho)\,\big(1+o_p(1)\big) \\ &=\; ML_{\beta,\rho}(k) + \frac{A^2(n/k)}{2\,\gamma\,(1-2\rho)}\,\big(1+o_p(1)\big) \;, \end{aligned} \tag{2.7}$$

(2.3) holds for $\overline{ML}$ as well. For models in (1.4), and directly from (2.2), we get

$$\begin{aligned} ML_{\beta,\rho}(k) \overset{d}{=} & \gamma + \frac{\gamma}{\sqrt{k}} Z_k^{(1)} + \frac{A(n/k)}{1-\rho} + \frac{\beta' A^2(n/k)}{\beta\gamma(1-2\rho)} \big(1+o_p(1)\big) + O_p\left(\frac{A(n/k)}{\sqrt{k}}\right) \\ & - \frac{A(n/k)}{\gamma}\left(\frac{\gamma}{1-\rho} + \frac{\gamma}{\sqrt{(1-2\rho)\,k}} Z_k^{(1-\rho)} + \frac{A(n/k)}{1-2\rho}\big(1+o_p(1)\big)\right). \end{aligned}$$

Working this expression, we finally obtain

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

i.e., (2.4) holds. Also directly from (2.4) and (2.7), (2.5) follows. Note that since $\sqrt{k}\, O_p\big(A(n/k)/\sqrt{k}\big) = O_p\big(A(n/k)\big) \to 0$, the summand $O_p\big(A(n/k)/\sqrt{k}\big)$ is totally irrelevant for the asymptotic bias in (2.6), that follows straightforwardly from the above obtained distributional representations. □

Remark 2.1. We know that the asymptotic variances of ML and $\overline{ML}$ are the same. Since $\lambda_A \geq 0$, $b_{\overline{ML}} = b_{ML} + \lambda_A/\big(2\gamma(1-2\rho)\big) \geq b_{ML}$. We may thus say that, asymptotically, the *ML*-statistic is expected to exhibit a better performance than $\overline{ML}$, provided the bias are both positive. Things work the other way round if the bias are both negative, i.e., the sample paths of $\overline{ML}$ are expected to be in average above the ones of ML.

Remark 2.2. For the Burr d.f. $F(x) = 1-(1+x^{-\rho/\gamma})^{1/\rho}$, $x \geq 0$, we have $U(t) = t^\gamma(1-t^\rho)^{-\gamma/\rho} = t^\gamma\big(1+\gamma\, t^\rho/\rho + \gamma(\gamma+\rho)\, t^{2\rho}/(2\rho^2) + o(t^{2\rho})\big)$, for $t \geq 1$. Consequently, (1.4) holds with $D_1 = \gamma/\rho$, $D_2 = \gamma(\gamma+\rho)/(2\rho^2)$, $\beta' = \beta = 1$ and $b_{ML} = 0$. A similar result holds for the *GP* d.f. $F(x) = 1-(1+\gamma x)^{-1/\gamma}$, $x \geq 0$. For this d.f., $U(t) = (t^\gamma - 1)/\gamma$, and (1.4) holds with $\rho = -\gamma$, $D_1 = -1$ and $D_2 = 0$. Hence $\beta = \beta' = 1$ and $b_{ML} = 0$. We thus expect a better performance of ML, comparatively to $\overline{ML}$, WH and $\overline{H}$ whenever the model underlying the data is close to Burr or to *GP* models, a situation that happens often in practice, and that is another point in favour of the *ML*-statistic.

2.2. Known ρ

We may state the following:

Theorem 2.2. *For models in* (1.4), *if* $k = k_n$ *is a sequence of intermediate integers, i.e.,* (1.7) *holds, and if* $\sqrt{k}\, A(n/k) \to \infty$, *with* $\sqrt{k}\, A^2(n/k)$ *converging towards* λ_A, *finite, as* $n \to \infty$, *then, with* $\hat\beta(k;\hat\rho)$, $ML_{\hat\beta,\hat\rho}(k)$ *and* $\overline{ML}_{\hat\beta,\hat\rho}(k)$ *given in* (1.14), (1.15) *and* (1.16), *respectively, the asymptotic variance of both* $ML^*(k) =$

$ML_{\hat\beta(k;\rho),\rho}(k)$ and $\overline{ML}^*(k) = \overline{ML}_{\hat\beta(k;\rho),\rho}(k)$ is equal to $\big(\gamma\,(1-\rho)/\rho\big)^2$, being their asymptotic bias given by

$$(2.8)\qquad b^*_{ML} = \frac{(\beta-\beta')\,(1-\rho)\,\lambda_A}{\beta\,\gamma\,(1-2\rho)\,(1-3\rho)} \quad \text{and} \quad b^*_{\overline{ML}} = \frac{\Big(\beta\,(3-5\rho) - 2\,\beta'(1-\rho)\Big)\,\lambda_A}{2\,\beta\,\gamma\,(1-2\rho)\,(1-3\rho)}\ ,$$

respectively, again with β and β' given in (1.5) *and* (1.6), *respectively.*

Proof: Following the steps in Gomes and Martins (2002), but working now with models in (1.4) and the distributional representation (2.2), we may write:

$$\begin{aligned} ML^*(k) &= H(k) - \frac{D_k(1-\rho)\,\Big\{D_k(1)\,\big(1+o(1)\big) - (1-\rho)\,D_k(1-\rho)\Big\}}{D_k(1-\rho)\,\big(1+o(1)\big) - (1-\rho)\,D_k(1-2\rho)} \\ &=: H(k) - \frac{\varphi_k(\rho)}{\psi_k(\rho)}\ , \end{aligned}$$

with $D_k(\alpha)$ given in (1.12). Directly from (2.2), we get

$$\frac{1}{\psi_k(\rho)} = -\frac{(1-\rho)\,(1-2\rho)}{\gamma\,\rho^2}\left(1 - \left\{\frac{2\,(1-\rho)\,A(n/k)}{\gamma\,(1-3\rho)} + O_p\Big(\frac{1}{\sqrt{k}}\Big)\right\}\,\big(1+o_p(1)\big)\right)$$

and, under the conditions on k imposed,

$$\begin{aligned} \varphi_k(\rho) &= \frac{\gamma^2}{\sqrt{k}}\left(\frac{Z_k^{(1)}}{1-\rho} - \frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}}\right) - \frac{\gamma\,\rho^2\,A(n/k)}{(1-\rho)^2\,(1-2\rho)} \\ &\quad - \frac{\rho^2\,A^2(n/k)}{(1-\rho)\,(1-2\rho)}\left(\frac{2\,\beta'}{\beta\,(1-3\rho)} + \frac{1}{1-2\rho}\right)\big(1+o_p(1)\big)\ . \end{aligned}$$

Consequently,

$$\begin{aligned} \frac{\varphi_k(\rho)}{\psi_k(\rho)} &= -\frac{\gamma}{\rho^2\,\sqrt{k}}\Big((1-2\rho)\,Z_k^{(1)} - (1-\rho)\,\sqrt{1-2\rho}\;Z_k^{(1-\rho)}\Big) + \frac{A(n/k)}{1-\rho} \\ &\quad + \frac{A^2(n/k)}{\gamma}\left(\frac{2\,(\beta'-\beta)}{\beta\,(1-3\rho)} + \frac{1}{1-2\rho}\right)\big(1+o_p(1)\big)\ . \end{aligned}$$

Then, with

$$\overline{Z}_k = \left(\frac{1-\rho}{\rho}\right)^2 Z_k^{(1)} - \left(\frac{(1-\rho)\,\sqrt{1-2\rho}}{\rho^2}\right) Z_k^{(1-\rho)}\ ,$$

$$ML^*(k) = ML_{\hat\beta(k;\rho),\rho}(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\,\overline{Z}_k - \frac{(\beta'-\beta)\,(1-\rho)\,A^2(n/k)}{\beta\,\gamma\,(1-2\rho)\,(1-3\rho)}\,\big(1+o_p(1)\big)\ ,$$

and the result in (2.8) follows for $ML^*(k)$. Also, since the asymptotic covariance between $Z_k^{(1)}$ and $Z_k^{(1-\rho)}$ is given by $\sqrt{1-2\rho}/(1-\rho)$, the asymptotic variance of $\overline{Z}_k$ is given by

$$\left(\frac{1-\rho}{\rho}\right)^4 + \frac{(1-\rho)^2\,(1-2\rho)}{\rho^4} - \frac{2\,(1-\rho)^3\,\sqrt{1-2\rho}}{\rho^4} \times \frac{\sqrt{1-2\rho}}{1-\rho} = \left(\frac{1-\rho}{\rho}\right)^2.$$

Hence, the asymptotic variance $\gamma^2\{(1-\rho)/\rho\}^2$, stated in the theorem. If we consider $\overline{ML}_{\hat\beta(k;\rho),\rho}(k)$, since $\sqrt{k}\,A(n/k) \to \infty$, $\hat\beta(k;\rho)$ converges in probability towards β and a result similar to (2.7) holds, i.e.,

$$\overline{ML}^*(k) \;=\; \overline{ML}_{\hat\beta(k;\rho),\rho}(k) \;=\; ML_{\hat\beta(k;\rho),\rho}(k) + \frac{A^2(n/k)}{2\gamma\,(1-2\rho)}\,\big(1+o_p(1)\big)\ .$$

The result in the theorem follows thus straightforwardly. □

Remark 2.3. For models in (1.4) and $\lambda_A \neq 0$ in Theorem 2.2, $b^*_{ML} = 0$ if and only if $\beta = \beta'$. Again, this holds for Burr and *GP* underlying models.

Remark 2.4. When we look at Theorems 2.1 and 2.2, we see that, for (β,ρ) known, despite the increasing in the asymptotic variance, $(b_{ML}/b^*_{ML})^2 = \big((1-3\rho)/(1-\rho)\big)^2$ is an increasing function of $|\rho|$, always greater than one, for $\rho < 0$, i.e., there is here again a compromise between bias and variance.

2.3. Asymptotic comparison at optimal levels

We now proceed to an asymptotic comparison of *ML* and *ML** at their optimal levels in the lines of de Haan and Peng (1998), Gomes and Martins (2001) and Gomes et al. (2005b, 2006), among others, but now for second order reduced-bias estimators. Suppose $\widehat\gamma^\bullet_n(k)$ is a general semi-parametric estimator of the extreme value index γ, for which the distributional representation

$$\widehat\gamma^\bullet_n(k) \;\overset{d}{=}\; \gamma + \frac{\sigma_\bullet}{\sqrt{k}}\,Z^\bullet_n + b_\bullet\,A^2(n/k) + o_p\big(A^2(n/k)\big) \tag{2.9}$$

holds for any intermediate k, and where $Z^\bullet_n$ is an asymptotically standard normal random variable. Then we have

$$\sqrt{k}\,\big[\widehat\gamma^\bullet_n(k)-\gamma\big] \;\overset{d}{\to}\; N(\lambda_A b_\bullet, \sigma^2_\bullet), \qquad \text{as } \; n\to\infty\ ,$$

provided k is such that $\sqrt{k}\,A^2(n/k) \to \lambda_A$, finite, as $n \to \infty$. In this situation we may write $Bias_\infty[\widehat\gamma^\bullet_n(k)] := b_\bullet\,A^2(n/k)$ and $Var_\infty[\widehat\gamma^\bullet_n(k)] := \sigma^2_\bullet/k$. The so-called Asymptotic Mean Squared Error (*AMSE*) is then given by

$$AMSE\,[\widehat\gamma^\bullet_n(k)] \;:=\; \frac{\sigma^2_\bullet}{k} + b^2_\bullet\,A^4(n/k)\ .$$

Using regular variation theory (Bingham et al., 1987), it may be proved that, whenever $b_\bullet \neq 0$, there exists a function $\varphi(n)$, dependent only on the underlying model, and not on the estimator, such that

$$\lim_{n\to\infty} \varphi(n)\,AMSE\,[\widehat\gamma^\bullet_{n0}] \;=\; C(\rho)\,(\sigma^2_\bullet)^{-\frac{4\rho}{1-4\rho}}\,(b^2_\bullet)^{\frac{1}{1-4\rho}} \;=:\; LMSE\,[\widehat\gamma^\bullet_{n0}]\ , \tag{2.10}$$

where $\widehat{\gamma}^{\bullet}_{n0} := \widehat{\gamma}^{\bullet}_n(k^{\bullet}_0(n))$, $k^{\bullet}_0(n) := \arg\inf_k AMSE\,[\widehat{\gamma}^{\bullet}_n(k)]$, is the estimator $\widehat{\gamma}^{\bullet}_n(k)$ computed at its optimal level, the level where its $AMSE$ is minimum.

It is then sensible to consider the usual:

Definition 2.1. Given two second order reduced-bias estimators, $\widehat{\gamma}^{(1)}_n(k)$ and $\widehat{\gamma}^{(2)}_n(k)$, for which distributional representations of the type (2.9) hold, with constants (σ_1, b_1) and (σ_2, b_2), $b_1, b_2 \neq 0$, respectively, both computed at their optimal levels, the Asymptotic Root Efficiency ($AREFF$) of $\widehat{\gamma}^{(1)}_{n0}$ relatively to $\widehat{\gamma}^{(2)}_{n0}$ is $AREFF_{1|2} \equiv AREFF_{\widehat{\gamma}^{(1)}_{n0}|\widehat{\gamma}^{(2)}_{n0}} := \big(LMSE\big[\widehat{\gamma}^{(2)}_{n0}\big]/LMSE\big[\widehat{\gamma}^{(1)}_{n0}\big]\big)^{1/2}$, with $LMSE$ given in (2.10).

Remark 2.5. This measure was devised so that the higher the $AREFF$ measure the better the estimator 1 is, comparatively to the estimator 2.

Proposition 2.1. *For every $\beta \neq \beta'$, if we compare $ML = ML_{\beta,\rho}$ and $ML^* = ML_{\hat{\beta}(k;\rho),\rho}$, we get $AREFF_{ML|ML^*} = (1-\rho)^2 \left((1-3\rho)\,\rho^{-4\rho}\right)^{-1/(1-4\rho)} > 1$ for all $\rho < 0$.*

We may also say that $AREFF_{ML|\overline{ML}} > 1$, for all ρ, β and β'. This indicator depends then not only of ρ, but also of β and β'. This result, together with the result in Proposition 2.1, provides again a clear indication on an overall better performance of the ML estimator, comparatively to $\overline{ML}$ and ML^*.

3. EXTREME VALUE INDEX ESTIMATION BASED ON THE ESTIMATION OF THE SECOND ORDER PARAMETERS β AND ρ

Again for $\alpha \geq 1$, let us further introduce the following extra notations:

$$W^{(\alpha)}_k = (2\alpha-1)\sqrt{(2\alpha-1)\,k/2}\left(\frac{1}{k}\sum_{i=1}^{k}\Big(\frac{i}{k}\Big)^{\alpha-1}\ln\Big(\frac{i}{k}\Big)E_i + \frac{1}{\alpha^2}\right), \tag{3.1}$$

$$D'_k(\alpha) = \frac{d\,D_k(\alpha)}{d\,\alpha} := \frac{1}{k}\sum_{i=1}^{k}\Big(\frac{i}{k}\Big)^{\alpha-1}\ln\Big(\frac{i}{k}\Big)U_i\ , \tag{3.2}$$

with U_i and $D_k(\alpha)$ given in (1.9) and (1.12), respectively.

Again with the same kind of reasoning as in Gomes et al. (2005a), we state:

Lemma 3.1. *Under the second order framework in* (1.2), *for intermediate k-sequences, i.e., whenever* (1.7) *holds, and with* U_i *given in* (1.9), *we may guarantee that, for any real* $\alpha \geq 1$ *and with* $D'_k(\alpha)$ *given in* (3.2),

$$D'_k(\alpha) \stackrel{d}{=} -\frac{\gamma}{\alpha^2} + \frac{\gamma\, W_k^{(\alpha)}}{(2\alpha-1)\sqrt{(2\alpha-1)\,k/2}} - \frac{A(n/k)}{(\alpha-\rho)^2}\big(1+o_p(1)\big)\ , \tag{3.3}$$

where $W_k^{(\alpha)}$, *in* (3.1), *are asymptotically standard normal r.v.'s.*

3.1. Estimation of both second order parameters β and ρ at a lower threshold

Let us assume first that we estimate both β and ρ externally at a level k_1 of a larger order than the level k at which we compute the extreme value index estimator, now assumed to be an intermediate level k such that $\sqrt{k}\,A(n/k) \to \lambda$, finite, as $n \to \infty$, with $A(t)$ the function in (1.2). We may state the following:

Theorem 3.1. *Under the initial conditions of Theorem 2.1, let us consider the class of extreme value index estimators* $\widetilde{ML}_{\hat\beta,\hat\rho}(k)$, *with* $\widetilde{ML}$ *denoting again either the ML estimator in* (1.15) *or the* $\overline{ML}$ *estimator in* (1.16), *with* $\hat\beta$ *and* $\hat\rho$ *consistent for the estimation of* β *and* ρ, *respectively, and such that*

$$(\hat\rho-\rho)\,\ln n = o_p(1), \qquad \text{as } n\to\infty\ . \tag{3.4}$$

Then, $\sqrt{k}\,\big\{\widetilde{ML}_{\hat\beta,\hat\rho}(k)-\gamma\big\}$ *is asymptotically normal with null mean value and variance* $\sigma_1^2=\gamma^2$, *not only when* $\sqrt{k}\,A(n/k)\to 0$, *but also whenever* $\sqrt{k}\,A(n/k)\to \lambda\neq 0$, *finite.*

Proof: With the usual notation $X_n \stackrel{p}{\sim} Y_n$ if and only if X_n/Y_n goes in probability to 1, as $n\to\infty$, we may write

$$\frac{\partial \widetilde{ML}_{\beta,\rho}}{\partial\beta} \stackrel{p}{\sim} -\Big(\frac{n}{k}\Big)^{\rho} D_k(1-\rho) = -\frac{A(n/k)\, D_k(1-\rho)}{\beta\,\gamma} \stackrel{p}{\sim} -\frac{A(n/k)}{\beta(1-\rho)}$$

and

$$\begin{aligned}\frac{\partial \widetilde{ML}_{\beta,\rho}}{\partial\rho} &\stackrel{p}{\sim} -\frac{A(n/k)}{\gamma}\left(\ln\Big(\frac{n}{k}\Big)\, D_k(1-\rho) - D'_k(1-\rho)\right)\\ &\stackrel{p}{\sim} -\frac{A(n/k)}{1-\rho}\left(\ln\Big(\frac{n}{k}\Big) + \frac{1}{1-\rho}\right)\ .\end{aligned}$$

If we estimate consistently ρ and β through the estimators $\hat\beta$ and $\hat\rho$ in the conditions of the theorem, we may use Taylor's expansion series, and we obtain

$$\widetilde{ML}_{\hat\beta,\hat\rho}(k) - \widetilde{ML}_{\beta,\rho}(k) \stackrel{p}{\sim} -\frac{A(n/k)}{1-\rho}\left\{\left(\frac{\hat\beta-\beta}{\beta}\right) + (\hat\rho-\rho)\left(\ln(n/k) + \frac{1}{1-\rho}\right)\right\}\ . \tag{3.5}$$

Consequently, taking into account the conditions in the theorem,

$$\widetilde{ML}_{\hat\beta,\hat\rho}(k) - \widetilde{ML}_{\beta,\rho}(k) = o_p\big(A(n/k)\big) \ .$$

Hence, if $\sqrt{k}\,A(n/k) \to \lambda$, finite, Theorem 2.1 enables us to guarantee the results in the theorem. □

3.2. Estimation of the second order parameter ρ only at a lower threshold

If we consider γ and β estimated at the same level k, we are going to have an increase in the asymptotic variance of our final extreme value index estimators, but we no longer need to assume that condition (3.4) holds. Indeed, as stated in Corollary 2.1 of Theorem 2.1 in Gomes and Martins (2002), for the estimator in (1.13), Theorem 3.2 in Gomes et al. (2004b), for the estimator $WH_{\hat\beta(k;\hat\rho),\hat\rho}$ and Theorem 3.2 in Caeiro et al. (2005), for the estimator $\overline{H}_{\hat\beta(k;\hat\rho),\hat\rho}$, we may state:

Theorem 3.2. (Gomes and Martins, 2002; Gomes et al., 2004b; Caeiro et al., 2005) *Under the second order framework in* (1.2), *if* $k = k_n$ *is a sequence of intermediate integers, i.e.,* (1.7) *holds, and if* $\lim_{n\to\infty} \sqrt{k}\,A(n/k) = \lambda$, *finite, then, with* UH *denoting any of the statistics* ML, $\overline{ML}$, WH *or* $\overline{H}$ *in* (1.15), (1.16), (1.17) *and* (1.18), *respectively,* $\hat\rho$ *any consistent estimator of the second order parameter* ρ, *and* $\hat\beta(k;\hat\rho)$ *the* β*-estimator in* (1.14),

$$\text{(3.6)} \qquad \sqrt{k}\Big(UH_{\hat\beta(k;\hat\rho),\hat\rho}(k) - \gamma\Big) \ \underset{n\to\infty}{\overset{d}{\longrightarrow}} \ Normal\left(0,\ \sigma_2^2 := \gamma^2\Big(\frac{1-\rho}{\rho}\Big)^2\right),$$

i.e., the asymptotic variance of $UH_{\hat\beta(k;\hat\rho),\hat\rho}(k)$ *increases of a factor* $\big((1-\rho)/\rho\big)^2 > 1$ *for every* $\rho < 0$.

Remark 3.1. If we compare Theorem 3.1 and Theorem 3.2, we see that, as expected, the estimation of the two parameters γ and β at the same level k induces an increase in the asymptotic variance of the final γ-estimator of a factor given by $\big((1-\rho)/\rho\big)^2$, greater than 1. The estimation of the three parameters γ, β and ρ at the same level k may still induce an extra increase in the asymptotic variance of the final γ-estimator, as may be seen in Feuerverger and Hall (1999) (where the three parameters are indeed computed at the same level k). These authors get an asymptotic variance ruled by $\sigma_{FH}^2 := \gamma^2\big((1-\rho)/\rho\big)^4$, and we have $\sigma_1 < \sigma_2 < \sigma_{FH}$ for all $\rho < 0$. Consequently, and taking into account asymptotic variances, it seems convenient to estimate both β and ρ "externally", at a level k_1 of a larger order than the level k used for the estimation of the extreme value index γ.

3.3. How to estimate the second order parameters

We now provide some details on the type of second order parameters' estimators we think sensible to use in practice, together with their distributional properties.

3.3.1. The estimation of ρ

Several classes of ρ-estimators are available in the literature. Among them, we mention the ones introduced in Hall and Welsh (1985), Drees and Kaufman (1998), Peng (1998), Gomes et al. (2002) and Fraga Alves et al. (2003). The one working better in practice and for the most common heavy-tailed models, is the one in Fraga Alves et al. (2003). We shall thus consider here particular members of this class of estimators. Under adequate general conditions, and for $\rho < 0$, they are semi-parametric asymptotically normal estimators of ρ, which show highly stable sample paths as functions of k_1, the number of top o.s.'s used, for a wide range of large k_1-values. Such a class of estimators has been first parameterized by a tuning parameter $\tau > 0$, but τ may be more generally considered as a real number (Caeiro and Gomes, 2004), and is defined as

$$\hat\rho(k_1;\tau) \equiv \hat\rho_\tau(k_1) \equiv \hat\rho_n^{(\tau)}(k_1) := -\left|\frac{3\left(T_n^{(\tau)}(k_1)-1\right)}{T_n^{(\tau)}(k_1)-3}\right| , \tag{3.7}$$

where

$$T_n^{(\tau)}(k_1) := \frac{\left(M_n^{(1)}(k_1)\right)^\tau - \left(M_n^{(2)}(k_1)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k_1)/2\right)^{\tau/2} - \left(M_n^{(3)}(k_1)/6\right)^{\tau/3}} , \qquad \tau \in \mathbb{R} ,$$

with the notation $a^{b\tau} = b \ln a$, whenever $\tau = 0$ and with

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left\{\ln\frac{X_{n-i+1:n}}{X_{n-k:n}}\right\}^j, \quad j \geq 1 \qquad \left[M_n^{(1)} \equiv H \quad \text{in } (1.10)\right] .$$

We shall here summarize a particular case of the results proved in Fraga Alves et al. (2003):

Proposition 3.1 (Fraga Alves et al., 2003)**.** *Under the second order framework in* (1.2)*, if k_1 is an intermediate sequence of integers, and if $\sqrt{k_1}\,A(n/k_1) \to \infty$, as $n \to \infty$, the statistics $\hat\rho_n^{(\tau)}(k_1)$ in* (3.7) *converge in probability towards ρ, as $n \to \infty$, for any real τ. Moreover, for models in* (1.4)*, if we further assume* *that* $\sqrt{k_1}\, A^2(n/k_1) \longrightarrow \lambda_{A_1}$, *finite,* $\hat\rho_\tau(k_1) \equiv \hat\rho_n^{(\tau)}(k_1)$ *is asymptotically normal with a bias proportional to* λ_{A_1}, *and* $\{\hat\rho_\tau(k_1) - \rho\} = O_p\big(1/(\sqrt{k_1}\, A(n/k_1))\big)$. *If* $\sqrt{k_1}\, A^2(n/k_1) \to \infty$, $\{\hat\rho_\tau(k_1) - \rho\} = O_p\big(A(n/k_1)\big)$.

Remark 3.2. Note that if we choose for the estimation of ρ a level k_1 under the conditions that assure, in Proposition 3.1, asymptotic normality with a non-null bias, we may guarantee that $k_1 = O\big(n^{-4\rho/(1-4\rho)}\big)$ and consequently $\sqrt{k_1}\, A(n/k_1) = O\big(n^{-\rho/(1-4\rho)}\big)$. Hence, $\hat\rho_\tau(k_1) - \rho = O_p\big(1/(\sqrt{k_1}\, A(n/k_1))\big) = O_p\big(n^{\rho/(1-4\rho)}\big) = o_p(1/\ln n)$ provided that $\rho < 0$, i.e., (3.4) holds whenever we assume $\rho < 0$.

Remark 3.3. The adaptive choice of the level k_1 suggested in Remark 3.2 is not straightforward in practice. The theoretical and simulated results in Fraga Alves et al. (2003), together with the use of these ρ-estimators in the Generalized Jackknife statistics of Gomes et al. (2000), as done in Gomes and Martins (2002), has led these authors to advise the choice $k_1 = \min\big(n-1, [2n/\ln\ln n]\big)$, to estimate ρ. Note however that with such a choice of k_1, $\sqrt{k_1}\, A^2(n/k_1) \to \infty$ and $\{\hat\rho_\tau(k_1) - \rho\} = O_p\big(A(n/k_1)\big) = O_p\big((\ln\ln n)^\rho\big)$. Consequently, without any further restrictions on the behavior of the ρ-estimators, we may no longer guarantee that (3.4) holds.

Remark 3.4. Here, and inspired in the results in Gomes et al. (2004*b*) for the estimator in (1.17), we advise the consideration of a level of the type

$$k_1 = \big[n^{1-\epsilon}\big], \qquad \text{for some } \ \epsilon > 0, \ \text{ small }, \tag{3.8}$$

where $[x]$ denotes, as usual, the integer part of x. When we consider the level k_1 in (3.8), $\sqrt{k_1}\, A^2(n/k_1) \to \infty$, if and only if $\rho > \frac{1}{4} - \frac{1}{4\epsilon} \to -\infty$, as $\epsilon \to 0$, and such a condition is an almost irrelevant restriction in the underlying model, provided we choose a small value of ϵ. For instance, if we choose $\epsilon = 0.001$, we get $\rho > -249.75$. Then, and with such an irrelevant restriction in the models in (1.4), if we work with any of the ρ-estimators in this section, computed at the level k_1, $\{\hat\rho - \rho\}$ is of the order of $A(n/k_1) = O(n^{\epsilon\times\rho})$, which is of smaller order than $1/\ln n$. This means that, again, condition (3.4) holds, being the choice in (3.8) a very adequate choice in practice.

We advise practitioners not to choose blindly the value of τ in (3.7). It is sensible to draw some sample paths of $\hat\rho(k;\tau)$, as functions of k and for a few τ-values, electing the value of $\tau \equiv \tau^*$ which provides the highest stability for large k, by means of any stability criterion, like the ones suggested in Gomes et al. (2004*a*), Gomes and Pestana (2004) and Gomes et al. (2005*a*). Anyway, in all the Monte Carlo simulations we have considered the level k_1 in (3.8), with

$\epsilon = 0.001$, and

$$\hat{\rho}_\tau := -\left|\frac{3\Big(T_n^{(\tau)}(k_1) - 1\Big)}{T_n^{(\tau)}(k_1) - 3}\right| , \qquad \tau = \begin{cases} 0 & \text{if } \rho \geq -1 , \\ 1 & \text{if } \rho < -1 . \end{cases} \tag{3.9}$$

Indeed, an adequate stability criterion, like the one used in Gomes and Pestana (2004), has practically led us to this choice for all models simulated, whenever the sample size n is not too small. Note also that the choice of the most adequate value of τ, let us say the tuning parameter $\tau = \tau^*$ mentioned before, is much more relevant than the choice of the level k_1, in the ρ-estimation and everywhere in the paper, whenever we use second order parameters' estimators in order to estimate the extreme value index.

From now on we shall generally use the notation $\hat{\rho} \equiv \hat{\rho}_\tau = \hat{\rho}(k_1;\tau)$ for any of the estimators in (3.7) computed at a level k_1 in (3.8).

3.3.2. The estimation of β based on the scaled log-spacings

We have here considered the estimator of β obtained in Gomes and Martins (2002), already defined in (1.14), and based on the scaled log-spacings U_i in (1.9), $1 \leq i \leq k$. The first part of the following result has been proved in Gomes and Martins (2002) and the second part, related to the behavior of $\hat{\beta}(k;\hat{\rho}(k;\tau))$, has been proved in Gomes et al. (2004b):

Proposition 3.2 (Gomes and Martins, 2002; Gomes et al., 2004b)**.** *If the second order condition* (1.2) *holds, with* $A(t) = \beta\,\gamma\,t^\rho$, $\rho < 0$, *if* $k = k_n$ *is a sequence of intermediate positive integers, i.e.* (1.7) *holds, and if* $\lim_{n\to\infty} \sqrt{k}\,A(n/k) = \infty$, *then* $\hat{\beta}(k;\rho)$, *defined in* (1.14), *converges in probability towards* β, *as* $n \to \infty$. *Moreover, if* (3.4) *holds,* $\hat{\beta}(k;\hat{\rho})$ *is consistent for the estimation of* β. *We may further say that*

$$\hat{\beta}\big(k;\hat{\rho}(k;\tau)\big) - \beta \;\overset{p}{\sim}\; -\beta \ln(n/k) \left(\hat{\rho}(k;\tau) - \rho\right) , \tag{3.10}$$

with $\hat{\rho}(k;\tau)$ *given in* (3.7). *Consequently,* $\hat{\beta}\big(k;\hat{\rho}(k;\tau)\big)$ *is consistent for the estimation of* β *whenever* (1.7) *holds and* $\sqrt{k}\,A(n/k)/\ln(n/k) \to \infty$. *For models in* (1.4), $\hat{\beta}\big(k;\hat{\rho}(k;\tau)\big) - \beta = O_p\big(\ln(n/k)/(\sqrt{k}\,A(n/k))\big)$ *whenever* $\sqrt{k}\,A^2(n/k) \to \lambda_{_A}$, *finite. If* $\sqrt{k}\,A^2(n/k) \to \infty$, *then* $\hat{\beta}\big(k;\hat{\rho}(k;\tau)\big) - \beta = O_p\big(\ln(n/k)\,A(n/k)\big)$.

An algorithm for second order parameter estimation, in a context of high quantiles estimation, can be found in Gomes and Pestana (2005).

4. FINITE SAMPLE BEHAVIOR OF THE ESTIMATORS

4.1. Simulated models

In the simulations we have considered the following underlying parents: the *Fréchet* model, with d.f. $F(x) = \exp(-x^{-1/\gamma})$, $x \geq 0$, $\gamma > 0$, for which $\rho = -1$, $\beta = 1/2$, $\beta' = 5/6$; and the *GP* model, with d.f. $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, $x \geq 0$, $\gamma > 0$, for which $\rho = -\gamma$, $\beta = 1$, $\beta' = 1$.

4.2. Mean values and mean squared error patterns

We have here implemented simulation experiments with 5000 runs, based on the estimation of β at the level k_1 in (3.8), with $\epsilon = 0.001$, the same level we have used for the estimation of ρ. We use the notation $\hat{\beta}_{j1} = \hat{\beta}(k_1; \hat{\rho}_j)$, $j = 0, 1$, with $\beta(k; \hat{\rho})$ and $\hat{\rho}_\tau$, $\tau = 0, 1$, given in (1.14) and (3.9), respectively. Similarly to what has been done in Gomes et al. (2004b) for the *WH*-estimator, in (1.17), and in Caeiro et al. (2005) for the $\overline{H}$-estimator, in (1.18), these estimators of ρ and β have been also incorporated in the $\widetilde{ML}$-estimators, leading to $\widetilde{ML}_0(k) \equiv \widetilde{ML}_{\hat{\beta}_{01}, \hat{\rho}_0}(k)$ or to $\widetilde{ML}_1(k) \equiv \widetilde{ML}_{\hat{\beta}_{11}, \hat{\rho}_1}(k)$, with $\widetilde{ML}$ denoting both ML and $\overline{ML}$ in (1.15) and (1.16), respectively.

The simulations show that the extreme value index estimators $UH_j(k) \equiv UH_{\hat{\beta}_{j1}, \hat{\rho}_j}(k)$, with UH denoting again either ML or $\overline{ML}$ or WH or $\overline{H}$, j equal to either 0 or 1, according as $|\rho| \leq 1$ or $|\rho| > 1$, seem to work reasonably well, as illustrated in Figures 1, 2 and 3. In these figures we picture for the above mentioned underlying models, and a sample of size $n = 1000$, the mean values ($E[\bullet]$) and the mean squared errors ($MSE[\bullet]$) of the Hill estimator H, together with UH_j (*left*), $UH_j^* \equiv UH_{\hat{\beta}(k; \hat{\rho}_j), \hat{\rho}_j}$ (*right*), with $j = 0$ or $j = 1$, according as $|\rho| \leq 1$ or $|\rho| > 1$ and the r.v.'s $UH \equiv UH_{\beta,\rho}$ (*center*). The discrepancy, in some of the models, between the behavior of the estimators proposed in this paper, the ones in the left figures, and the r.v.'s, in the central ones, suggests that some improvement in the estimation of second order parameters β and ρ is still welcome.

Remark 4.1. For the Fréchet model (Figure 1), the $UH_{\hat{\beta}, \hat{\rho}}$ estimators exhibit a negative bias up to moderate values of k and consequently, as hinted in Remark 2.1, the ML statistic is the one exhibiting the worst performance in terms of bias and minimum mean squared error. The $\overline{ML}_0$ estimator, always quite close to WH_0, exhibits the best performance among the statistics considered.

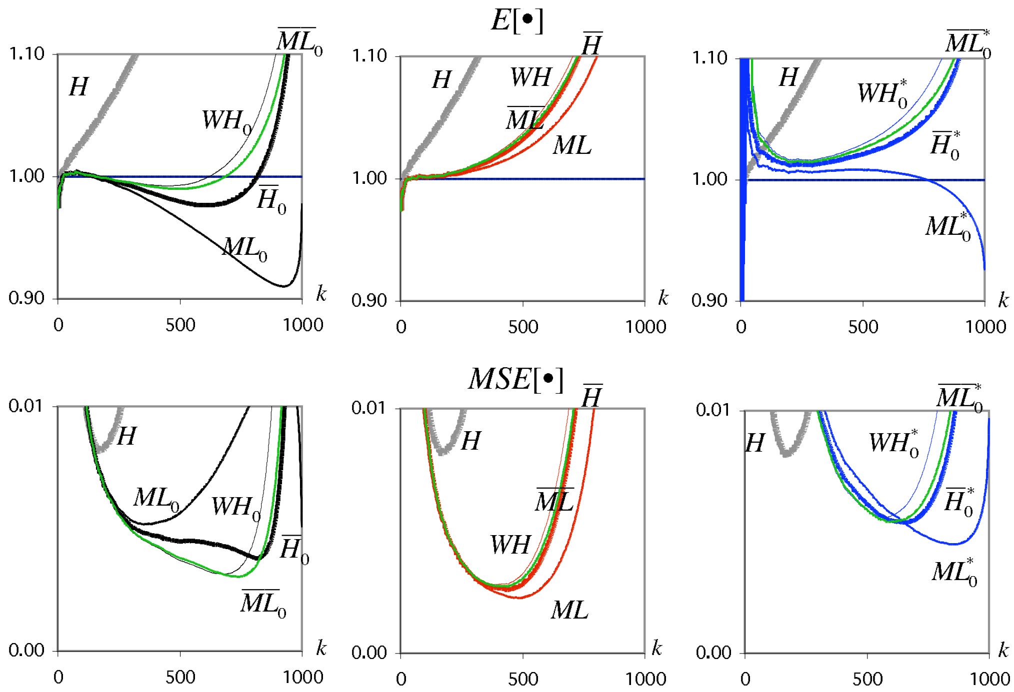

Figure 1: Underlying *Fréchet* parent with $\gamma = 1$ $(\rho = -1)$.

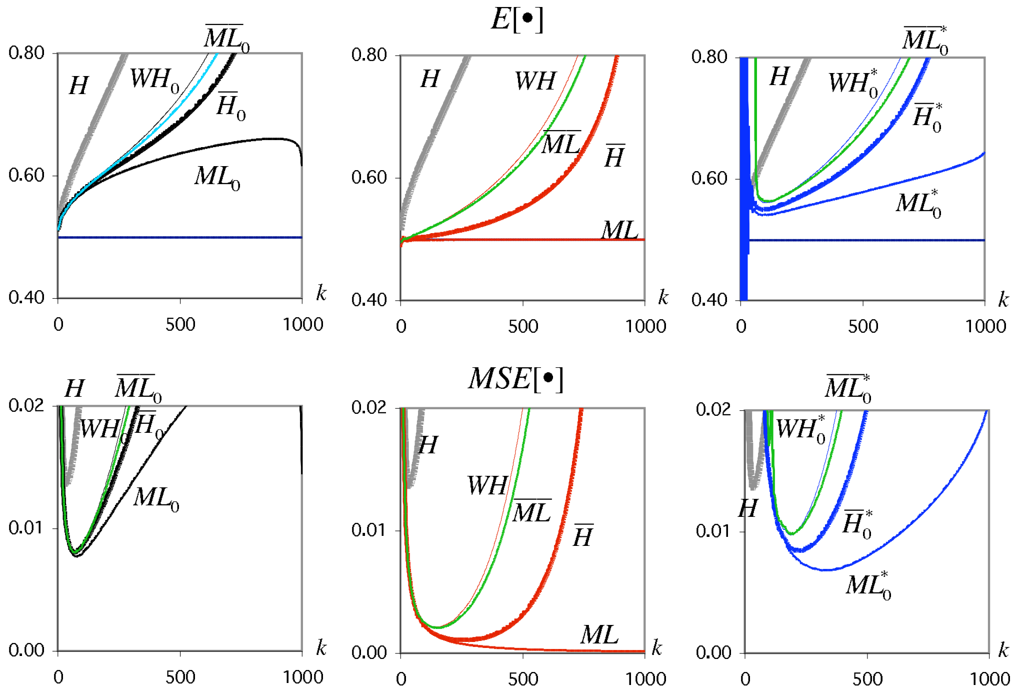

Figure 2: Underlying *GP* parent with $\gamma = 0.5$ $(\rho = -0.5)$.

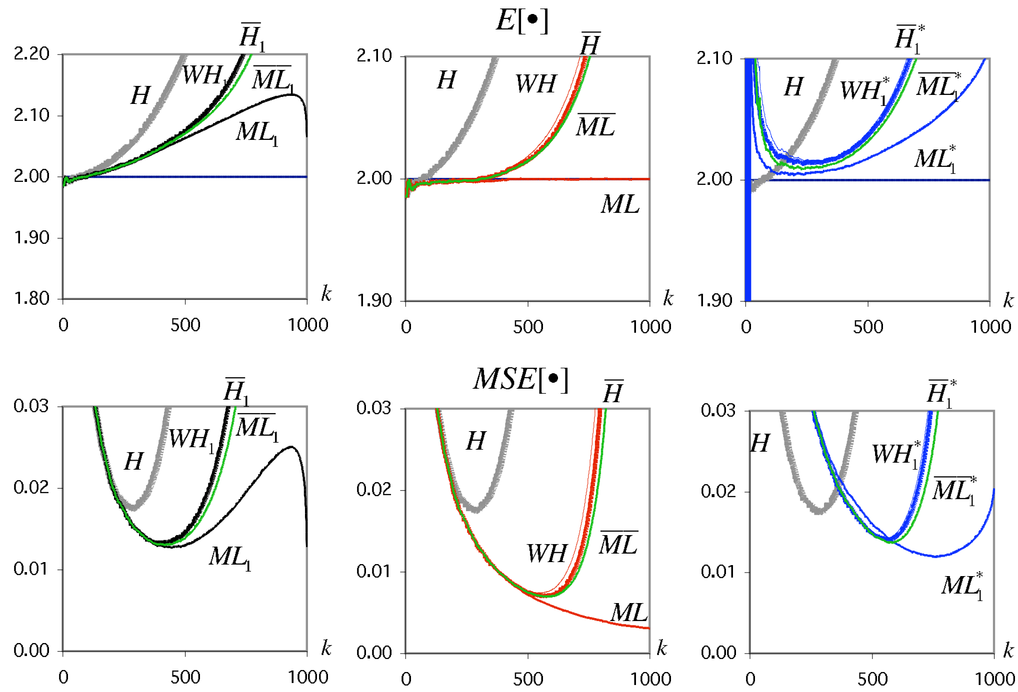

Figure 3: Underlying *GP* parent with $\gamma = 2$ ($\rho = -2$).

Things work the other way round, either with the r.v.'s UH (Figure 1, center) or with the statistics UH_0^* (Figure 1, right). The ML_0^* statistic is then the one with the best performance.

Remark 4.2. For a *GP* model, we make the following comments:

1) The ML statistic behaves indeed as a "really unbiased" estimator of γ, should we get to know the true values of β and ρ (see the central graphs of Figures 2 and 3). Indeed $b_{ML} = 0$ (see Remark 2.2), but we believe that more than this happens, although we have no formal proof of the unbiasedness of $ML(k)$ for all k and for Burr and *GP* models, among other possible parents.

2) For values of $\rho > -1$ (Figure 2), the estimators exhibit a positive bias, overestimating the true value of the parameter, and the ML-statistic is better than $\overline{H}$, which on its turn behaves better than $\overline{ML}$, this one better than WH, both regarding bias and mean squared error and in all situations (either when β and ρ are known or when β and ρ are estimated at the larger level k_1 or when only ρ is estimated at a larger level k_1, with β estimated at the same level than the extreme value index).

3) For $\rho < -1$ (Figure 3), we need to use $\hat{\rho}_1$ (instead of $\hat{\rho}_0$) or an hybrid estimator like the one suggested in Gomes and Pestana (2004). In all the simulated cases the ML_1-statistic is always the best one, being $\overline{ML}_1$, $\overline{H}_1$ and WH_1 almost equivalent.

4.3. Simulated comparative behavior at optimal levels

In Table 1, for the above mentioned Fréchet$(\gamma=1)$, $GP(\gamma=.5)$ and $GP(\gamma=2)$ parents and for the r.v.'s $UH \equiv UH_{\beta,\rho}$, we present the simulated values of the following characteristics at optimal levels: the optimal sample sample fraction (OSF)/ mean value (E) (*first row*) and the mean squared error (MSE)/ Relative Efficiency ($REFF$) indicator (*second row*). The simulated output is now based on a multi-sample simulation of size 1000×10, and standard errors, although not shown, are available from the authors. The OSF is, for any $T_n(k)$,

$$OSF_{T} \equiv \frac{k_0^{(T)}(n)}{n} := \frac{\arg\min_k MSE\big(T_n(k)\big)}{n} ,$$

and, relatively to the Hill estimator $H_n(k)$ in (1.10), the $REFF$ indicator is

$$REFF_{T} := \sqrt{MSE\Big[H_n\big(k_0^{(H)}(n)\big)\Big] \,/\, MSE\Big[T_n\big(k_0^{(T)}(n)\big)\Big]} .$$

For any value of n, and among the four r.v.'s, the largest $REFF$ (equivalent to smallest MSE) is **in bold and underlined**.

It is clear from Table 1 the overall best performance of ML estimator, whenever (β,ρ) is assumed to be known. Indeed, since $b_{ML}=0$, we were intuitively expecting this type of performance. The choice is not so clear-cut when we consider the estimation of the second order parameters, and either the statistics UH_j or the statistics UH_j^*. Tables 2, 3 and 4 are similar to Table 1, but for the extreme value index estimators UH_j and UH_j^*, $j=0$ or 1 according as $|\rho|\leq 1$ or $|\rho>1$. Again, for any value of n, and among any four estimators of the same type, the largest $REFF$ (equivalent to smallest MSE) is also in **bold and underlined** if it attains the largest value among all estimators, or only in **bold** if it attains the largest value among estimators of the same type.

A few remarks:

- For Fréchet parents, and among the UH_0^* estimators, the best performance is associated to $\overline{ML}_0^*$ for $n<500$ and to ML_0^* for $n\geq 500$. Among the UH_0 estimators, $\overline{ML}_0$ exhibits the best performance for all n.

- For GP parents with $\gamma = 0.5$, ML_0 exhibits the best performance among the UH_0 statistics. ML_0^* is also the best among the UH_0^* statistics, behaving ML_0^* better than ML_0, for all n.
- For GP parents with $\gamma = 2$, ML_1 exhibits the best performance among the UH_1 statistics. ML_1^* is also the best among the UH_1^* statistics. Now, ML_1^* behaves better than ML_1, for $n \geq 500$ and for $n < 500$ ML_1 performs better than ML_1^*.

Table 1: Simulated OSF/E (first row) and $MSE/REFF$ (second row) at optimal levels of the r.v.'s under study.

n	100	200	500	1000	2000
Fréchet parent, $\gamma = 1$ $(\rho = -1)$					
ML	0.642 / 0.986 0.015 / **<u>1.678</u>**	0.599 / 1.017 0.009 / **<u>1.734</u>**	0.517 / 1.037 0.004 / **<u>1.832</u>**	0.473 / 1.039 0.002 / **<u>1.909</u>**	0.429 / 1.012 0.001 / **<u>2.001</u>**
$\overline{ML}$	0.608 / 0.971 0.016 / 1.647	0.544 / 1.008 0.010 / 1.662	0.477 / 1.045 0.005 / 1.727	0.416 / 1.040 0.003 / 1.782	0.367 / 1.007 0.002 / 1.855
WH	0.580 / 0.960 0.018 / 1.539	0.513 / 1.019 0.011 / 1.577	0.450 / 1.052 0.005 / 1.658	0.395 / 1.041 0.003 / 1.723	0.357 / 1.003 0.002 / 1.805
$\overline{H}$	0.587 / 0.963 0.018 / 1.560	0.537 / 1.012 0.010 / 1.609	0.482 / 1.048 0.005 / 1.710	0.436 / 1.041 0.003 / 1.786	0.379 / 1.008 0.001 / 1.874
GP **parent**, $\gamma = 0.5$ $(\rho = -0.5)$					
ML	0.987 / 0.507 0.002 / **<u>5.813</u>**	0.985 / 0.513 0.001 / **<u>6.567</u>**	0.991 / 0.504 0.000 / **<u>7.831</u>**	0.990 / 0.504 0.000 / **<u>9.184</u>**	0.997 / 0.503 0.000 / **<u>10.487</u>**
$\overline{ML}$	0.295 / 0.565 0.009 / 2.529	0.240 / 0.545 0.006 / 2.561	0.183 / 0.530 0.003 / 2.591	0.157 / 0.531 0.002 / 2.697	0.124 / 0.523 0.001 / 2.753
WH	0.273 / 0.573 0.012 / 2.246	0.221 / 0.566 0.007 / 2.332	0.174 / 0.537 0.004 / 2.419	0.146 / 0.533 0.002 / 2.542	0.117 / 0.530 0.001 / 2.624
$\overline{H}$	0.391 / 0.549 0.007 / 2.918	0.353 / 0.537 0.004 / 3.128	0.302 / 0.536 0.002 / 3.367	0.262 / 0.5200 0.001 / 3.597	0.208 / 0.521 0.001 / 3.835
GP **parent**, $\gamma = 2$ $(\rho = -2)$					
ML	0.990 / 2.065 0.032 / **<u>1.923</u>**	0.994 / 1.921 0.016 / **<u>2.030</u>**	0.995 / 1.992 0.006 / **<u>2.211</u>**	0.993 / 2.011 0.00 / **<u>2.382</u>**	0.999 / 2.015 0.002 / **<u>2.541</u>**
$\overline{ML}$	0.731 / 2.111 0.050 / 1.530	0.677 / 1.956 0.027 / 1.544	0.633 / 2.033 0.012 / 1.573	0.588 / 2.047 0.007 / 1.602	0.549 / 2.063 0.004 / 1.640
WH	0.659 / 2.091 0.058 / 1.420	0.633 / 1.977 0.031 / 1.450	0.576 / 2.036 0.014 / 1.496	0.540 / 2.057 0.008 / 1.528	0.505 / 2.062 0.004 / 1.573
$\overline{H}$	0.669 / 2.103 0.058 / 1.423	0.647 / 1.976 0.030 / 1.470	0.604 / 2.047 0.013 / 1.525	0.574 / 2.053 0.007 / 1.570	0.533 / 2.057 0.004 / 1.622

Table 2: Simulated OSF/E (first row) and $MSE/REFF$ (second row) at optimal levels of the different estimators and r.v.'s under study, for Fréchet parents with $\gamma=1$ ($\rho=-1$, $\beta=0.5$).

n	100	200	500	1000	2000
H	0.326 / 1.026 0.044 / 1.000	0.281 / 1.069 0.026 / 1.000	0.222 / 1.056 0.013 / 1.000	0.174 / 1.055 0.008 / 1.000	0.138 / 1.031 0.005 / 1.000
ML_0	0.569 / 0.820 0.037 / 1.084	0.592 / 0.966 0.021 / 1.113	0.826 / 0.977 0.010 / 1.185	0.808 / 1.010 0.005 / 1.269	0.999 / 0.985 0.003 / 1.402
$\overline{ML}_0$	0.847 / 0.959 0.019 / **1.518**	0.802 / 1.027 0.012 / **1.485**	0.758 / 1.008 0.006 / **1.538**	0.727 / 1.026 0.003 / **1.641**	0.709 / 0.998 0.002 / **1.766**
WH_0	0.816 / 0.963 0.020 / 1.494	0.756 / 1.014 0.012 / 1.467	0.702 / 1.004 0.006 / 1.517	0.678 / 1.030 0.003 / 1.616	0.650 / 1.001 0.001 / 1.731
$\overline{H}_0$	0.877 / 0.951 0.024 / 1.358	0.841 / 1.005 0.015 / 1.331	0.819 / 0.998 0.007 / 1.376	0.808 / 1.026 0.004 / 1.469	0.808 / 0.973 0.002 / 1.576
ML_0^*	0.947 / 0.849 0.037 / 1.092	0.920 / 0.973 0.020 / 1.139	0.870 / 0.992 0.009 / **1.239**	0.855 / 1.019 0.005 / **1.349**	0.834 / 0.979 0.002 / **1.480**
$\overline{ML}_0^*$	0.858 / 0.988 0.027 / **1.277**	0.787 / 1.054 0.017 / **1.234**	0.676 / 1.064 0.009 / 1.222	0.603 / 1.058 0.005 / 1.230	0.530 / 1.001 0.003 / 1.246
WH_0^*	0.811 / 0.992 0.030 / 1.211	0.736 / 1.062 0.018 / 1.194	0.647 / 1.069 0.009 / 1.194	0.567 / 1.057 0.006 / 1.208	0.511 / 1.003 0.003 / 1.224
H_0^*	0.856 / 0.973 0.031 / 1.191	0.795 / 1.048 0.019 / 1.183	0.711 / 1.059 0.009 / 1.205	0.643 / 1.057 0.005 / 1.231	0.579 / 0.994 0.003 / 1.261

Table 3: Simulated OSF/E (first row) and $MSE/REFF$ (second row) at optimal levels of the different estimators and r.v.'s under study, for GP parents with $\gamma=0.5$ ($\rho=-0.5$, $\beta=1$).

n	100	200	500	1000	2000
H	0.103 / 0.742 0.058 / 1.000	0.077 / 0.646 0.037 / 1.000	0.051 / 0.632 0.020 / 1.000	0.040 / 0.602 0.014 / 1.000	0.028 / 0.585 0.009 / 1.000
ML_0	0.306 / 0.636 0.023 / **1.572**	0.216 / 0.633 0.017 / **1.474**	0.107 / 0.606 0.011 / **1.383**	0.076 / 0.583 0.008 / **1.339**	0.051 / 0.558 0.006 / **1.274**
$\overline{ML}_0$	0.211 / 0.674 0.029 / 1.418	0.149 / 0.618 0.019 / 1.383	0.101 / 0.606 0.011 / 1.338	0.073 / 0.588 0.008 / 1.310	0.049 / 0.558 0.006 / 1.258
WH_0	0.202 / 0.669 0.029 / 1.416	0.144 / 0.614 0.019 / 1.382	0.100 / 0.607 0.011 / 1.336	0.071 / 0.586 0.008 / 1.308	0.049 / 0.558 0.006 / 1.257
$\overline{H}_0$	0.234 / 0.641 0.029 / 1.418	0.165 / 0.640 0.019 / 1.384	0.103 / 0.607 0.011 / 1.339	0.073 / 0.588 0.008 / 1.310	0.049 / 0.557 0.006 / 1.257
ML_0^*	0.795 / 0.652 0.022 / **1.612**	0.636 / 0.628 0.016 / **1.525**	0.421 / 0.602 0.010 / **1.452**	0.310 / 0.578 0.007 / **1.420**	0.240 / 0.568 0.005 / **1.370**
$\overline{ML}_0^*$	0.449 / 0.720 0.049 / 1.090	0.350 / 0.654 0.030 / 1.114	0.251 / 0.610 0.015 / 1.148	0.192 / 0.600 0.010 / 1.185	0.140 / 0.579 0.006 / 1.199
WH_0^*	0.450 / 0.732 0.051 / 1.068	0.334 / 0.649 0.030 / 1.110	0.245 / 0.612 0.015 / 1.149	0.191 / 0.600 0.010 / 1.187	0.138 / 0.576 0.006 / 1.205
H_0^*	0.464 / 0.697 0.040 / 1.211	0.389 / 0.634 0.024 / 1.240	0.289 / 0.600 0.012 / 1.261	0.226 / 0.599 0.009 / 1.280	0.169 / 0.558 0.006 / 1.271

Table 4: Simulated OSF/E (first row) and $MSE/REFF$ (second row) at optimal levels of the different estimators and r.v.'s under study, for GP parents with $\gamma = 2$ ($\rho = -2,\ \beta = 1$).

n	100	200	500	1000	2000
H	0.415 / 2.179 0.117 / 1.000	0.359 / 1.968 0.064 / 1.000	0.319 / 2.018 0.030 / 1.000	0.290 / 2.068 0.018 / 1.000	0.251 / 2.069 0.010 / 1.000
ML_1	0.817 / 2.184 0.071 / **1.282**	0.647 / 2.012 0.043 / **1.221**	0.663 / 2.048 0.021 / **1.194**	0.657 / 2.077 0.013 / **1.173**	1.000 / 2.094 0.007 / **1.180**
$\overline{ML}_1$	0.631 / 2.140 0.079 / 1.215	0.558 / 2.008 0.046 / 1.184	0.478 / 2.044 0.022 / 1.168	0.399 / 2.050 0.013 / 1.158	0.358 / 2.040 0.008 / 1.153
WH_1	0.623 / 2.155 0.081 / 1.197	0.554 / 2.024 0.047 / 1.171	0.470 / 2.048 0.023 / 1.159	0.396 / 2.051 0.013 / 1.153	0.349 / 2.041 0.008 / 1.149
$\overline{H}_1$	0.618 / 2.167 0.083 / 1.186	0.545 / 2.041 0.047 / 1.165	0.470 / 2.050 0.023 / 1.156	0.396 / 2.051 0.013 / 1.152	0.349 / 2.041 0.008 / 1.148
ML_1^*	0.990 / 2.194 0.072 / **1.272**	0.935 / 2.000 0.044 / **1.211**	0.828 / 2.034 0.021 / **1.204**	0.768 / 2.077 0.012 / **1.197**	0.681 / 2.055 0.007 / **1.191**
$\overline{ML}_1^*$	0.751 / 2.199 0.089 / 1.143	0.696 / 1.993 0.050 / 1.129	0.624 / 2.044 0.024 / 1.123	0.571 / 2.065 0.014 / 1.125	0.519 / 2.041 0.008 / 1.130
WH_1^*	0.711 / 2.240 0.100 / 1.079	0.652 / 2.002 0.054 / 1.087	0.595 / 2.038 0.025 / 1.098	0.548 / 2.070 0.014 / 1.105	0.510 / 2.045 0.008 / 1.115
$\overline{H}_1^*$	0.710 / 2.240 0.10 / 1.0780	0.657 / 2.001 0.054 / 1.088	0.604 / 2.041 0.025 / 1.101	0.561 / 2.071 0.014 / 1.109	0.513 / 2.041 0.008 / 1.120

4.4. An overall conclusion

The main advantage of the estimators UH_j, and particularly of the ML_j estimators in this paper, the ones with an overall better performance, lies on the fact that we may estimate β and ρ adequately through $\hat{\beta}$ and $\hat{\rho}$ so that the MSE of the new estimator is smaller than the MSE of Hill's estimator for all k, even when $|\rho| > 1$, a region where it has been difficult to find alternatives for the Hill estimator. And this happens together with a higher stability of the sample paths around the target value γ. These new estimators work indeed better than the Hill estimator for all values of k, contrarily to the alternatives so far available in the literature, like the alternatives UH_j^*, $j = 0$ or 1, also considered in this paper for comparison.

5. CASE-STUDIES IN THE FIELDS OF FINANCE AND INSURANCE

5.1. Euro-UK Pound daily exchange rates

We shall first consider the performance of the above mentioned estimators in the analysis of the Euro-UK Pound daily exchange rates from January 4, 1999 until December 14, 2004. This data has been collected by the European System of Central Banks, and was obtained from `http://www.bportugal.pt/rates/cambtx/`. In Figure 4 we picture the Daily Exchange Rates x_t over the above mentioned period and the Log-Returns, $r_t = 100\times(\ln x_t - \ln x_{t-1})$, the data to be analyzed. Indeed, although conscious that the log-returns of any financial time-series are not i.i.d., we also know that the semi-parametric behavior of estimators of rare event parameters may be generalized to weak dependent data (see Drees, 2002, and references therein). Semi-parametric estimators of extreme events' parameters, devised for i.i.d. processes, are usually based on the tail empirical process, and remain consistent and asymptotically normal in a large class of weakly dependent data.

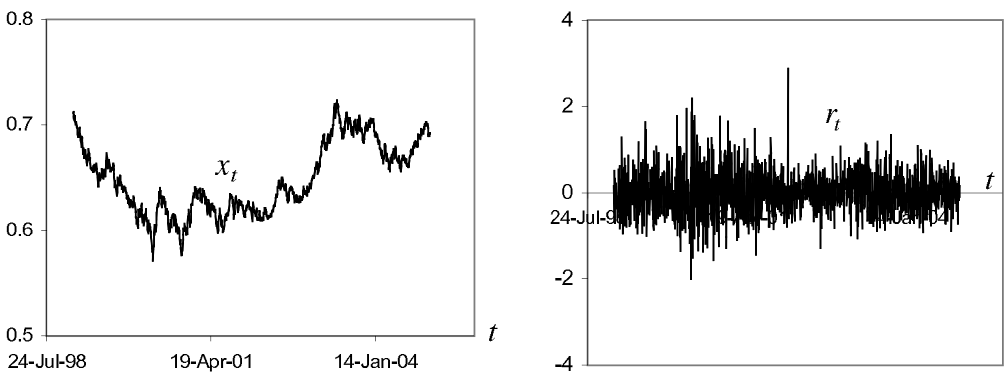

Figure 4: Daily Exchange Rates (*left*) and Daily Log-Returns (*right*) on Euro-UK Pound Exchange Rate.

The histogram in Figure 5 points to a heavy right tail. Indeed, the empirical counterparts of the usual skewness and kurtosis coefficients are $\hat\beta_1 = 0.424$ and $\hat\beta_2 = 1.835$, clearly greater than 0, the target value for an underlying normal parent.

In Figure 6, and working with the $n_0 = 725$ positive log-returns, we now picture the sample paths of $\hat\rho(k;\tau)$ in (3.7) for $\tau = 0$, and 1 (*left*), as functions of k. The sample paths of the ρ-estimates associated to $\tau = 0$ and $\tau = 1$ lead us

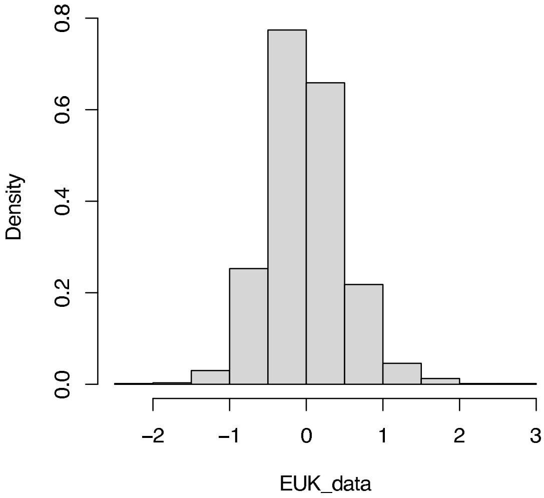

Figure 5: Histogram of the Daily Log-Returns on the Euro-UK Pound.

to choose, on the basis of any stability criterion for large values of k, the estimate associated to $\tau = 0$. In Figure 6 we thus present the associated second order parameters estimates, $\hat\rho_0 = \hat\rho_0(721) = -0.65$ (*left*) and $\hat\beta_0 = \hat\beta_{\hat\rho_0}(721) = 1.03$, together with the sample paths of $\hat\beta(k;\hat\rho_0)$ in (1.14), for $\tau = 0$ (*center*). The sample paths of the classical Hill estimator in (1.10) (H) and of three of reduced-bias, second order extreme value index estimates discussed in this paper, associated to $\hat\rho_0 = -0.65$ and $\hat\beta_0 = 1.03$, are also pictured in Figure 6 (*right*). We do not picture the statistic WH_0 because that statistic practically overlaps $\overline{ML}_0$.

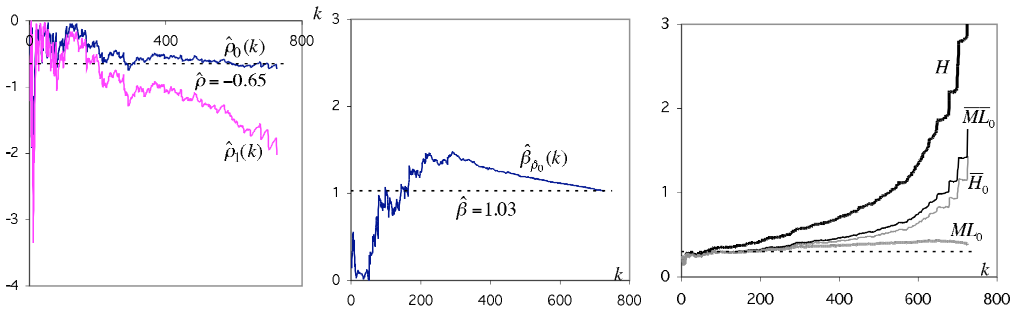

Figure 6: Estimates of the second order parameter ρ (*left*), of the second order parameter β (*center*) and of the extreme value index (*right*), for the Daily Log-Returns on the Euro-UK Pound.

The Hill estimator exhibits a relevant bias, as may be seen from Figure 6, and we are for sure a long way from the strict Pareto model. The other estimators, ML_0, $\overline{ML}_0$ and $\overline{H}_0$, which are "asymptotically unbiased", reveal without doubt a bias much smaller than that of the Hill. All these statistics enable us to take a decision upon the estimate of γ to be used, with the help of any stability criterion, but the *ML* statistic is without doubt the one with smallest bias, among the statistics considered. More important than this: we know that any estimate considered on the basis of $ML_0(k)$ (or any of the other three reduced-bias statistics) performs for sure better than the estimate based on $H(k)$ for any level k. Here, we represent the estimate $\widehat{\gamma} \equiv \widehat{\gamma}_{ML} = 0.30$, the median of the *ML* estimates, for thresholds k between $\left[n_0^{-2\hat{\rho}/(1-2\hat{\rho})}/4\right] = 10$ and $\left[4 \times n_0^{-2\hat{\rho}/(1-2\hat{\rho})}\right] = 165$, chosen in an heuristic way. If we use this same criterion on the estimates $\overline{ML}$, *WH* and $\overline{H}$ we are also led to the same estimate, $\widehat{\gamma}_{\overline{ML}} \equiv \widehat{\gamma}_{WH} \equiv \widehat{\gamma}_{\overline{H}} = 0.30$. The development of adequate techniques for the adaptive choice of the optimal threshold for this type of second order reduced-bias extreme value index estimators is needed, being indeed an interesting topic of research, but is outside the scope of the present paper.

5.2. Automobile claims

We shall next consider an illustration of the performance of the above mentioned estimators, through the analysis of automobile claim amounts exceeding 1,200.000 Euros, over the period 1988–2001, and gathered from several European insurance companies co-operating with the same re-insurer (Secura Belgian Re). This data set has already been studied, for instance, in Beirlant et al. (2004). Figure 7 is similar to Figure 5, but for the Secura data. It is now quite clear the heaviness of the right tail. The empirical skewness and kurtosis coefficients are $\hat{\beta}_1 = 2.441$ and $\hat{\beta}_2 = 8.303$. Here, the existence of left-censoring is also clear, begin the main reason for the high skewness and kurtosis values.

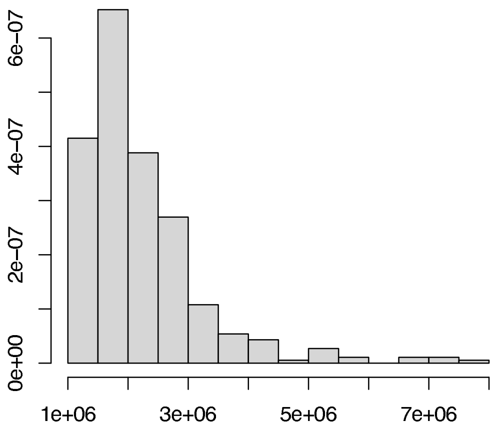

Figure 7: Histogram or the Secura data.

Finally, in Figure 8, working with the $n = 371$ automobile claims exceeding 1,200.000 Euro, we present the sample path of the $\hat{\rho}_\tau$ (*left*), $\hat{\rho}_\tau$ (*center*) estimates, as function of k, for $\tau = 0$ and $\tau = 1$, together with the sample paths of estimates of the extreme value index γ, provided by the Hill estimator, H, the M-estimator and the $\overline{M}$ estimator (*right*).

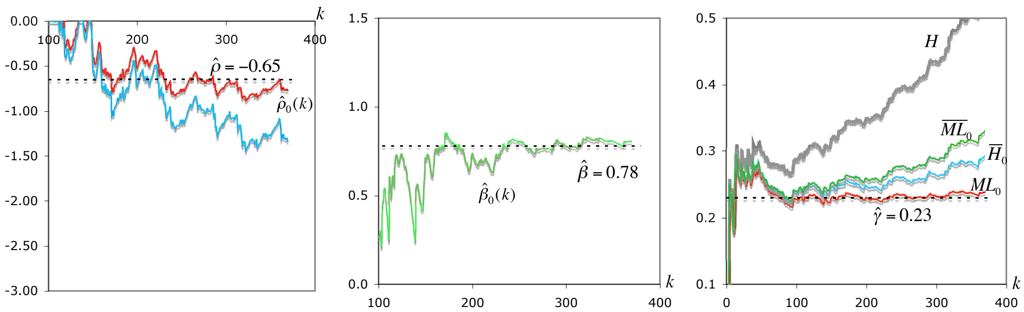

Figure 8: Estimates of the second order parameter ρ (*left*) and of the extreme value index γ (*right*) for the automobile claims.

Again, the ML_0 statistic is the one exhibiting the best performance, leading us to the estimate $\widehat{\gamma} = 0.23$.

ACKNOWLEDGMENTS

Research partially supported by FCT / POCTI and POCI / FEDER.

REFERENCES

[1] Bingham, N.H.; Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press.

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

[3] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J. *Statistics of Extremes. Theory and Applications*, Wiley, New York.

[4] Caeiro, F. and Gomes, M.I. (2004). A new class of estimators of the "scale" second order parameter. To appear in *Extremes*.

[5] Caeiro, F.; Gomes, M.I. and Pestana, D.D. (2005). Direct reduction of bias of the classical Hill estimator, *Revstat*, **3**, 2, 113–136.

[6] Drees, H. (1998). A general class of estimators of the tail index, *J. Statist. Planning and Inference*, **98**, 95–112.

[7] Drees, H. (2002). *Tail empirical processes under mixing conditions.* In "Empirical Process Techniques for Dependent Data" (Dehling *et al.*, Eds.), Birkhăuser, Boston, 325–342.

[8] Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation, *Stoch. Proc. and Appl.*, **75**, 149–172.

[9] Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution, *Ann. Statist.*, **27**, 760–781.

[10] Fraga Alves, M.I.; Gomes, M.I. and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica*, **60**, 1, 193–213.

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

[12] Gomes, M.I.; Caeiro, F. and Figueiredo, F. (2004a). Bias reduction of a tail index estimator trough an external estimation of the second order parameter, *Statistics*, **38**, 6, 497–510.

[13] Gomes, M.I.; Figueiredo, F. and Mendonça, S. (2005a). Asymptotically best linear unbiased tail estimators under a second order regular variation condition, *J. Statist. Planning and Inference*, **134**, 2, 409–433.

[14] Gomes, M.I., de Haan, L. and Peng, L. (2002). Semi-parametric estimation of the second order parameter — asymptotic and finite sample behavior, *Extremes*, **5**, 4, 387–414.

[15] Gomes, M.I.; Haan, L. de and Rodrigues, L. (2004b). *Tail index estimation through accommodation of bias in the weighted log-excesses*, Notas e Comunicações, CEAUL, 14/2004 (submitted).

[16] Gomes, M.I. and Martins, M.J. (2001). Generalizations of the Hill estimator — asymptotic versus finite sample behaviour, *J. Statistical Planning and Inference*, **93**, 161–180.

[17] Gomes, M.I. and Martins, M.J. (2002). "Asymptotically unbiased" estimators of the extreme value index based on external estimation of the second order parameter, *Extremes*, **5**, 1, 5–31.

[18] Gomes, M.I. and Martins, M.J. (2004). Bias reduction and explicit estimation of the extreme value index, *J. Statist. Planning and Inference*, **124**, 361–378.

[19] Gomes, M.I.; Martins, M.J. and Neves, M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events — the Jackknife methodology. *Extremes*, **3**, 3, 207–229.

[20] Gomes, M.I.; Miranda, C. and Pereira, H. (2005b). Revisiting the role of the Jackknife methodology in the estimation of a positive tail index, *Comm. in Statistics — Theory and Methods*, **34**, 1–20.

[21] Gomes, M.I.; Miranda, C. and Viseu, C. (2006). Reduced bias tail index estimation and the Jackknife methodology, *Statistica Neerlandica*, **60**, 4, 1–28.

[22] Gomes, M.I. and Oliveira, O. (2000). The bootstrap methodology in Statistical Extremes — choice of the optimal sample fraction, *Extremes*, **4**, 4, 331–358.

[23] Gomes, M.I. and Pestana, D. (2004). A simple second order reduced-bias extreme value index estimator. To appear in *J. Statist. Comput. and Simulation*.

[24] Gomes, M.I. and Pestana, D. (2005). A sturdy reduced-bias extreme quantile (*VaR*) estimator. To appear in *J. American Statist. Assoc.*.

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

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

[27] Peng, L. (1998). Asymptotically unbiased estimator for the extreme-value index, *Statistics and Probability Letters*, **38**, 2, 107–115.

REVSTAT – Statistical Journal
Volume 5, Number 2, June 2007, 209–226

INTERFAILURE DATA WITH CONSTANT HAZARD FUNCTION IN THE PRESENCE OF CHANGE-POINTS

Authors: Jorge Alberto Achcar
– Departamento de Estatística, Universidade Federal de São Carlos, São Carlos-SP, Brazil
jachcar@power.ufscar.br

Selene Loibel
– Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, S.P., Brazil
sloibel@terra.com.br

Marinho G. Andrade
– Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo-USP, São Carlos, S.P., Brazil
marinho@icmc.usp.br

Received: October 2006 Revised: February 2007 Accepted: April 2007

Abstract:

- Markov Chain Monte Carlo (MCMC) methods are used to perform a Bayesian analysis for interfailure data with constant hazard function in the presence of one or more change-points. We also present some Bayesian criteria to discriminate different models. The methodology is illustrated with a data set originally reported in Maguire, Pearson and Wynn [8].

Key-Words:

- *constant hazard; change-points; Gibbs sampling; MCMC algorithms.*

AMS Subject Classification:

- 62J02, 62F10, 62F03.

1. INTRODUCTION

Applications of change-point models are given in many areas of interest. For example, medical researchers usually have interest to know if a new therapy of leukemia produces a departure from the usual experience of a constant relapse rate after the induction of a remission (see for example, Matthews and Farewell [9], Matthews *et al.* [10] or Henderson and Matthews [6]). Bayesian analysis for change-point models has been introduced by many authors. A Bayesian analysis for a homogeneous Poisson process with a change-point has been introduced by Raftery and Akman [11]. A Bayesian interval estimator has been derived for a change-point in a Poisson process by West and Ogden [15] and a Bayesian approach for lifetime data with a constant hazard function and censored data in the presence of a change point by Achcar and Bolfarine [1]. Recently Loschi and Cruz [7] presented a Bayesian approach to the multiple change point identification problem in Poisson data.

In this paper, we consider the presence of two or more change-point for lifetime with constant hazards, generalizing previous work (see for example, Achcar and Bolfarine [1]).

Consider a homogeneous Poisson process with one or more change-points at unknown times. With a single change-point, the rate of occurrence at time s is given by

$$\lambda(s) = \begin{cases} \lambda_1, & 0 \le s \le \tau\,, \\ \lambda_2, & s > \tau\,. \end{cases} \tag{1.1}$$

The analysis of the Poisson process is based on the counting data in the period $[0,T]$, where $N(T)=n$ is the number of events that occur at the ordered times $t_1, t_2, ..., t_n$.

With two change-points at unknown times τ_1 and τ_2 the rate of occurrences are given by

$$\lambda(s) = \begin{cases} \lambda_1, & 0 < s \le \tau_1\,, \\ \lambda_2, & \tau_1 < s \le \tau_2\,, \\ \lambda_3, & \tau_2 < s \le T\,. \end{cases} \tag{1.2}$$

We also could have homogeneous Poisson processes with more than two change-points.

The use of Bayesian methods has been considered by many authors for homogeneous or nonhomogeneous Poisson processes in the presence of one change-point (see for example, Raftery and Akman [11] or Ruggeri and Sivaganesan [13]).

Observe that times between failures for a homogeneous Poisson process follow an exponential distribution.

In this paper, we present a Bayesian analysis for interfailure data with constant hazard function assuming more than one change-point and using MCMC methods (see for example [4]).

The paper is organized as follows: in Section 2, we introduce the likelihood function; in Section 3, we introduce a Bayesian analysis for the model, in Section 4, we present some consideration on model selection; in Section 5, we introduce an example with real data and finally, in Section 6, we present some conclusions.

2. THE LIKELIHOOD FUNCTION

Let $x_i = t_i - t_{i-1}$, $i = 1, 2, ..., n$ where $t_0 = 0$, be the interfailure times and assume a single-change-point model (1.1). In this way, we observe that x_i has an exponential distribution with parameter λ_1 for $\sum_{k=1}^{i} x_k \leq \tau$ and an exponential distribution with parameter λ_2 for $\sum_{k=1}^{i} x_k > \tau$, $i = 1, 2, ..., n$. Assuming that the change-point τ is taking the values t_i, the likelihood function for λ_1, λ_2 and τ is given by

$$L(\lambda_1, \lambda_2, \tau) = \prod_{i=1}^{N(T)} (\lambda_1 e^{-\lambda_1 x_i})^{\epsilon_i} (\lambda_2 e^{-\lambda_2 x_i})^{1-\epsilon_i} \tag{2.1}$$

where $\epsilon_i = 1$ if $\sum_{j=1}^{i} x_j \leq \tau$ and $\epsilon_i = 0$ if $\sum_{j=1}^{i} x_j \geq \tau$. That is,

$$L(\lambda_1, \lambda_2, \tau) = \lambda_1^{N(\tau)} e^{-\lambda_1 \tau} \lambda_2^{N(T)-N(\tau)} e^{-\lambda_2 (T-\tau)} \tag{2.2}$$

where $N(\tau) = \sum_{i=1}^{N(T)} \epsilon_i$, $N(T) = n$, $\tau = \sum_{i=1}^{N(T)} x_i \epsilon_i$ and $T - \tau = \sum_{i=1}^{N(T)} x_i (1 - \epsilon_i)$.

Let us assume a two-change-point model (1.2) with the change-points τ_1 and τ_2 taking discrete values $\tau_1 = t_i$, $\tau_2 = t_j$ ($t_i < t_j$, $i \neq j$) with $k_1 = N(\tau_1)$ and $k_2 = N(\tau_2)$. The likelihood function for λ_1, λ_2, λ_3, τ_1 and τ_2 is given by

$$L(\lambda_1, \lambda_2, \lambda_3, \tau_1, \tau_2) = \prod_{i=1}^{n} (\lambda_1 e^{-\lambda_1 x_i})^{\epsilon_{1,i}} (\lambda_2 e^{-\lambda_2 x_i})^{\epsilon_{2,i}} (\lambda_3 e^{-\lambda_3 x_i})^{\epsilon_{3,i}} \tag{2.3}$$

where

$$\epsilon_{1,i} = \begin{cases} 1 & \text{if } \sum_{k=1}^{i} x_k \leq \tau_1 , \\ 0 & \text{if } \sum_{k=1}^{i} x_k > \tau_1 , \end{cases} \tag{2.4}$$

$$\epsilon_{2,i} = \begin{cases} 1 & \text{if } \tau_1 < \sum_{k=i+1}^{j} x_k \leq \tau_2 , \\ 0 & \text{if } \sum_{k=i+1}^{j} x_k \leq \tau_1 \text{ or } \sum_{k=i+1}^{j} x_k > \tau_2 , \end{cases} \tag{2.5}$$

$$\epsilon_{3,i} = \begin{cases} 1 & \text{if } \tau_2 < \sum_{k=j+1}^{n} x_k , \\ 0 & \text{if } \tau_2 \geq \sum_{k=j+1}^{n} x_k . \end{cases} \tag{2.6}$$

That is,

$$L(\lambda_1, \lambda_2, \lambda_3, \tau_1, \tau_2) = \tag{2.7}$$
$$= \lambda_1^{N(\tau_1)} e^{-\lambda_1\tau_1}\, \lambda_2^{N(\tau_2)-N(\tau_1)} e^{-\lambda_2(\tau_2-\tau_1)}\, \lambda_3^{N(T)-N(\tau_2)} e^{-\lambda_3(T-\tau_2)}$$

where $\sum_{i=1}^{N(T)} \epsilon_{1,i} = N(\tau_1)$, $\sum_{i=1}^{N(T)} \epsilon_{2,i} = N(\tau_2) - N(\tau_1)$, $\sum_{i=1}^{N(T)} \epsilon_{3,i} = N(T) - N(\tau_2)$ and $N(T) = n$. Observe that $\tau_1 = \sum_{i=1}^{N(T)n} x_i\, \epsilon_{1,i}$, $\tau_2 - \tau_1 = \sum_{i=1}^{N(T)n} x_i\, \epsilon_{2,i}$ and $T - \tau_2 = \sum_{i=1}^{N(T)n} x_i\, \epsilon_{3,i}$.

In the same way, we could generalize for more than two change-points.

3. A BAYESIAN ANALYSIS

Assume the change-point model (1.1) with a single change-point τ.

Assume that τ is independent from λ_1 and λ_2, and also that λ_1 is conditionally independent from λ_2, given $\tau = t_i$. Considering a noninformative prior distribution for λ_1 and λ_2 given τ (see for example, Box and Tiao [2]), we have

$$\pi\big(\lambda_1, \lambda_2, \tau = t_i\big) = \pi\big(\lambda_1, \lambda_2 \,|\, \tau = t_i\big)\, \pi\big(\tau = t_i\big) \propto \frac{1}{\lambda_1\lambda_2}\, \pi\big(\tau = t_i\big) \tag{3.1}$$

where $\lambda_1, \lambda_2 > 0$.

Assuming an uniform prior distribution $\pi_0(\tau = t_i) = 1/n$, the joint posterior distribution for λ_1, λ_2 and τ is given by

$$\pi\big(\lambda_1, \lambda_2, \tau \,|\, \mathcal{D}\big) \propto \lambda_1^{N(\tau)-1} e^{-\lambda_1\tau}\, \lambda_2^{n-N(\tau)-1} e^{-\lambda_2(T-\tau)} \tag{3.2}$$

where $\mathcal{D}$ denotes the data set.

Observe that we are using a data dependent prior distribution for the discrete change-point (see for example Achcar and Bolfarine [1]). Also observe that the event $\{\tau = t_i\}$ is equivalent to $\{N(t_i) = i\}$, where the t_i are the ordered occurrence epochs of failures. We also could consider an informative gamma prior distribution for the parameters λ_1 and λ_2.

The marginal posterior distribution for τ is, from (3.2), given by

$$\pi\big(\tau \,|\, \mathcal{D}\big) \propto \frac{\Gamma\big[N(\tau)\big]\; \Gamma\big[n - N(\tau)\big]}{\tau^{N(\tau)}\,(T-\tau)^{n-N(\tau)}}\ . \tag{3.3}$$

Assuming $\tau = \tau^*$ known, the marginal posterior distribution for λ_1 and λ_2 are given by

$$\begin{aligned} &\text{(i)} \quad \lambda_1 | \tau^*, \mathcal{D} \sim \text{Gamma}\big[N(\tau^*), \tau^*\big]\,, \\ &\text{(ii)} \quad \lambda_2 | \tau^*, \mathcal{D} \sim \text{Gamma}\big[n - N(\tau^*), T - \tau^*\big]\,, \end{aligned} \tag{3.4}$$

where Gamma$[a,b]$ denotes a gamma distribution with mean a/b and variance a/b^2.

Assuming τ unknown, since the marginal posterior distribution for τ is obtained analytically (see(3.3)), we use a mixed Gibbs sampling and Metropolis–Hastings algorithm to generate the posterior distributions of λ_1 and λ_2. The conditional posterior distributions for the Gibbs sampling algorithm are given by

$$\text{(3.5)} \qquad \begin{aligned} &\text{(i)} \quad \lambda_1 \,|\, \lambda_2,\, \tau,\, \mathcal{D} \sim \text{Gamma}\big[N(\tau), \tau\big] ,\\ &\text{(ii)} \quad \lambda_2 \,|\, \lambda_1,\, \tau,\, \mathcal{D} \sim \text{Gamma}\big[n - N(\tau), T - \tau\big] . \end{aligned}$$

Starting with initial values $\lambda_1^{(0)}$ and $\lambda_2^{(0)}$, we follow the steps:

(i) Generate $\tau^{(i)}$ from (3.3).

(ii) Generate $\lambda_1^{(i+1)}$ from $\pi\big(\lambda_1 \,|\, \lambda_2^{(i)}, \tau^{(i)}, \mathcal{D}\big)$.

(iii) Generate $\lambda_2^{(i+1)}$ from $\pi\big(\lambda_2 \,|\, \lambda_1^{(i+1)}, \tau^{(i)}, \mathcal{D}\big)$.

We could monitor the convergence of the Gibbs samples using Gelman and Rubin's method that uses the analysis of variance technique to determine whether further iterations are needed (see [5] for details).

A great simplification to get the posterior summaries of interest for the constant hazard function model in the presence of a change-point is to use the *WinBugs* software (see, Spiegelhalter *et al.* [14]) which requires only the specification of the distribution for the data and prior distributions for the parameters.

Consider now, the change-point model (1.2) with two change-points τ_1 and τ_2 (with $\tau_1 < \tau_2$). The prior density for λ_1, λ_2, λ_3, τ_1 and τ_2 is given by

$$\text{(3.6)} \qquad \begin{aligned} \pi\big(\lambda_1, \lambda_2, \lambda_3, \tau_1, \tau_2\big) &= \\ &= \pi\big(\lambda_1, \lambda_2, \lambda_3 \,|\, \tau_1 = t_i, \tau_2 = t_j\big)\; \pi_0\big(\tau_1 = t_i, \tau_2 = t_j\big)\, I_{\{t_i < t_j\}} , \end{aligned}$$

given $\tau_1 = t_i$, $\tau_2 = t_j$, $(t_i < t_j,\ i \neq j)$.

Assuming τ_1 and τ_2 independent from λ_1, λ_2 and λ_3, and also that λ_1, λ_2 and λ_3 are conditionally independent given τ_1 and τ_2, a noninformative joint prior distribution for λ_1, λ_2, λ_3 and τ_1 and τ_2 is given by

$$\text{(3.7)} \qquad \pi\big(\lambda_1, \lambda_2, \lambda_3, \tau_1 = t_i, \tau_2 = t_j\big) \propto \frac{1}{\lambda_1 \lambda_2 \lambda_3}\, \pi_0\big(\tau_1 = t_i, \tau_2 = t_j\big)\, \mathrm{I}_{\{t_i < t_j\}}$$

where $\lambda_1, \lambda_2, \lambda_3 > 0$, $\mathrm{I}_{\{t_i < t_j\}} = 1$ if $t_i < t_j$ and $\mathrm{I}_{\{t_i < t_j\}} = 0$ otherwise, for all $i \neq j$.

Assuming an uniform prior distribution for the discrete variables $\tau_1 = t_i$ and $\tau_2 = t_j$, where $t_i < t_j$, $i, j = 1, ..., n$, that is $\pi_0(\tau_1 = t_i, \tau_2 = t_j) = 2/n(n-1)$, the joint posterior distribution for λ_1, λ_2, λ_3, τ_1 and τ_2 is given by

$$\text{(3.8)} \qquad \begin{aligned} \pi\big(\lambda_1, \lambda_2, \lambda_3, \tau_1, \tau_2 \,|\, \mathcal{D}\big) &\propto \\ &\propto \lambda_1^{N(\tau_1)-1} e^{-\lambda_1 \tau_1}\, \lambda_2^{N(\tau_2)-N(\tau_1)-1} e^{-\lambda_2(\tau_2 - \tau_1)}\, \lambda_3^{N(T)-N(\tau_2)-1} e^{-\lambda_3(T - \tau_2)} \end{aligned}$$

where $\lambda_1, \lambda_2, \lambda_3 > 0$ and $\tau_1 < \tau_2$.

The joint marginal posterior distribution for τ_1 and τ_2 is given by

$$\pi(\tau_1, \tau_2 | \mathcal{D}) = \frac{\Gamma\big[N(\tau_1)\big]\,\Gamma\big[N(\tau_2) - N(\tau_1)\big]}{\tau_1^{N(\tau_1)} (\tau_2 - \tau_1)^{N(\tau_2) - N(\tau_1)}} \, \frac{\Gamma\big[N(\tau_2) - N(\tau_1)\big]}{(T - \tau_2)^{N(T) - N(\tau_2)}} \,. \tag{3.9}$$

We use the Metropolis–Hastings algorithm to generate τ_1, τ_2 from the joint marginal posterior distribution (3.9) and the Gibbs sampling algorithm to generate λ_1, λ_2 and λ_3. The conditional posterior distribution for the Gibbs sampling algorithm are given by

$$\lambda_1 |\ \lambda_2, \lambda_3, \tau_1, \tau_2, \mathcal{D} \sim \text{Gamma}\big[N(\tau_1), \tau_1\big] \,, \tag{3.10}$$

$$\lambda_2 |\ \lambda_1, \lambda_3, \tau_1, \tau_2, \mathcal{D} \sim \text{Gamma}\big[N(\tau_2) - N(\tau_1), \tau_2 - \tau_1\big] \,, \tag{3.11}$$

$$\lambda_3 |\ \lambda_1, \lambda_2, \tau_1, \tau_2, \mathcal{D} \sim \text{Gamma}\big[N(T) - N(\tau_2), T - \tau_2\big] \,. \tag{3.12}$$

This marginalization process should be made with careful choice of the lower and upper limits of summation as well as of the number of minimum points between τ_1 and τ_2. We consider $\tau_1 = t_i$ for $i = 1, ..., m_1 - 1$, $\tau_2 = t_i$ for $i = m_2 + 1, ..., n$, where $\tau_1 < \tau_2$ and m_j $(j = 1, 2)$ is a positive integer such that $t_{m_j} = \tau_j$. Note that once τ_1, (τ_2) is known, possible candidates of τ_1, (τ_2) are limited within $\{t_1, ..., t_{m_1-1}\}$, $(\{t_{m_2+1}, ..., t_n\})$.

Starting with the initial values $\lambda_1^{(0)}$, $\lambda_2^{(0)}$ and $\lambda_3^{(0)}$, we follow the steps:

(i) Generate $\tau_1^{(i)}$ and $\tau_2^{(i)}$ from the marginal posterior distributions (3.9).

(ii) Generate $\lambda_1^{(i+1)}$ from $\pi\big(\lambda_1 | \lambda_2^{(i)}, \lambda_3^{(i)}, \tau_1^{(i)}, \tau_2^{(i)}, \mathcal{D}\big)$.

(iii) Generate $\lambda_2^{(i+1)}$ from $\pi\big(\lambda_2 | \lambda_1^{(i+1)}, \lambda_3^{(i)}, \tau_1^{(i)}, \tau_2^{(i)}, \mathcal{D}\big)$.

(iv) Generate $\lambda_3^{(i+1)}$ from $\pi\big(\lambda_3 | \lambda_1^{(i+1)}, \lambda_2^{(i+1)}, \tau_1^{(i)}, \tau_2^{(i)}, \mathcal{D}\big)$.

Observe that the choices for m_1 and m_2 could have been made empirically based on a preliminary analysis of the data set (empirical Bayesian methods). In this way, we could use plots of the accumulated number of failures against time of occurrence to get some information on the change-point.

4. SOME CONSIDERATIONS ON MODEL SELECTION

For model selection, we could use the predictive density for the interfailure time x_i given $\underset{\sim}{x}_{(i)} = (x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$. The predictive density for x_i given $\underset{\sim}{x}_{(i)}$ is

$$c_i = f\big(x_i | \underset{\sim}{x}_{(i)}\big) = \int f\big(x_i | \underset{\sim}{\theta}\big)\; \pi\big(\underset{\sim}{\theta} | \underset{\sim}{x}_{(i)}\big)\; d\underset{\sim}{\theta} \tag{4.1}$$

where $\pi\big(\underset{\sim}{\theta} | \underset{\sim}{x}_{(i)}\big)$ is the posterior density for a vector of parameters $\underset{\sim}{\theta}$ given the data $\underset{\sim}{x}_{(i)}$.

Using the Gibbs samples, (4.1) can be approximated by its Monte Carlo estimates,

$$\hat{f}\big(x_i \,|\, \underset{\sim}{x}_{(i)}\big) \;=\; \frac{1}{M}\sum_{j=1}^{M} f\big(x_i \,|\, \underset{\sim}{\theta}^{(j)}\big)\;, \tag{4.2}$$

where $\underset{\sim}{\theta}^{(j)}$ are the generated Gibbs samples, $j = 1, 2, ..., M$.

We can use $c_i = \hat{f}\big(x_i \,|\, \underset{\sim}{x}_{(i)}\big)$ in model selection. In this way, we consider plots of c_i versus i ($i = 1, 2, ..., n$) for different models; large values of c_i (in average) indicates a better model. We could also have choosen the model such that $P_l = \prod_{i=1}^{n} c_i(l)$ is maximum (l indexes models). We could also have considered (see Raftery [12]) the marginal likelihood of the whole data set $\mathcal{D}$ for a model M_l given by

$$P(\mathcal{D}\,|\,M_l) \;=\; \int_{\theta_l} L(\mathcal{D}\,|\,\theta_l, M_l)\; \pi(\theta_l\,|\,M_l)\; d\theta_l \tag{4.3}$$

where $\mathcal{D}$ is the data, M_l is the model specification (the number of change points), θ_l is the vector of the parameters in M_l, $L(\mathcal{D}\,|\,\theta, M_l)$ is the likelihood function and $\pi(\theta_l\,|\,M_l)$ is the prior.

The Bayes factor criterion prefers model M_1 to model M_2 if $P(\mathcal{D}\,|\,M_2) < P(\mathcal{D}\,|\,M_1)$. A Monte Carlo estimate for the marginal likelihood $P(\mathcal{D}\,|\,M_l)$ is given by

$$\hat{P}(\mathcal{D}\,|\,M_l) \;=\; \frac{1}{M}\sum_{j=1}^{M} L\big(\mathcal{D}\,|\,\theta_l^{(j)}, M_l\big) \tag{4.4}$$

where $\theta_l^{(j)}$, $j = 1, 2, ..., M$, could have been generated through the use of importance sampling. The simplest estimator of this type results from taking the prior as the importance sampling function (see Raftery [12]).

Other ways to estimate the marginal likelihood $P(\mathcal{D}\,|\,M_l)$ are proposed by Raftery [12].

Considering a sample from the posterior distribution, we have

$$\hat{P}(\mathcal{D}\,|\,M_l) \;=\; \left(\frac{1}{M}\sum_{j=1}^{M} \frac{1}{L\big(\mathcal{D}\,|\,\theta_l^{(j)}, M_l\big)}\right)^{-1}. \tag{4.5}$$

In this case, the importance-sampling function is the posterior distribution.

A modification of the harmonic mean estimator (4.5) is proposed by Gelfand and Dey [3], given by

$$\hat{P}(\mathcal{D}\,|\,M_l) \;=\; \left(\frac{1}{M}\sum_{j=1}^{M} \frac{f\big(\theta_l^{(j)}\big)}{L\big(\mathcal{D}\,|\,\theta_l^{(j)}, M_l\big)\; \pi_0\big(\theta_l^{(j)}\big)}\right)^{-1} \tag{4.6}$$

where $f(\theta_l)$ is any probability density and $\pi_0(\theta_l)$ is a prior probability density.

5. AN EXAMPLE

In this section, we analyze a data set related to the number of mine accidents in England from 1875 to 1951. To analyze this data set, we have assumed the validity of a homogeneous Poisson process in the presence of change-points. Considering the time intervals between explosions in mines, we introduced a Bayesian analysis to get inference for the parameter of the exponential distributions and for the finite change-points.

In Table 1, we have the time intervals (in days) between explosions in mines, involving more than 10 men killed, from December 6, 1875 to May 29, 1951 (data introduced by Maguire, Pearson and Wynn [8]).

Table 1: Time intervals in days between explosions in mines.

378	36	15	31	215	11	137	4	15	72	96
124	50	120	203	176	55	93	59	315	59	61
1	13	189	345	20	81	286	114	108	188	233
28	22	61	78	99	326	275	54	217	113	32
23	151	361	312	354	58	275	78	17	1205	644
467	871	48	123	457	498	49	131	182	255	195
224	566	390	72	228	271	208	517	1613	54	326
1312	348	745	217	120	275	20	66	291	4	369
338	336	19	329	330	312	171	145	75	364	37
19	156	47	129	1630	29	217	7	18	1357	

From a plot of $N(t_i)$ versus t_i, $i=1,2,...,109$ (see Figure 1), we observe the presence of two or more change-points. We could also have assumed the presence of a random number of change-points (see for example, Ruggeri and Sivaganesan [13]) but this case is beyond the scope of this paper. As an illustration of the proposed model introduced in Section 1, we assume the presence of two change-points. Assuming the two change-points model (1.2) to analyze the data set of Table 1 and from Figure 1, we see that these two change-points are approximately $\widehat{\tau}_1 = t_{45} = 5231$ and $\widehat{\tau}_2 = t_{81} = 19053$. We also assume the presence of only one change-point and use Bayesian discrimination methods to decide for the best model.

In Figure 1, we also have empirical estimates for the rates λ_j, $j=1,2,3$, obtained from the usual definition of the homogeneous Poisson processes $N(t) \propto \lambda t + o(n)$, where $N(t)$ is the accumulated number of occurrences in the interval $(0,t)$.

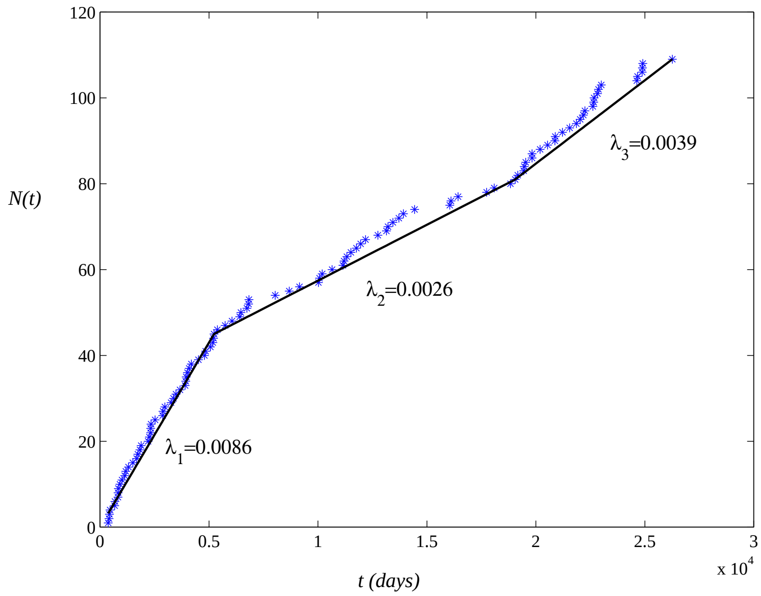

Figure 1: Plot of $N(t_i)$ versus $t_i(days)$.

If we assume the change-point model (1.1) with a single change-point τ with an uniform discrete prior, the mode of the marginal posterior distribution for τ (see (3.3)) is given by $\tau^* = 5382$ (see Figure 2). Assuming τ^* known, the mean of the marginal posterior distributions (3.4) are given by $\widetilde{\lambda}_1 = 0.008361$ and $\widetilde{\lambda}_2 = 0.003065$.

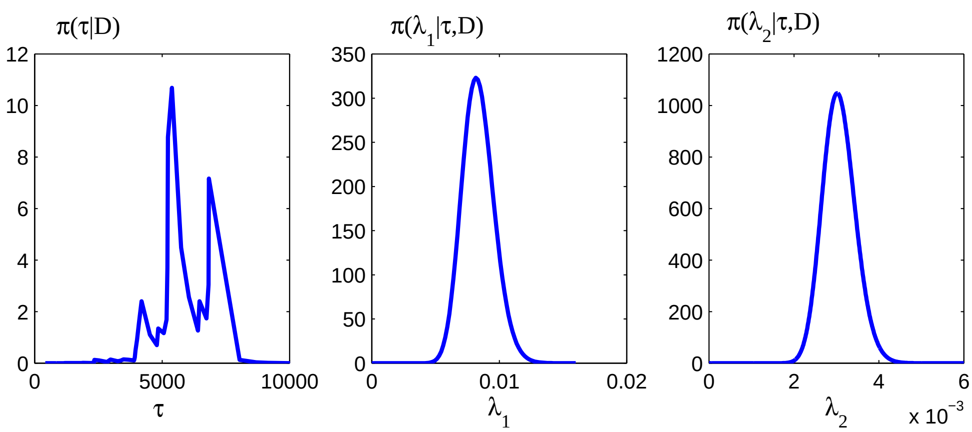

Figure 2: Marginal posterior distribution for τ and, λ_1 and λ_2 with $\tau = \tau^*$.

Assuming one or two unknown change-points, we have obtained posterior summaries (see Tables 2, 3, 4 and 5) through the use of MCMC algorithms.

In all cases, we have considered a "burn-in-sample" of size 5,000; after this, we have simulated 50,000 mixed Metropolis–Hastings and Gibbs samples taking every 10^{th} sample, to get approximated uncorrelated samples. The convergence of the mixed algorithms was monitored using graphical methods and standard existing indexes (see, for example, Gelman and Rubin [5]).

Considering the change-point model (1.1) with only one change-point τ, we have in Table 2, the posterior summaries for the parameters τ, λ_1 and λ_2 assuming the noninformative prior (3.1). In Figure 3, we have the approximate marginal posterior densities.

Table 2: Posterior summaries (change-point model 1.1).

	Mean	S.D.	95% Cred. Inter.
τ	5813	932	(4086 ; 7364)
λ_1	0.008059	0.001285	(0.005814 ; 0.010786)
λ_2	0.003047	4.011E-4	(0.002289 ; 0.003884)

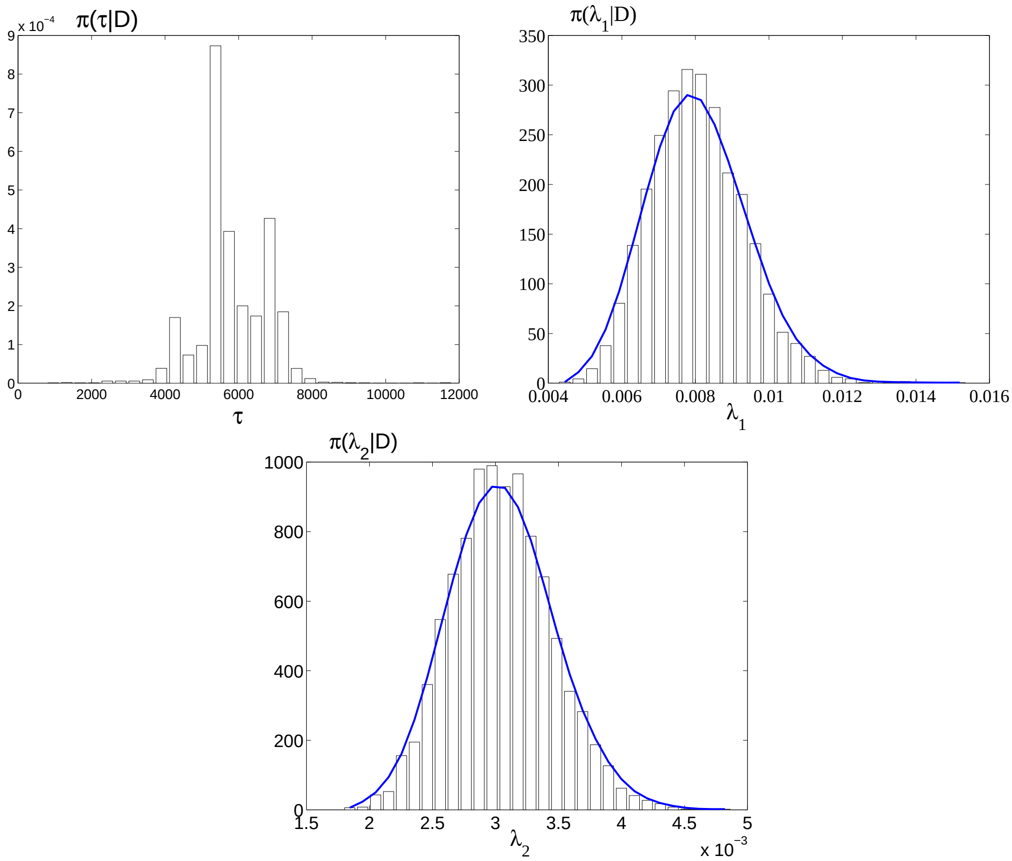

Figure 3: Marginal posterior distribution (change-point model 1.1).

Similar results could also have been obtained from the parametrization $k = N(t_k)$, λ_1 and λ_2. Assuming an uniform prior distribution for $N(t_i)$ taking the values $\{1, 2, ..., n\}$ and Gamma$(0.1, 0.1)$ prior distributions for λ_1 and λ_2, we obtain by Gibbs sampling algorithms the approximate marginal posterior densities for τ, λ_1 and λ_2. In Table 3 we have the posterior summaries of interest using the *WinBugs* software. The code of the *WinBugs* program is given in Appendix 1, assuming $k = N(t_k)$. Observe that $k \cong 46$ corresponds to $\tau = 5382$. That is, we have obtained results similar to the previous ones.

Table 3: Posterior summaries (gamma priors for λ_1 and λ_2).

	Mean	S.D.	95% Cred. Inter.
k	45.63	5.186	(35.0 ; 53.0)
λ_1	0.008322	0.001315	(0.006085 ; 0.01120)
λ_2	0.003056	3.975E-4	(0.002344 ; 0.003892)

In Figure 4, we have the approximated marginal posterior densities considering the 5,000 generated Gibbs samples.

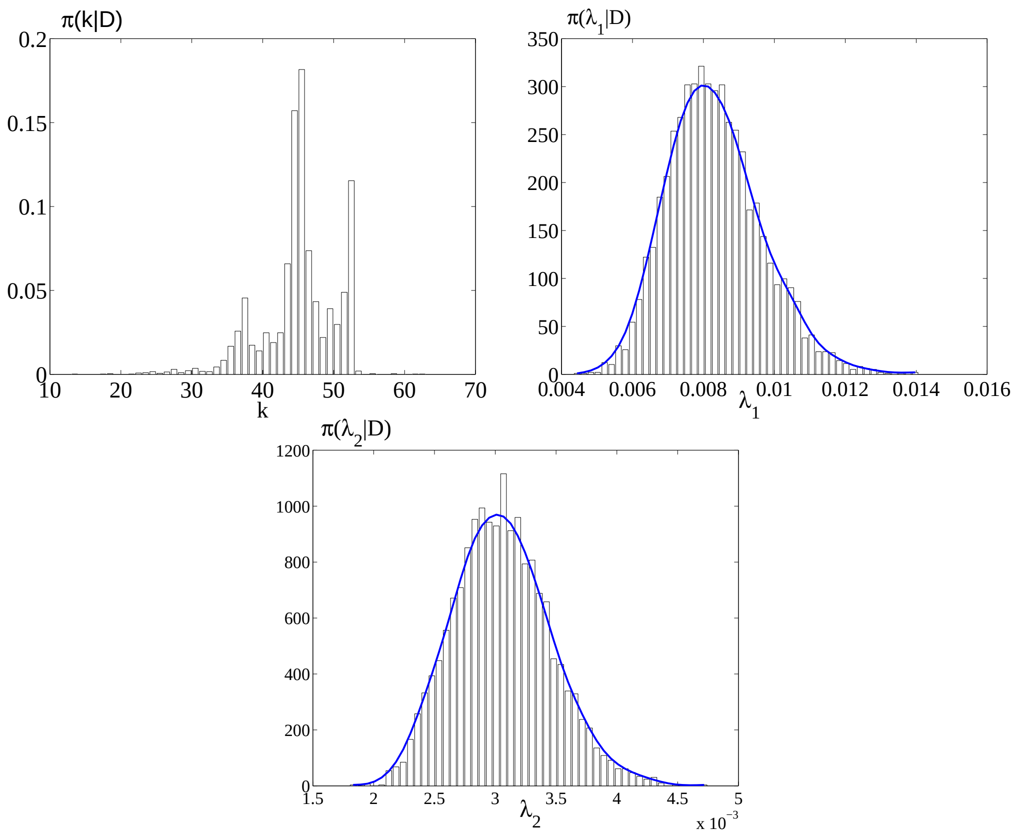

Figure 4: Marginal posterior distribution (gamma prior distribution for λ_1 and λ_2).

Assuming the two change-point model (1.2), we have in Table 4, the posterior summaries for the parameters λ_1, λ_2, λ_3, τ_1 and τ_2 obtained from the 5,000 generated Gibbs samples using the conditional posterior distributions (3.10)–(3.12). In Figure 5 we have the approximate marginal posterior densities.

Table 4: Posterior summaries (change-point model 1.2).

	Mean	S.D.	95% Cred. Inter.
τ_1	5990	876	(4176 ; 7354)
τ_2	17459	3162	(11287 ; 22741)
λ_1	0.008036	0.001262	(0.005765 ; 0.010703)
λ_2	0.002713	6.080E-4	(0.001655 ; 0.004053)
λ_3	0.003450	7.646E-4	(0.002103 ; 0.005082)

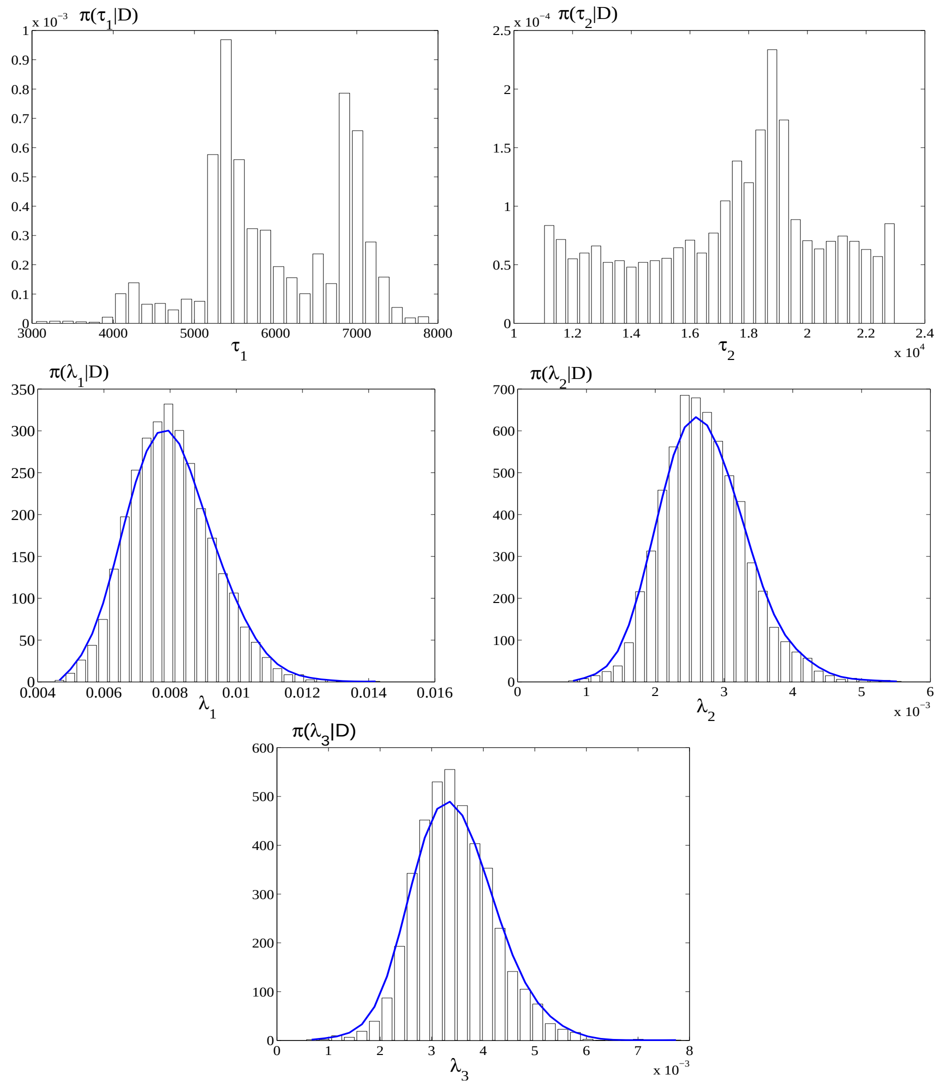

Figure 5: Marginal posterior distributions (change-point model 1.2).

Similar results have been obtained from the parametrization $k_1 = N(t_{k_1})$, $k_2 = N(t_{k_2})$, λ_1, λ_2 and λ_3. In Table 5, we have the posterior summaries of interest, obtained using the *WinBugs* software (code in Appendix 1), informative discrete prior distributions for the two change-points and independent Gamma(0.1, 0.1) prior distributions for λ_1, λ_2 and λ_3. Observe that $k_1 \cong 46$ corresponds to $\tau_1 = 5382$ and $k_2 \cong 78$ corresponds to $\tau_2 = 17743$. In Figure 6, we have the approximate marginal posterior distributions considering the 5,000 generated Gibbs samples.

Table 5: Posterior summaries (two change-point and gamma priors for λ_1, λ_2 and λ_3).

	Mean	S.D.	95% Cred. Inter.
k_1	46.22	4.237	(37.0 ; 53.0)
k_2	78.29	10.45	(58.0 ; 97.0)
λ_1	0.008349	0.001298	(0.006077 ; 0.01115)
λ_2	0.002780	6.378E-4	(0.001606 ; 0.004134)
λ_3	0.003445	7.392E-4	(0.002195 ; 0.005079)

In Figure 7, we have plots of the predictive densities $c_i = f(x_i | \underset{\sim}{x}_{(i)})$, $i = 1, 2, ..., n$, approximated by the Monte Carlo estimates (4.2) for both models M_1 (a single change-point model) and M_2 (two change-points model). For model M_1, we have $P_1 = \prod_{i=1}^{n} \widehat{c}_{1i} = 7.896 \times 10^{-303}$ and for model M_2 we have $P_2 = \prod_{i=1}^{n} \widehat{c}_{2i} = 9.5536 \times 10^{-302}$. The ratio of these values is given by $P_2/P_1 = 12.09$.

In Table 6, we have different estimates (see (4.5) and (4.6)) for the marginal likelihood functions considering models M_1 (single change-point model) and M_2 (two change-point model).

Table 6: Estimate values of the marginal likelihood.

Model	$P(\mathcal{D}\|M_l)$ using (4.5)	$P(\mathcal{D}\|M_l)$ using (4.6)
M_1	7.7716×10^{-305}	4.6420×10^{-304}
M_2	3.1256×10^{-304}	2.5020×10^{-302}

From Table 6, we calculate the Bayes factors $B_{ij} = P(\mathcal{D}|M_i)/P(\mathcal{D}|M_j)$, $i, j = 1, 2$. The Bayes factors are given by $B_{21} = 4.02$ (using (4.5)) and $B_{21} = 53.9$ (using (4.6)). If compared to one change-point model, we observe a better fit of the two change-point model M_2 for the data set of Table 1, considering the three model selection procedures.

It is important to point out that better models also could be considered to analyze the data set of the Table 1, considering more than two change-points.

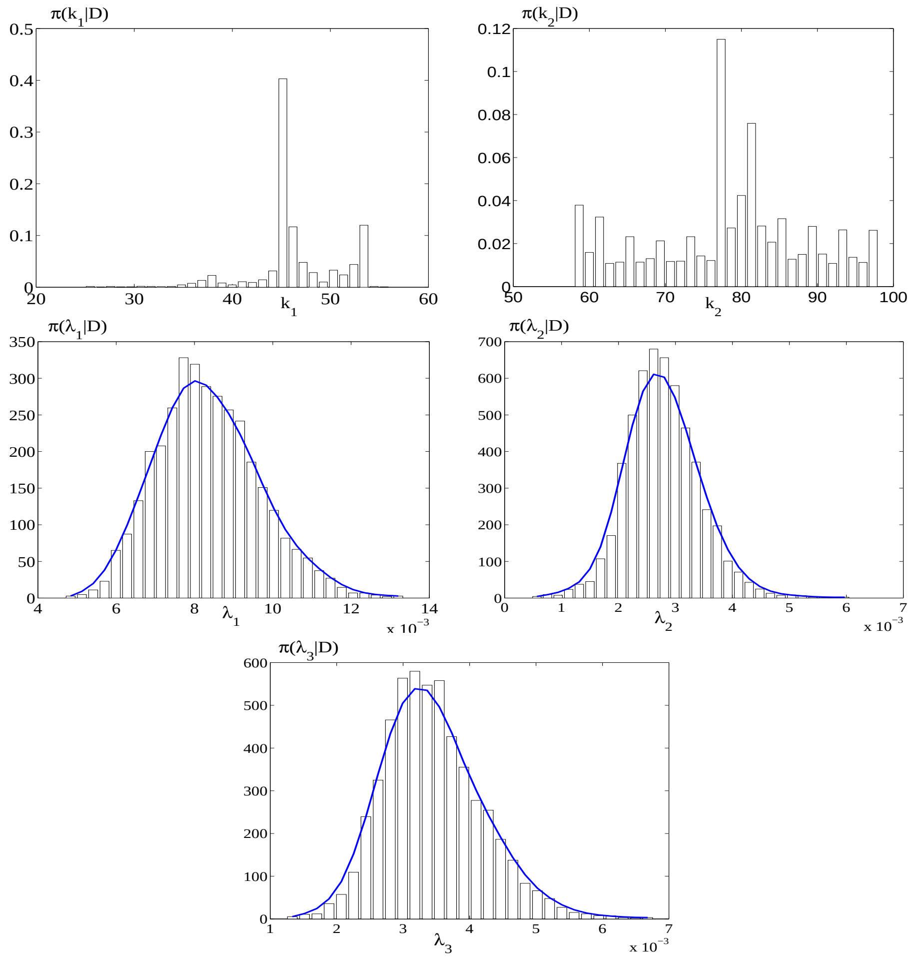

Figure 6: Marginal posterior distributions (gamma prior distributions for λ_1, λ_2 and λ_3 an informative discrete prior distribution for τ_1 and τ_2).

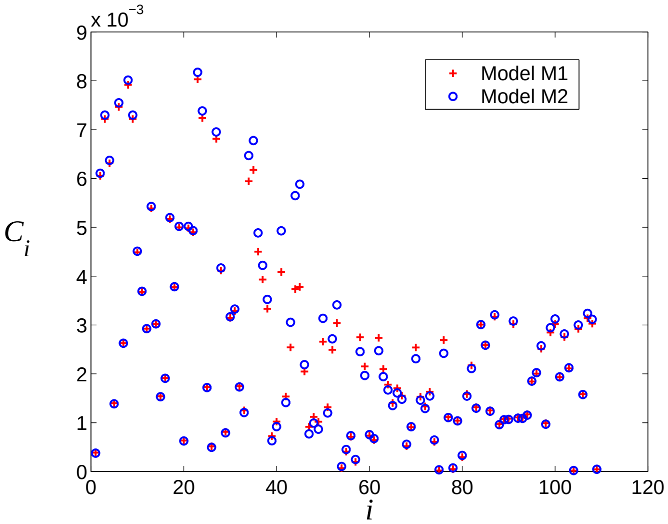

Figure 7: Plot of c_i versus i (M_1: $+$, M_2: $\circ$).

6. CONCLUDING REMARKS

In this paper, we have observed that Bayesian inference for the parameters of change-point models is easily obtained through the use of Markov Chain Monte Carlo methods.

The use of recent software, such as *WinBugs*, to simulate samples for the joint posterior distribution of interest gives a great simplification in the computational work. It is important to point out that the usual classical inference procedures usually are not appropriate for change-point models (see for example, Mattews *et al.* [10]).

The proposed Bayesian methodology could also have been considered directly using the counting data modeled by homogeneous Poisson processes in the presence of one or more change-points in place of the inter-failure data (see for example, Raftery and Akman [11]).

Similar results could have been obtained for interfailure data with constant hazards and more than two change-points.

The use of Monte Carlo estimates for the predictive densities $f\left(x_i \,|\, \underset{\sim}{x}_{(i)}\right)$, $i=1,2,...,n$, or for the marginal likelihood of the whole data set $\mathcal{D}$ for a model M_l, gives simple ways to discriminate the different change-point models, a problem of great practical interest.

APPENDIX

A. *WinBugs* code (one change-point)

```
Model
      {
      for(i in 1 : N) {
          t[i] ∽ dexp(lambda[J[i]])
          J[i]< −1+step(i−k−0.5)
          punif[i]<−1/N
                          }
      for(j in 1 : 2) {
          lambda[j]∽ dgamma(0.1, 0.1)
                          }
      k ∽ dcat(punif[ ])
      }
```

```
list(t=c(378, 36, 15, 31, 215, 11, 137, 4, 15, 72, 96, 124, 50, 120, 203, 176, 55, 93,
59, 315, 59, 61, 1, 13, 189, 345, 20, 81, 286, 114, 108, 188, 233, 28, 22, 61, 78, 99,
326, 275, 54, 217, 113, 32, 23, 151, 361, 312, 354, 58, 275, 78, 17, 1205, 644, 467,
871, 48, 123, 457, 498, 49, 131, 182, 255, 195, 224, 566, 390, 72, 228, 271, 208,
517, 1613, 54, 326, 1312, 348, 745, 217, 120, 275, 20, 66, 291, 4, 369, 338, 336, 19,
329, 330, 312, 171, 145, 75, 364, 37, 19, 156, 47, 129, 1630, 29, 217, 7, 18, 1357),
N=109)
    list(k=50, lambda=c(0.5, 0.5))
```

B. ***WinBugs*** code (two change-point)

```
Model
    {
    for(i in 1 : N) {
        t[i] ∽ dexp(lambda[J[i]])
        J[i]< −1+step(i−k1−0.5)+step(i−k2−0.5)
                    }
    for(j in 1 : 3) {
        lambda[j]∽ dgamma(0.1, 0.1)
                    }
    k1∽dcat(p1[ ])
    k2∽dcat(p2[ ])
    }
list(t=c(378, 36, 15, 31, 215, 11, 137, 4, 15, 72, 96, 124, 50, 120, 203, 176, 55, 93,
59, 315, 59, 61, 1, 13, 189, 345, 20, 81, 286, 114, 108, 188, 233, 28, 22, 61, 78, 99,
326, 275, 54, 217, 113, 32, 23, 151, 361, 312, 354, 58, 275, 78, 17, 1205, 644, 467,
871, 48, 123, 457, 498, 49, 131, 182, 255, 195, 224, 566, 390, 72, 228, 271, 208,
517, 1613, 54, 326, 1312, 348, 745, 217, 120, 275, 20, 66, 291, 4, 369, 338, 336, 19,
329, 330, 312, 171, 145, 75, 364, 37, 19, 156, 47, 129, 1630, 29, 217, 7, 18, 1357),
p1=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.0225, 0.0225,
0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225,
0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.1, 0.0225, 0.0225,
0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225,
0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0), p2=(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225,
0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.10,
0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225,
0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0.0225, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0), N=109)
    list(k1=45,k2=78,lambda=c(0.5,0.5,0.5))
```

ACKNOWLEDGMENTS

The authors acknowledge helpful suggestions by the editor and the referees of this paper.

REFERENCES

[1] Achcar, J.A. and Bolfarine, H. (1989). Constant hazard against a change-point alternative: a Bayesian approach with censored data, *Communications in Statistics — Theory and Methods*, **18**, 10, 3801–3819.

[2] Box, G.E. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*, Addison-Wesley, New York.

[3] Gelfand, A.E. and Dey, D.K. (1994). Bayesian model choice: asymptotics and exact calculations, *Journal of the Royal Statistical Society, B*, **56**, 501–514.

[4] Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**, 398–409.

[5] Gelman, A.E. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457–472.

[6] Henderson, R. and Matthews, J.N.S. (1993). An investigation of change-points in the annual number of case of haemolytic uraemic sydrom, *Applied Statistics*, **42**, 461–471.

[7] Loschi, R.H. and Cruz, F.R.B. (2005). Byesian identification of multiple change points in Poisson data, *Advances in Complex Systems*, **8**, 4, 465–482.

[8] Maguire, B.A.; Pearson, E.S. and Wynn, A.H.A. (1952). The time intervals between industrial accidents, *Biometrika*, **39**, 168–180.

[9] Matthews, D.E. and Farewell, V.T. (1982). On testing for a constant hazard against a change-point alternative, *Biometrics*, **38**, 463–468.

[10] Matthews, D.E.; Farewell, V.T. and Pyke, R. (1985). Asymptotic score statistic processes and tests for constant hazard against a change-point alternative, *Annals of Statistics*, **13**, 2, 583–591.

[11] Raftery, A.E. and Akman, V.E. (1986). Bayesian analysis of a Poisson process with a change-point, *Biometrika*, **73**, 1, 85–89.

[12] Raftery, A.E. (1996). *Hypothesis testing and model selection.* In "Markov Chain Monte Carlo in Practice" (W.R. Gilks, S. Richardson and D.J. Spiegelhalter, Eds.), Chapman & Hall, London, 163–187.

[13] Ruggeri, F. and Sivaganesan, S. (2005). On modelling change-point in non-homogeneous Poisson process, *Statistical Inference for Stochastic Processes*, **8**, 311–329.

[14] Spiegelhalter, D.J.; Thomas, A. and Best, N.G. (1999). *WinBugs: Bayesian Inference Using Gibbs Sampling*, Cambridge: MCR Biostatistics Unit.

[15] West, R.W. and Ogden, R.T. (1997). Continous-time estimation of a change-point in a Poisson process, *Journal of Statistical Computational Simulation*, **56**, 293–302.

REVSTAT – STATISTICAL JOURNAL

Background

Statistical Institute of Portugal (INE), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT – STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

All research articles will be refereed by at least two persons, one from the Editorial Board and another, external.

The only working language allowed will be English.

Three volumes are scheduled for publication, one in March, one in June and the other in November.

On average, four articles will be published per issue.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract/indexed in

REVSTAT is expected to be abstracted/indexed at least in *Current Index to Statistics*, *Mathematical Reviews*, *Statistical Theory and Method Abstracts*, and *Zentralblatt für Mathematic*.

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

— By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

— By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Submitted manuscripts (text, tables and figures) should be typed only in black, on one side, in double spacing, with a left margin of at least 3 cm and not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style.

This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to *PC Windows System* (Zip format), *Mackintosh, Linux* and *Solaris Systems* (StuffIt format), and *Mackintosh System* (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's Website: http://www.ine.pt/revstat.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: liliana.martins@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Also send always the final paper version to:

Maria José Carrilho
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright and Reprints

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, the INE, in order to ensure the widest possible dissemination of information, namely through the National Statistical Institute's Website (http://www.ine.pt).

After assigning the transfer copyright form, authors may use their own material in other publications provided that the REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

Authors of articles published in the REVSTAT will be entitled to one free copy of the respective issue of the Journal and twenty-five reprints of the paper are provided free. Additional reprints may be ordered at expenses of the author(s), and prior to publication.