A Prediction Error Criterion for Choosing the Lower Quantile in Pareto Index Estimation

Debbie J. Dupuis  Maria-Pia Victoria-Feser

First version: July 2003
Current version: July 2004

This research has been carried out within the NCCR FINRISK project on "Financial Econometrics for Risk Management".
A PREDICTION ERROR CRITERION
FOR CHOOSING THE LOWER QUANTILE
IN PARETO INDEX ESTIMATION

Debbie J. DUPUIS
Maria-Pia VICTORIA-FESER
A Prediction Error Criterion
for Choosing the Lower Quantile
in Pareto Index Estimation

by

Debbie J. Dupuis
Department of Statistical and Actuarial Sciences
University of Western Ontario
London, Ontario
N6A 5B7 Canada.
email: dupuis@stats.uwo.ca, fax:(519) 661-3813

and

Maria-Pia Victoria-Feser
Faculty of Psychology and Education
Faculty of Economics and Social Sciences (HEC)
University of Geneva
40, Bd du Pont d’Arve
CH-1211 Geneva 4
email: Maria-Pia.VictoriaFeser@hec.unige.ch, fax: +41 22 705 8104

July 2003

ACKNOWLEDGEMENT

The first author acknowledges ongoing grant support from the Natural Sciences and Engineering Research Council of Canada. The second author acknowledges the financial support of the Swiss National Fund (grant no 610-057883.99 and the NCCR FINRISK).
A Prediction Error Criterion
for Choosing the Lower Quantile
in Pareto Index Estimation

SUMMARY

Successful estimation of the Pareto tail index from extreme order statistics relies heavily on the procedure used to determine the number of extreme order statistics that are used for the estimation. Most of the known procedures are based on the minimization of (an estimate of) the asymptotic mean square error of the Hill estimator for the Pareto tail index. The principal drawback of these approaches is that they involve the estimation of nuisance parameters, and therefore lead to complicated selection procedures. Instead, we propose to use Pareto quantile plots and build a prediction error estimator. The latter depends on the quantile at which the data are truncated, and it is minimized to find the optimal number of extreme order statistics used for estimating the Pareto tail index. The main advantages of the new approach are computational simplicity and the absence of estimation of nuisance parameters. The prediction error estimator is actually a generalization of Mallows’ $C_p$ to non-normal regression models. Through a simulation study involving several data generating models, we show that our prediction error criterion performs very well in terms of MSE compared to other procedures. The proposed method is also successfully applied to alluvial diamond deposits, finance, income, and insurance data.

KEY WORDS: Hill estimator, Mallows’ $C_p$, extreme order statistics, influence function
1 INTRODUCTION

Whether it is in insurance, finance, economics, engineering, or environmental issues, there is much interest in managing risk. An important aspect of the latter is the analysis, modeling, and prediction of extreme events. Over the last twenty years, many innovative techniques based on stochastic extreme value theory have been devised in an attempt to draw reasonable inferences on extreme events based only on a limited amount of data. It is easily argued that only the extreme large (or small) data should be used in tail modeling, however the selection of a threshold above (below) which ‘tail behavior’ of the data can be justified is not straightforward. Selecting a threshold too close to the central data will cause bias and selecting too extreme a threshold will yield large variances for any resulting estimates. Here, we focus on estimation of the upper-tail, but results apply to the lower-tail after proper relabelling.

Let \( X_1, X_2, \ldots, X_n \) be a sequence of positive independent and identically distributed random variables, each with distribution function \( F \). We are interested here in the upper tail of the distribution, namely \( X_{[n-k]}, X_{[n-k+1]}, \ldots, X_{[n]} \) with \( X_i \) denoting the \( ith \) order statistic. Our interest lies in Pareto type tails and we suppose that for sufficiently large quantiles \( x \), \( F \) is such that there exists a positive constant \( \theta \) for which
\[
1 - F(x) = x^{-\theta} l(x)
\]
where \( l(x) \) is a so-called slowly varying function at infinity (see e.g. Beirlant et al. 1996b).

Estimation of the Pareto index \( \theta > 0 \) (or \( 1/\theta \)) has received considerable attention in extreme value statistics. Hill (1975) proposed an estimator based on maximum likelihood arguments. While other estimators have been proposed, see Csörgő and Viharos (1998) for a literature review, Hill’s estimator remains the most popular for Pareto tails. All of the proposed estimators are based on the assumption that \( k \), the number of observations in the upper-tail to be included, is known. In practice, \( k \) is unknown. A general approach in determining \( k \) is the minimization of the asymptotic mean squared error (AMSE) of the Hill estimator. Beirlant et al. (1996b) obtain an estimate for the optimal \( k \) by minimizing a nonparametric estimate of the AMSE of

3
the Hill estimator for $1/\theta$. Beirlant, Dierckx, Goegebeur, and Matthys (1999) choose $k$ to minimize a plug-in estimator of the theoretical expression of AMSE. Hall (1990) and Danielsson, de Haan, Peng, and de Vries (2001) use subsample bootstrapping to estimate the AMSE of the Hill estimator. In a slightly different approach, Beirlant, Dierckx, Guillou, and Stărică (2002) find an asymptotic representation of $k_{n,\text{opt}}$ and an estimator for $k_{n,\text{opt}}$ is derived based on that representation. All of these approaches, along with those proposed by Drees and Kaufmann (1998) and Guillou and Hall (2001), are somewhat complicated in that they require the estimation of nuisance parameters. Moreover, as pointed out by Beirlant et al. (1996b), $\hat{\theta}$ is a very volatile function of $k$, which makes the optimization procedure very difficult numerically.

The remainder of the paper is organized as follows. In Section 2, we review the Pareto approach. In Section 3, a general formulation for the prediction error is presented and then developed more precisely for our Pareto approach. This leads to our criterion for determining $k$. In Section 4, we compare our criterion to other existing methods from a theoretical viewpoint. A small simulation study is presented in Section 5 and applications to alluvial diamond deposits, finance, income, and insurance data are presented in Section 6. Finally, Section 7 contains a few closing remarks.

## 2 PARETO APPROACH

In the context of income distribution, Pareto (1896) analyzed the characteristic of regularity in the upper tail of the observed income distribution. He found empirically that there was a decreasing linear relationship between the income-power (i.e. the logarithm of the income variable) and the logarithm of the backward cumulative distribution function. Formally, the relation observed by Pareto is given by

$$\log \left(1 - F_{(n)}(x)\right) = \beta - \theta \log \left(\frac{x}{x_0}\right)$$

(1)

with all $x > x_0$ and $F_{(n)}$ the empirical distribution in the upper-tail. Note that since $1 - F_{(n)}(x_0) = 1$ and $\log \left(\frac{x}{x_0}\right) |_{x = x_0} = 0$, then $\beta = 0$. This lead him to propose the
so-called Pareto model

\[ F_\theta(x) = 1 - \left( \frac{x}{x_0} \right)^{-\theta} \]  \hspace{1cm} (2)

with density

\[ f(x; \theta) = \theta x^{-(\theta+1)} x_0^\theta, \quad x \geq x_0. \]  \hspace{1cm} (3)

Given a sample from the upper-tail, e.g. say \( X_{[n-k+1]}, \ldots, X_{[n]} \) from Section 1, the maximum likelihood estimator (MLE) of \( \theta \) is the Hill estimator

\[ \hat{\theta} = \left[ \frac{1}{k} \sum_{i=1}^{k} \log X_{[n-i+1]} - \log X_{[n-k]} \right]^{-1}. \]  \hspace{1cm} (4)

We also have that \( \sqrt{n}(\hat{\theta} - \theta) \) is asymptotically normal with mean 0 and variance \( \theta^2 \).

As it is often done, the Pareto model can be viewed as a regression model (see e.g. Beirlant et al. 1996b). Indeed, first let (for notational convenience) \( X_{[i]}^*, i = 1, \ldots, k \), be the ordered largest \( k \) observations. We suppose that \( X_{[i]}^* \) has probability density function as in (3), with cumulative distribution function in (2) so that

\[ F_\theta^{-1}(q) = x_0 (1 - q)^{-1/\theta}. \]

Let

\[ Q(F; q) = \inf \{x | F(x) \geq q\} \]

so that

\[ X_{[i]}^* = Q(F_{[n]}; i/(k + 1)) \]

The relationship (1) can be written as

\[ \log \left( \frac{x}{x_0} \right) = \frac{1}{\theta} \log \left( 1 - F_{[n]}(x) \right) \]  \hspace{1cm} (5)

and at \( x = X_{[i]}^*, i = 1, \ldots, k \) we get

\[ \log \left( \frac{Q(F_{[n]}; i/(k + 1))}{x_0} \right) = \frac{1}{\theta} \log \left( \frac{k + 1 - i}{k + 1} \right), \quad i = 1, \ldots, k. \]  \hspace{1cm} (6)

The plot of \( \log \left( X_{[i]} \right) \) versus \( -\log \left( \frac{n+1-i}{n+1} \right), i = 1, \ldots, n \) is often used to detect graphically the quantile \( X_{[i]} \) above which the Pareto relationship is valid, i.e. the point above which the plot yields a straight line. In Section 3 we use this relationship in order to develop a prediction error criterion for choosing the value of \( x_0 \), and hence indirectly the value of \( k \).
3 PREDICTION ERROR CRITERION

We first review prediction error in a general context, then apply it to the special case of the Pareto model.

3.1 General Formulation

Let \( Y = (Y_1, \ldots, Y_n)^T \) be a random sample of observations from a distribution \( F \). Consider the most general case of \( Y \) having covariance \( \Sigma \). Given a model, let \( \hat{Y} = (\hat{Y}_1, \ldots, \hat{Y}_n)^T \) be the predicted values for \( Y \). One can define a prediction error criterion as

\[
\Gamma = \frac{1}{n} \text{tr} \left\{ \mathbf{E} \left[ \Sigma^{-1} (\hat{Y}_i - \mathbf{E}[Y_i]) (\hat{Y}_i - \mathbf{E}[Y_i])^T \right] \right\}.
\]

(7)

In the case of independent observations, (7) reduces to

\[
\Gamma = \frac{1}{n} \sum_{i=1}^{n} \mathbf{E} \left[ \left( \frac{\hat{Y}_i - \mathbf{E}[Y_i]}{\sigma_i} \right)^2 \right],
\]

(8)

where \( \sigma_i^2 = \text{var}(Y_i) \). In many situations (7) is numerically intractable given an \( n \times n \) covariance matrix \( \Sigma \). Therefore, even when observations are dependent, one might resort to (8) as a prediction error criterion.

We seek an unbiased estimator of \( \Gamma \) and find it helpful to rewrite (8). Let

\[
\delta_i = \hat{Y}_i - \mathbf{E}[Y_i], \quad r_i = Y_i - \hat{Y}_i, \quad \text{and} \quad RSS = \sum_{i=1}^{n} r_i^2
\]

so that we now have

\[
n \Gamma = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \mathbf{E} \left[ \delta_i^2 \right] = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \text{var} \left[ \hat{Y}_i \right] + \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \mathbf{E} \left[ \delta_i^2 \right].
\]

(9)

We know that

\[
\mathbf{E}[RSS] = \sum_{i=1}^{n} \mathbf{E} \left[ r_i^2 \right] = \sum_{i=1}^{n} \text{var} \left[ r_i \right] + \sum_{i=1}^{n} \mathbf{E} \left[ r_i^2 \right]
\]

and since

\[
\sum_{i=1}^{n} \mathbf{E} \left[ r_i^2 \right] = \sum_{i=1}^{n} \mathbf{E} \left[ \delta_i^2 \right]
\]

6
we easily see that
\[ \sum_{i=1}^{n} \mathbb{E} [\delta_i]^2 = \mathbb{E} [R.S.S] - \sum_{i=1}^{n} \text{var} [r_i]. \]  
(10)

So, substituting (10) into (9), we find
\[
n\Gamma = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \text{var} [\hat{Y}_i] + \mathbb{E} \left[ \sum_{i=1}^{n} \frac{1}{\sigma_i^2} (Y_i - \hat{Y}_i)^2 \right] - \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \text{var} [r_i]
\]
\[
= \mathbb{E} \left[ \sum_{i=1}^{n} \frac{1}{\sigma_i^2} (Y_i - \hat{Y}_i)^2 \right] + \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \text{var} [\hat{Y}_i] - \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \text{var} [Y_i - \hat{Y}_i]
\]
\[
= \mathbb{E} \left[ \sum_{i=1}^{n} \left( \frac{Y_i - \hat{Y}_i}{\sigma_i} \right)^2 \right] + 2 \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \text{cov} [Y_i, \hat{Y}_i] - n
\]
and choose
\[ C = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \hat{Y}_i}{\sigma_i} \right)^2 + 2 \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \text{cov} [Y_i, \hat{Y}_i] - 1. \]  
(11)

with suitable estimators for \( \sigma_i^2 \) and \( \text{cov} [Y_i, \hat{Y}_i] \) as an estimator of \( \Gamma \) with relative bias to \( O(n^{-3/2}) \).

3.2 A prediction error criterion for the Pareto Model

The following theorem gives an estimator of the prediction error \( \Gamma \) for the Pareto model, by making use of the regression type relationship given in (6).

**Theorem 1.** Let
\[ Y_i = \log \left( \frac{X_{ik}}{x_0} \right) = \log \left( \frac{Q(F_{i(k); i/(k + 1)})}{x_0} \right) \]  
(12)

and for the predicted values
\[ \hat{Y}_i = -\frac{1}{\theta} \log \left( \frac{k + 1 - i}{k + 1} \right) \]  
(13)
where \( \hat{\theta} \) is the Hill estimator

\[
\hat{\theta} = \left[ \frac{1}{k} \sum_{i=1}^{k} \log X_{i}^{*} - \log x_{0} \right]^{-1},
\]

an estimator of \( \Gamma \) is given by

\[
C = \hat{\theta}^{2} \frac{1}{k} \sum_{i=1}^{k} \frac{(k+1-i)}{i} \left[ \log \left( \frac{X_{i}^{*}}{x_{0}} \right) + \frac{1}{\hat{\theta}} \log \left( \frac{k+1-i}{k+1} \right) \right]^{2} + \frac{2}{k} \sum_{i=1}^{k} \frac{(k+1-i)}{i} \log \left( \frac{k+1-i}{k+1} \right) - 1.
\]

(14)

**Proof.** Substituting the values for \( Y_{i} \) and \( \hat{Y}_{i} \) in (11), we obtain

\[
C = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{\sigma_{i}^{2}} \left( \log \left[ \frac{Q(F_{n}; i/(k+1))}{x_{0}} \right] + \frac{1}{\hat{\theta}} \log \left[ \frac{k+1-i}{k+1} \right] \right)^{2} + \frac{2}{k} \sum_{i=1}^{k} \frac{1}{\sigma_{i}^{2}} \text{cov} \left( Y_{i}, \hat{Y}_{i} \right) - 1.
\]

(15)

One can use an influence function approach to compute the required covariances. It is shown in the Appendix that

\[
\sigma_{i}^{2} = \frac{1}{\hat{\theta}^{2}} \frac{i}{k+1-i}
\]

(16)

and

\[
\text{cov} \left( Y_{i}, \hat{Y}_{i} \right) \approx \frac{1}{\theta^{2}} \log \left( \frac{k+1-i}{k+1} \right)^{2}.
\]

(17)

Substituting (16) and (17) into (15), replacing \( \theta \) by the Hill estimator in (4), one obtains the estimated prediction error (14).

We propose to choose \( x_{0} \), and thus indirectly \( k \), so as to minimize (14) and refer to this as the \( C \)-criterion.

## 4 OTHER CRITERIA

As stated in Section 1, other approaches have been suggested to determine \( k \). Here, we briefly explain some of these methods and draw comparisons with our \( C \)-criterion where appropriate.

Beirlant et al. (1996b) obtain an estimate for the optimal \( k \) by minimizing a nonparametric estimate of the AMSE of the Hill estimator for \( 1/\theta \). The weighted MSE
expression minimized is *(in our notation)*

\[
\text{MSE}_{\text{opt}}(k) = \frac{1}{k} \sum_{i=1}^{k} w_{i,k}^{\text{opt}} \left( \log \frac{X_{[i]}^*}{x_0} + \frac{1}{\theta} \log \left( \frac{k + 1 - i}{k + 1} \right) \right)^2
\]

for some sequence of weights \(w_{i,k}^{\text{opt}}\) which depends on \(\rho\), a non-positive index that is assumed to characterize the slowly varying function and which must also be estimated. The index \(\rho\) is also estimated nonparametrically and thus estimation of the AMSE, and the optimal \(k\), is an iterative procedure. Note that the factor \(1/\hat{\theta}\) is the estimate of \(1/\theta\) obtained by using the Hill estimator with \(k\) observations. Beirlant et al. (1996b) rely on probabilistic deductions to establish optimal weights \(w_{i,k}^{\text{opt}}\). Essentially, we have

\[
w_{i,k}^{\text{opt}} = \delta_{1,k} w_{i,k}^{(1)} + \delta_{2,k} w_{i,k}^{(2)}
\]

where \(\delta_{1,k}\) and \(\delta_{2,k}\) are scaling constants that depend on \(k\) and \(\rho\), and \(w_{i,k}^{(1)}\) and \(w_{i,k}^{(2)}\) are chosen weight functions. Beirlant et al. (1996b) used \(w_{i,k}^{(1)} = 1\) and \(w_{i,k}^{(2)} = (k + 1 - i)/(k + 1)\). The values of \(\delta_{1,k}\) and \(\delta_{2,k}\) must be obtained numerically as even for simple choices of \(w_{i,k}\), and assuming a fixed value of \(\rho\), we cannot get an analytical expression. However, \(\delta_{1,k}\) and \(\delta_{2,k}\) will be functions of \(k\), and we keep this notation in an attempt to compare \(\text{MSE}_{\text{opt}}(k)\) with the \(C\) in (14). We thus have

\[
\text{MSE}_{\text{opt}}(k) = \frac{1}{k} \sum_{i=1}^{k} \left( \delta_{1,k} + \delta_{2,k} \left( \frac{k + 1 - i}{k + 1} \right) \right) \left( \log \frac{X_{[i]}^*}{x_0} + \frac{1}{\theta} \log \left( \frac{k + 1 - i}{k + 1} \right) \right)^2
\]

\[
= \frac{\delta_{2,k}}{k(k + 1)} \sum_{i=1}^{k} \left( k + 1 - i \right) \left( \log \frac{X_{[i]}^*}{x_0} + \frac{1}{\theta} \log \left( \frac{k + 1 - i}{k + 1} \right) \right)^2 + \frac{\delta_{1,k}}{k} \sum_{i=1}^{k} \left( \log \frac{X_{[i]}^*}{x_0} + \frac{1}{\theta} \log \left( \frac{k + 1 - i}{k + 1} \right) \right)^2
\]

and the following may be noted about the first term in \(\text{MSE}_{\text{opt}}(k)\): 1) the sum is scaled by \(\delta_{2,k}/k(k + 1)\), an implicit function of the estimated AMSE(\(1/\hat{\theta}\), compared to \(\hat{\theta}^2/k\) for \(C\); and 2) observations in increasing distance from \(x_0\) are not as severely downweighted as in \(C\) with \(k + 1 - i\) versus \((k + 1 - i)/i\). An expansion of the second term in \(\text{MSE}_{\text{opt}}(k)\) leads to

\[
\frac{\delta_{1,k}}{k} \frac{1}{\hat{\theta}^2} \sum_{i=1}^{k} \log \left( \frac{k + 1 - i}{k + 1} \right)^2 + \frac{\delta_{1,k}}{k} \sum_{i=1}^{k} \log \left( \frac{X_{[i]}^*}{x_0} \right)^2 + \frac{2\delta_{1,k}}{k(\theta)} \sum_{i=1}^{k} \log \left( \frac{X_{[i]}^*}{x_0} \right) \log \left( \frac{k + 1 - i}{k + 1} \right)
\]

(19)
The following may be noted about the first term in (19): 1) the sum is scaled by $\delta_{1,k} \, k \hat{\theta}^2$ versus the simple $2/k$ in the second term in $C$; and 2) the squared log terms all have weight 1 versus the variable weights $(k + 1 - i)/i$ in $C$. Note that the variable weights in $C$ are such that weights decrease with increased size in the contribution from the squared log term. The third term in $C$ is the constant -1. From (19) it is easily seen that $\text{MSE}_{\text{opt}}(k)$ is somewhat more complex with the presence of its remaining terms

$$\frac{\delta_{1,k}}{k} \sum_{i=1}^{k} \log \left( \frac{X_i^s}{x_0} \right)^2 + 2 \frac{\delta_{1,k}}{k \hat{\theta}} \sum_{i=1}^{k} \log \left( \frac{X_i^s}{x_0} \right) \log \left( \frac{k + 1 - i}{k + 1} \right).$$

Now minimizing the latter is minimizing

$$\frac{\delta_{1,k}}{k} \left( \sum_{i=1}^{k} Y_i^2 - 2 \sum_{i=1}^{k} Y_i \hat{Y}_i \right)$$

in our regression setting. The direct impact of the latter is unclear. More complex weight functions and optimal weighting schemes were introduced in Beirlant et al. (1996a). The more complex weighting scheme, which does not always produce better results, makes it such that we cannot carry out a term by term analytical comparison as above.

A theoretical comparison of our $C$-criterion to other recently suggested methods for selecting $k$ is not so easily derived. We describe these other methods briefly here and include them in a simulation study in Section 5. Beirlant, Dierckx, Guillou, and Stărică (2002) find an asymptotic representation of $k_{n,\text{opt}}$ and derive an estimator for $k_{n,\text{opt}}$ based on that representation. The automatic method for selection is somewhat ad-hoc as it is really the median of $\hat{k}_{n,k_0}$ for $k_0 = 3, \ldots, n/2$ that is the recommended threshold. The choice is said to be practical, but is not justified mathematically. The estimator requires consistent estimators for $\rho$, $1/\theta$, and $g((n + 1)/(k_0 + 1))$, where $g$ is a rate function satisfying $g(x) \to 0$ as $x \to \infty$ characterizing the slowly-varying function $l$. Least squares estimators based on regression models with additive noise are used. Both Hall (1990) and Danielsson, de Haan, Peng, and de Vries (2001) use subsample bootstrapping to estimate the MSE of the Hill estimator. Drees and Kaufmann (1998) present a sequential procedure, based on ‘stopping times’ for the sequence $H_{k,n}$ of Hill
estimators that are asymptotically equivalent to a deterministic sequence, to select the optimal $k_{n,opt}$. Beirlant, Dierckx, Goegebeur, and Matthys (1999) choose $k$ to minimize a plug-in estimator of the theoretical expression of AMSE of the Hill estimator. A survival regression model with exponential responses is used to obtain MLE of $1/\theta$, $\rho$, and $\hat{g}_{n,k}$, the expected value of $g(U_{k+1,n}^{-1})$ where $U_{k+1,n}$ denotes the $(k + 1)^{th}$ order statistic of a uniform$(0,1)$ sample of size $n$. MLE are the plug-in values. Matthys and Beirlant (2003) present a refinement of Beirlant et al. (1999) where $\theta > 0$ is no longer required and terms of second order regular variation are considered. Guillou and Hall (2001) propose to choose $H_{k,n}$ where $k$ is the smallest value of $k \in [n^a, n^b]$ for which a scaled least squares estimate of $g((n + 1)/(k + 1))$ is larger than a critical value.

5 SIMULATIONS

In order to evaluate the usefulness of the $C$-criterion, we carry out a simulation study and compare results with those of others. The computational complexity of other selection procedures make it such that they could not easily be implemented for comparison. Beirlant et al. (2002) carried out an extensive simulation study and we will refer to some of their results, along with those of Beirlant et al. (1996b) and Matthys and Beirlant (2003), here. More specifically, methods compared to the $C$-criterion are:

- Method 1 - Beirlant et al. (1996b)
- Method 2 - Beirlant et al. (2002)
- Method 5 - Guillou and Hall (2001)

The root mean squared error (RMSE) of $\hat{\theta}^{-1}$ is reported. All results for the $C$-criterion are based on 200 replications and others are as stated below. Note that the $C$-criterion should further outperform these asymptotic-based criteria in smaller sample sizes, but the smallest sample size previously considered is $n = 500$. 

11
5.1 Simulations for the Burr Distribution

Consider a Burr distribution parametrized as

\[ F(x) = 1 - (1 + x^{-\rho})^{1/\rho} \]

for some parameter \( \rho < 0 \). Beirlant et al. (1996b) consider \( \rho = -0.5, -1, \) and \(-1.5 \) (all leading to \( \theta = 1 \)) and sample sizes of \( n = 500, 1000, \) and \( 1500 \). Their simulation results are based on 200 replications. Results based on 100 replications were reported by Beirlant et al. (2002) for Methods 2-5 and a sample size of \( n = 500 \). All these results, along with those for the \( C \)-criterion, are listed in Table 1. We can see that the \( C \)-criterion does quite well, especially for \( n = 500 \) and/or \( \rho = -1 \) and \(-1.5 \).

Table 1 here

5.2 Simulations for other Distributions

Beirlant et al. (2002) also consider

- a Fréchet distribution \( F(x) = \exp(-x^{-2}) \), so that \( \theta = 2 \);
- a Student \( t_6 \) distribution, so that \( \theta = 6 \); and
- a Loggamma distribution with density \( f(x) = x^{-2} \log(x) \), so that \( \theta = 1 \).

for Methods 2-5. Their results are for samples of size \( n = 500 \) and are based on 100 replications. All these results, along with those for the \( C \)-criterion, are listed in Table 2. Here, results are mixed with the Loggamma being somewhat problematic. It could be argued that with these distributions, in particular the Loggamma and probably also the \( t_6 \), the Pareto distribution is not a suitable model for the upper tail of the distribution. We discuss this further below.

Slight variations on the above are considered in Matthys and Beirlant (2003) for Method 6 and samples of size \( n = 500 \). These results, based on 100 replications, are included in Table 3. Again here, results are mixed, however there is a noticeable improvement over Method 6 for the very thick-tailed \( |t_2| \).
Tables 2 and 3 here

It is helpful to have a closer look at the tails of the distributions considered in Tables 2 and 3 in order to understand the performance of our $C$-criterion. Recall that our $C$-criterion is based on a Pareto model and thus its use for distributions that lie far away from this model will likely lead to a poor performance. We consider the index

$$
\tau_p = \frac{(G^{-1}(0.999) - G^{-1}(0.95))/(G^{-1}(0.99) - G^{-1}(0.95))}{(F^{-1}(0.999) - F^{-1}(0.95))/(F^{-1}(0.99) - F^{-1}(0.95))}
$$

for a distribution $G$ where $F$ is Pareto as in (2) with the corresponding value of $\theta$ for $G$. This index is in the spirit of the $\tau$ index for normal tails (Hoaglin et al. 1983), but is pushed deeper in the tail given our interest. Values of the tail index are given in Table 4. It is easily seen that the performance of the $C$-criterion may be explained by the distance of the true model from the Pareto model. Better performance could be obtained from a robust criterion and more discussion appears in Section 7.

5.3 Simulations for mixture Distributions

In order to have examples with a relatively clear distinction between the Pareto upper-tail and the rest of the distribution, samples of size $n = 1000$ from a uniform/Pareto mixture for the upper-tail were also considered. More precisely, we generated $\lfloor an \rfloor$ data (where $\lfloor x \rfloor$ denotes the integer part of $x$) from a Pareto distribution as in (2) and $n - \lfloor an \rfloor$ data from a uniform distribution with density $(1 - \alpha)/x_0$. Parameter settings were $\alpha = 0.1$, $x_0 = 2.5$, and $\theta = 2.5$. A simulated sample is shown in Figure 1, along with a Pareto regression plot. One can see that the line bends quite abruptly around $x_0 = 2.75$. Ideally, a $k$-selection criterion would properly capture the break. The behaviour of the $C$-criterion for this sample is presented in Figure 2 and one can see that the minimum lies where it was expected.

Figures 1 and 2 here

The boxplots of the parameter estimates from 100 replications of the above mixture are plotted in Figure 3. One can see that the estimator for $\theta$, the parameter of interest, is unbiased, hence showing again the good performance of the $C$-criterion.
6 DATA ANALYSIS

In this section, the newly proposed method is used to draw inferences on data from diverse applications.

6.1 Bougban data

As a first illustration, we consider data on alluvial diamond deposits in Guinea, South Africa. The data set covers a part of the Bougban River and consists of 683 stone sizes. A graphical representation in terms of a histogram and the Pareto-regression plot is given in Figure 4. On the Pareto-regression plot, one identifies a rough and reasonable interval for the optimal value of $x_0$ as $(\exp (-1.75), \exp (0)) \cong (0.17, 1)$. Figure 5 shows the behaviour of the $C$-criterion, as well that of $\hat{\theta}$ as a function of $x_0$, for the latter interval of values. The $C$-criterion shows a minimum at around 0.35. It should be noted that the $C$-criterion function is not very smooth. Although this is somewhat worrisome, one could argue that since the minimization problem is univariate, one can always calibrate graphically the interval in which the minimum lies and then run an optimization procedure. However, a smoother $C$-criterion will be sought in future research.

The values for $x_0$, $k$ and $\hat{\theta}$ provided by the $C$-criterion, are presented in Table 5, along with values found by Caers, Vynckier, Beirlant, and Rombouts (1996). These data were also analysed by Beirlant et al. (1996a) who found them problematic because of the somewhat outlying two largest observations. They had to modify their suggested procedure to obtain good inference, however their estimates are not comparable to the values in Table 5.

A fitted line given by

$$
\log \left( X^{*}_{[i]} \right) = \log \left( x_0 \right) + \hat{\theta}^{-1} \left( -\log \left( \frac{n+1-i}{n+1} \right) + \log \left( \frac{k+1}{n+1} \right) \right)
$$
is added to the Pareto-regression plot. For the $C$-criterion, the resulting fitted line and the corresponding value of $x_0$ is described in Figure 6 by the thickest solid and dashed lines. Our $C$-criterion is not affected by the presence of the somewhat outlying two largest values and yields a good fit directly.

**Table 5 and Figures 4, 5, and 6 here**

### 6.2 Finance data

In finance theory, one important issue is the ability to establish the value at risk (VaR) of an investment (an asset, a portfolio of assets,...etc). The latter can be defined as the level of loss (i.e. a quantile usually in the lower tail of the distribution of returns) on a portfolio which is expected to be equaled or exceeded with a given, (usually small) probability (see e.g. Jorion 1997). Since the returns in the tails of the distribution are sparse, it is therefore important to be able to model them. Empirical studies on the tails of daily log-returns have indicated that a Pareto-type model is usually suitable. As an example, we consider here a series of log-returns in 100% on alternative investments on a monthly basis between January 1997 and December 2002 (i.e. $n = 72$ observations). See Credit Suisse First Boston (CSFB) / Tremont hedge fund index at [www.hedgeindex.com](http://www.hedgeindex.com) and Perret-Gentil and Victoria-Feser (2003) for a more detailed description of the data. We actually take 100 minus the log-return for the evaluation of the downside risk.

A graphical representation in terms of a histogram and the Pareto-regression plot is given in Figure 7. This data are interesting in that the three largest returns are somewhat outlying, but we ignore this for the moment. On the Pareto-regression plot, one can see that a rough and reasonable interval for the optimal value of $x_0$ is $(\exp(4.58), \exp(4.62)) \cong (97.5, 101.5)$. The $C$-criterion shows a minimum at $x_0 = 98.872$, yielding $k = 49$ and $\hat{\theta} = 63.116$. The corresponding fitted Pareto-regression line (Hill Estimator) is given in Figure 8 along with a robust fit that will be explained below. One can see that the fit is satisfactory. The outliers do not seem to affect the choice of $x_0$ nor the estimate $\hat{\theta}$. To further test that, we repeat the analysis with the
three largest returns removed. Values of $x_0 = 98.897$, $k = 46$ and $\hat{\theta} = 81.654$ are found. The corresponding fitted Pareto-regression line appears in Figure 9. The fit is very similar, but the value of $\hat{\theta} = 81.654$ is very different from our original $\hat{\theta} = 63.116$. The value of $x_0$ remained relatively unchanged (and thus apparently not influenced by the extreme data). Note that given $x_0 = 98.872$ and $k = 49$, an optimal bias robust estimate (OBRE) of $\theta$ yields $\hat{\theta} = 65.450$. See Hampel, Ronchetti, Rousseeuw, and Stahel (1986) for the general theory of the OBRE, Victoria-Feser and Ronchetti (1994) for its application to the Pareto model, and Dupuis (1999) for its use in the peaks over threshold approach to extremes. Hill and OBRE fits appear on the Pareto-regression plot in Figure 8. The OBRE does discount the upper three points, but it is unclear if that is appropriate. A better sense of the impact of the fits may be obtained by comparing estimated VaR. The latter appear in Table 6 along with another financial risk measure, the expected shortfall (ES), defined as the mean return in the (usually lower) tail of the distribution, i.e.

$$ES(F; q) = \frac{1}{q} \int_{q(F; q)} x dF(x).$$

For the Pareto modeling of the upper tail, we have

$$F(x) = \begin{cases} 
F_{(n)}(x) & x \leq x_0 \\
1 - \alpha \left( \frac{x}{x_0} \right)^{-\theta} & x \geq x_0
\end{cases}$$

with $\alpha = 1 - \int_{x_0}^{\infty} dF(x)$ (see e.g. Cowell and Victoria-Feser 2000a). The expected shortfall is then

$$ES(F; q) = \frac{\alpha \theta x_0^\theta}{q} \int_{q(F; q)} x^{-\theta} dx = \alpha \frac{\theta}{\theta - 1} \frac{x_0}{q} (1 - q)^{\theta - 1}.$$ 

Note that when we estimate $x_0$ and $k$ based on all the data (as we have done), and subsequently use either the Hill estimator or the OBRE to estimate $\theta$, the estimated VaR and ES remain relatively the same. On the other hand, the change in the estimated quantities is more appreciable when the analysis is performed without the three largest (outlying) observations. A robust procedure for selection, and not just estimation, is therefore needed. This will be developed in future research.

**Table 6 and Figures 7, 8 and 9 here**
6.3 Income data

The data are monthly incomes \((n = 6706)\) in the UK in 1979 (see Department of Social Security 1992, and Cowell and Victoria-Feser 1996 for a more detailed description). With income data, the determination of the value \(x_0\) is an important issue, since traditionally upper tails are sparse and therefore are fitted with Pareto-type distributions for the estimation of inequality measures. The value of \(x_0\) can also be used as a cutting point for a semi-parametric approach to stochastic dominance comparisons (see Cowell and Victoria-Feser 2000a).

A graphical representation in terms of a histogram and the Pareto-regression plot is given in Figure 10. Our \(C\)-criterion finds a minimum at \(x_0 = 266.54 = \exp(5.59)\), and yields \(k = 981\) and of \(\hat{\theta} = 4.429\). The corresponding fitted Pareto-regression line appears in Figure 11. The fit is quite good.

Figures 10 and 11 here

6.4 Danish fire data

A well-known data set, originally provided by Mette Rytgaard of Copenhagen Re, the data are major Danish fire total losses from 1980 to 1990. Amounts are in millions of Danish Kroner adjusted for inflation (1985 prices) and appear as a resident data set in the \texttt{finmetrics} module of \texttt{Splus}. Our \(C\)-criterion reveals a clear minimum at \(x_0 = 1.376\), yielding \(k = 1562\) and \(\hat{\theta} = 1.402\). The corresponding fitted Pareto-regression line is given in Figure 12. The fit is good and does reconcile the variability in the upper-tail quite well.

Figure 12

7 CONCLUDING REMARKS

We have provided a methodology to determine the number of extreme order statistics to use in the estimation of the Pareto tail index. Our \(C\)-criterion is simple and performs well both in simulations and real data examples. It can be extended to more
complicated models (e.g., generalized Pareto) and more complicated estimators (e.g., OBRE). The criterion is not as smooth as we would like in real data examples, but solving a univariate minimization problem is easy. Establishing a smoothing procedure is the object of further research. Also, as shown by simulations away from the Pareto model in Section 5 and the example in Section 6.2, a robust procedure for selection, and not just estimation (e.g., as provided by the OBRE), is needed. As the $C$-criterion stems from a prediction error approach, a robust criterion should be possible by making use of robustness results obtained in the general area of prediction error. For example, Ronchetti and Staudte (1994) proposed a robust version of Mallows’ $C_p$. A robust $C$-criterion will be pursued.
APPENDIX

Consider the influence function (IF) defined for a statistic \( \hat{\alpha} \) as

\[
IF(z; \hat{\alpha}, F_\theta) = \left. \frac{\partial}{\partial \varepsilon} \hat{\alpha} (F_\varepsilon) \right|_{\varepsilon = 0}
\]

where

\[
F_\varepsilon = (1 - \varepsilon) F_\theta + \varepsilon \Delta_z
\]

and \( \Delta_z \) is the probability measure that puts mass 1 on the point \( z \). The IF can be used to compute variances and covariances between two statistics \( \hat{\alpha} \) and \( \hat{\beta} \). We get up to \( O(1/n) \)

\[
\text{cov} (\hat{\alpha}, \hat{\beta}) \simeq \int IF(z; \hat{\alpha}, F_\theta)IF(z; \hat{\beta}, F_\theta) dF_\theta(z),
\]

see e.g. Hampel et al. 1986.

First, we seek the variance of \( Y_i \) as defined in (12). The IF of \( Y_i \) is

\[
IF(z; Y_i, F_\theta) = \left. \frac{\partial}{\partial \varepsilon} \left[ \log \left( \frac{Q(F_\theta; i/(k + 1))}{x_0} \right) \right] \right|_{\varepsilon = 0}
\]

\[
= \frac{1}{Q(F_\theta; i/(k + 1))} \left. \frac{\partial}{\partial \varepsilon} Q(F_\varepsilon; i/(k + 1)) \right|_{\varepsilon = 0}
\]

\[
= \frac{i/(k + 1) - \iota(Q(F_\theta; i/(k + 1)) \geq z)}{Q(F_\theta; i/(k + 1)) f(Q(F_\theta; i/(k + 1)), \theta)}
\]

where we have used the Cowell and Victoria-Feser (2000b) result

\[
IF(z; Q(\cdot, q), F_\theta) = \frac{q - \iota(Q(F_\theta; q) \geq z)}{f(Q(F_\theta; q), \theta)}
\]

where \( \iota \) is the indicator function. The variance is then

\[
\sigma_i^2 = \int_{x_0}^\infty \left( \frac{i/(k + 1) - \iota(Q(F_\theta; i/(k + 1)) \geq z)}{Q(F_\theta; i/(k + 1)) f(Q(F_\theta; i/(k + 1)), \theta)} \right)^2 dF_\theta(z)
\]

\[
= \left[ Q(F_\theta; i/(k + 1))^2 f(Q(F_\theta; i/(k + 1)), \theta) \right]^{-1} \times \left[ (i/(k + 1))^2 - 2(i/(k + 1)) \int_{x_0}^{Q(F_\theta; i/(k + 1))} dF_\theta(z) + \int_{x_0}^{Q(F_\theta; i/(k + 1))} dF_\theta(z) \right].
\]

We have

\[
\int_{x_0}^{Q(F_\theta; i/(k + 1))} dF_\theta(z) = 1 - (Q(F_\theta; i/(k + 1))/x_0)^{-\theta},
\]

19
\[ Q(F_0; i/(k + 1)) = x_0 \left( 1 - i/(k + 1) \right)^{-1/\theta}, \]

and

\[ Q(F_0; i/(k + 1)) f(Q(F_0; i/(k + 1)), \theta) = \theta \left( 1 - i/(k + 1) \right), \]

so that

\[ \int_{x_0}^{Q(F_0; i/(k + 1))} dF_0(z) = i/(k + 1). \]

Finally, from the latter, it is easily shown that

\[ \sigma_i^2 = \frac{1}{\theta^2} \frac{i}{(k + 1 - i)}, \]

which is equation (16) in the main text.

Secondly, we seek \( \text{cov}(Y_i, \hat{Y}_i) \) and thus require the IF of \( \hat{Y}_i \). Since

\[
IF(z; \hat{Y}_i, F_0) = \frac{\partial}{\partial \varepsilon} \left[ -\frac{1}{\theta(F_0)} \log \left( \frac{k + 1 - i}{k + 1} \right) \right]_{\varepsilon=0}
\]

we will also need \( IF(z; \hat{\theta}, F_0) \). We know that

\[ IF(z; \hat{\theta}, F_0) = M(\theta)^{-1} s(z, \theta) \]

where

\[
s(z, \theta) = \frac{\partial}{\partial \theta} \log(f(z; \theta)) = \frac{1}{\theta} + \log x_0 - \log z
\]

and

\[
M(\theta) = \int_{x_0}^{\infty} s(z, \theta)^2 dF_0(z)
\]

\[ = \int_{x_0}^{\infty} \left( \frac{1}{\theta} + \log x_0 - \log z \right)^2 \theta x_0^\theta z^{-\theta-1} dz
\]

\[ = \left( \frac{1}{\theta} + \log x_0 \right)^2 \theta x_0^\theta \int_{x_0}^{\infty} z^{-\theta-1} dz
\]

\[ - 2 \left( \frac{1}{\theta} + \log x_0 \right) \theta x_0^\theta \int_{x_0}^{\infty} \log(z) z^{-\theta-1} dz
\]

\[ + \theta x_0^\theta \int_{x_0}^{\infty} (\log z)^2 z^{-\theta-1} dz. \]

Integrals are straightforward and subsequent simplifications yield

\[ M(\theta) = \frac{1}{\theta^2} \]

20
so that

\[ IF(z; \hat{\theta}, F_\theta) = \theta^2 \left( \frac{1}{\theta} + \log x_0 - \log z \right) \]

and

\[ IF(z; \hat{Y}_i, F_\theta) = \log \left( \frac{k + 1 - i}{k + 1} \right) \left( \frac{1}{\theta} + \log x_0 - \log z \right) . \]

Thus, we now have

\[
\text{cov} \left( Y_i, \hat{Y}_i \right) = \int_{x_0}^{\infty} \frac{i/(k + 1) - i(Q(F_\theta; i/(k + 1)) \geq z)}{Q(F_\theta; i/(k + 1)) f(Q(F_\theta; i/(k + 1)), \theta)} \log \left( \frac{k + 1 - i}{k + 1} \right) \left( \frac{1}{\theta} + \log x_0 - \log z \right) \theta x_0^\theta z^{-\theta - 1} dz
\]

\[= \frac{\log \left( \frac{k + 1 - i}{k + 1} \right)}{\theta} \left( \frac{i}{k + 1} \right) \frac{1}{\theta} + \frac{i}{(k + 1) \log x_0}
\]

\[-i/(k + 1) \int_{x_0}^{\infty} \log (z) \theta x_0^\theta z^{-\theta - 1} dz - \frac{1}{\theta} \int_{x_0}^{Q(F_\theta; i/(k + 1))} \theta x_0^\theta z^{-\theta - 1} dz
\]

\[-\log x_0 \int_{x_0}^{Q(F_\theta; i/(k + 1))} \theta x_0^\theta z^{-\theta - 1} dz + \int_{x_0}^{Q(F_\theta; i/(k + 1))} \log (z) \theta x_0^\theta z^{-\theta - 1} dz \]

but it can easily be shown that

\[\int_{x_0}^{Q(F_\theta; i/(k + 1))} \log (z) \theta x_0^\theta z^{-\theta - 1} dz = \frac{i}{k + 1} \log x_0 + \frac{1}{\theta} \left( 1 - \frac{i}{k + 1} \right) \log \left( 1 - \frac{i}{k + 1} \right) + \frac{1}{\theta} \frac{i}{k + 1} \]

and straightforward, although tedious, simplifications lead to

\[\text{cov} \left( Y_i, \hat{Y}_i \right) = \frac{1}{\theta^2} \log \left( \frac{k + 1 - i}{k + 1} \right)^2 , \]

which is equation (17) in the main text.
References


Figure 1: Histogram and Pareto regression line of a simulated sample from a uniform/Pareto mixture upper tail.

Figure 2: C-criterion for a simulated sample from a uniform/Pareto mixture upper tail.
Figure 3: Parameter estimates for 100 replications of a simulated sample of $n = 1000$ from a uniform/Pareto mixture upper tail.

Figure 4: (a) Histogram and (b) Pareto-regression plot of the Bougban data
Figure 5: (a) $C$-criterion as a function of $x_0$ and (b) $\hat{\theta}$ as a function of $x_0$ for the Bougban data

Figure 6: Pareto-regression plot for the Bougban data with the fitted regression lines based on two methods for choosing the optimal value for $k$
Figure 7: (a) Histogram and (b) Pareto-regression plot of the finance data

Figure 8: Pareto-regression plot for the finance data with the fitted (classical and robust) regression lines
Figure 9: Pareto-regression plot for the fiance data (without three extremes) with the fitted regression line

Figure 10: (a) Histogram and (b) Pareto-regression plot of the income data
Figure 11: Pareto-regression plot for the income data with the fitted regression line

Figure 12: Pareto-regression plot for Danish data with fitted regression line
Table 1: RMSE for $\hat{\theta}^{-1}$ with the Burr distribution ($\theta = 1$).

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$n$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500</td>
<td>1000</td>
<td>1500</td>
<td></td>
</tr>
<tr>
<td>$-0.5$</td>
<td>$C$ criterion</td>
<td>0.299</td>
<td>0.296</td>
<td>0.295</td>
</tr>
<tr>
<td></td>
<td>Method 1</td>
<td>0.334</td>
<td>0.238</td>
<td>0.220</td>
</tr>
<tr>
<td></td>
<td>Method 2</td>
<td>0.305</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Method 3</td>
<td>0.382</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Method 4</td>
<td>0.352</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Method 5</td>
<td>0.381</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>$-1.0$</td>
<td>$C$ criterion</td>
<td>0.235</td>
<td>0.143</td>
<td>0.128</td>
</tr>
<tr>
<td></td>
<td>Method 1</td>
<td>0.264</td>
<td>0.148</td>
<td>0.132</td>
</tr>
<tr>
<td>$-1.5$</td>
<td>$C$ criterion</td>
<td>0.116</td>
<td>0.094</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>Method 1</td>
<td>0.150</td>
<td>0.100</td>
<td>0.085</td>
</tr>
</tbody>
</table>

Table 2: RMSE for $\hat{\theta}^{-1}$ over several distributions, $n = 500$.

<table>
<thead>
<tr>
<th></th>
<th>Fréchet ($\theta = 2$)</th>
<th>Student ($\theta = 6$)</th>
<th>Loggamma ($\theta = 1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$ criterion</td>
<td>0.080</td>
<td>0.196</td>
<td>0.789</td>
</tr>
<tr>
<td>Method 2</td>
<td>0.072</td>
<td>0.105</td>
<td>0.245</td>
</tr>
<tr>
<td>Method 3</td>
<td>0.104</td>
<td>0.133</td>
<td>0.301</td>
</tr>
<tr>
<td>Method 4</td>
<td>0.064</td>
<td>0.114</td>
<td>0.280</td>
</tr>
<tr>
<td>Method 5</td>
<td>0.070</td>
<td>0.138</td>
<td>0.262</td>
</tr>
</tbody>
</table>
Table 3: RMSE for $\hat{\theta}^{-1}$ over several distributions, $n = 500$.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$\theta$</th>
<th>Method 6</th>
<th>$C$ criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burr, $\rho = -0.25$</td>
<td>1</td>
<td>0.780</td>
<td>0.660</td>
</tr>
<tr>
<td>Burr, $\rho = -0.5$</td>
<td>1</td>
<td>0.340</td>
<td>0.307</td>
</tr>
<tr>
<td>Burr, $\rho = -1$</td>
<td>1</td>
<td>0.110</td>
<td>0.156</td>
</tr>
<tr>
<td>Fréchet</td>
<td>1</td>
<td>0.130</td>
<td>0.145</td>
</tr>
<tr>
<td>Loggamma</td>
<td>1</td>
<td>0.200</td>
<td>0.789</td>
</tr>
<tr>
<td>$</td>
<td>t_2</td>
<td>$</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4: Distributions used in the simulation study and their corresponding tail index.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$\tau_p(G)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burr, $\rho = -1$</td>
<td>1</td>
</tr>
<tr>
<td>Burr, $\rho = -1.5$</td>
<td>0.999</td>
</tr>
<tr>
<td>Burr, $\rho = -0.5$</td>
<td>1.096</td>
</tr>
<tr>
<td>Burr, $\rho = -0.25$</td>
<td>1.829</td>
</tr>
<tr>
<td>Fréchet</td>
<td>0.996</td>
</tr>
<tr>
<td>Loggamma</td>
<td>1.272</td>
</tr>
<tr>
<td>$t_2$</td>
<td>0.976</td>
</tr>
<tr>
<td>$t_6$</td>
<td>0.910</td>
</tr>
</tbody>
</table>
Table 5: Pareto tail estimates for the Bougban data.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\theta$</th>
<th>$k$</th>
<th>$x_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$-criterion</td>
<td>1.598</td>
<td>241</td>
<td>0.348</td>
</tr>
<tr>
<td>Caers et al. (1996)</td>
<td>1.587</td>
<td>285</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Table 6: Estimated VaR and ES for the Finance data

VAR probabilities

<table>
<thead>
<tr>
<th></th>
<th>0.05</th>
<th>0.01</th>
<th>0.005</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>all data, Hill</td>
<td>103.04</td>
<td>103.70</td>
<td>106.87</td>
<td>109.63</td>
</tr>
<tr>
<td>3 largest removed, Hill</td>
<td>102.08</td>
<td>104.11</td>
<td>105.00</td>
<td>107.09</td>
</tr>
<tr>
<td>all data, OBRE</td>
<td>102.89</td>
<td>105.45</td>
<td>106.58</td>
<td>109.23</td>
</tr>
</tbody>
</table>

ES probabilities

<table>
<thead>
<tr>
<th></th>
<th>0.05</th>
<th>0.01</th>
<th>0.005</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>all data, Hill</td>
<td>1300</td>
<td>6769</td>
<td>13607</td>
<td>68304</td>
</tr>
<tr>
<td>3 largest removed, Hill</td>
<td>1269</td>
<td>6609</td>
<td>13284</td>
<td>66683</td>
</tr>
<tr>
<td>all data, OBRE</td>
<td>1299</td>
<td>6766</td>
<td>13599</td>
<td>68264</td>
</tr>
</tbody>
</table>

32
2002.01 Jean-Emile DENIS
Taking International Marketing Decisions under WTO Rules

2002.02 Jean-Luc PRIGENT, Olivier SCAILLET
Weak Convergence of Hedging Strategies of Contingent Claims

2002.03 Séverine CAUCHIE, Martin HOESLI, Dušan ISAKOV
The Determinants of Stock Returns: an Analysis of Industrial Sector Indices

2002.04 Ana C. CEBRIAN, Michel DENUIT, Olivier SCAILLET
Testing for Concordance Ordering

2002.05 M. DENUIT, Olivier SCAILLET
Nonparametric Tests for Positive Quadrant Dependence

2002.06 Bernard DUBOIS, Sandor CZELLAR
Prestige Brands or Luxury Brands? An Exploratory Inquiry on Consumer Perceptions

2002.07 John GUNSON, Jean-Paul DE BLASIS
Implementing ERP in Multinational Companies: their Effects on the Organization and Individuals at Work

2002.08 Michael GIBBERT, Gilbert J.B. PROBST
Anticipating and Managing Three Dilemmas in Knowledge Management: Insights from an In-Depth Case Study of a Major Diversified Firm

2002.09 Michael GIBBERT, Marius LEIBOLD, Gilbert J.B. PROBST
Five Styles of Customer Knowledge Management and how Smart Companies put them into Action

2002.10 Gilbert J.B. PROBST, Michael GIBBERT, Steffen RAUB
Wissensmanagement

2002.11 Claude TADONKI, Jean-Philippe VIAL
The Linear Separation Problem Revisited with ACCPM

2002.12 Philippe GAUD, Elion JANI
Détectants et dynamique de la structure du capital des entreprises suisses: une étude empirique

2002.13 Foort HAMELINK, Martin HOESLI
What Factors Determine International Real Estate Security Returns?

2002.14 Sandor CZELLAR
Consumer Attitude Towards Brand Extensions: an Integrative Model and Research Propositions
2002.15  Michelle BERGADAA
Achats en ligne : un nouveau cadre temporel d'analyse de l'éthique dans la relation achat-vente

2002.16  Michelle BERGADAA
Comment les mutations mondiales sont-elles vécues par les grandes entreprises en Europe ?

2002.17  Michelle BERGADAA, Leila AMRAOUI
Les mutations du Marketing : recherche naturelle conduite auprès de sept entreprises suisses en 2001

2002.18  Michelle BERGADAA, Samad LAAROUSSI
Proposition d'une typologie d'interface entreprise-client induite d'une recherche qualitative par groupes de focus

2002.19  Michelle BERGADAA, Mohamed Jamil HEBALI
Les seniors utilisateurs d'Internet : typologie induite d'une recherche qualitative en-ligne

2002.20  Patricia GARCIA-PRIETO, Erwan BELLARD, Susan C. SCHNEIDER
Experiencing Diversity, Conflict and Emotions in Teams

2002.21  Natacha AYMARD-MARTINOT, Zofia HUBER SWINARSKI, Bernard MORARD
Balanced Scorecard : Evaluation de la performance dans un établissement médico-social

2002.22  Sandor CZELLAR
A Link between Self-monitoring and Prestige-related Consumer Knowledge? Preliminary Evidence from North America and Europe

2002.23  Olivier DU MERLE, Jean-Philippe VIAL
Proximal ACCPM, a Cutting Plane Method for Column Generation and Lagrangian Relaxation : Application to the $P$-Median Problem

2002.24  Michelle BERGADAA
L’utilisation de multi-cas d’entreprises dans un contexte de changement et de crise : compte-rendu d’expérience

2002.25  Stéphane GRABER, Sandor CZELLAR, Jean-Emile DENIS
Using Partial Least Squares Regression in Marketing Research
2003.01 Stéphane GRABER, Sandor CZELLAR, Jean-Emile DENIS
Premières évidences empiriques sur les antécédents de l’orientation vers le marché
dans la gestion bancaire privée

2003.02 John GUNSON, Jean-Paul DE BLASIS
ERP implementation project management : an art as well as a science

2003.03 Paolo BATTOCCHIO, Francesco MENONCIN, Olivier SCAILLET
Optimal asset allocation for pension funds under mortality risk during the
accumulation and decumulation phases

2003.04 Dušan ISAKOV, Frédéric SONNEY
Are practitioners right ? On the relative importance of industrial factors in international
stock returns

2003.05 Martin HOESLI
Pourquoi les institutionnels investissent-ils si peu en immobilier ?

2003.06 Aharon BEN TAL, Boaz GOLANY, Arcadi NEMIROVSKI, Jean-Philippe VIAL
Supplier-Retailer Flexible Commitments Contracts : A Robust Optimization Approach

2003.07 Gaël VETTORI
Economies d’échelle – Les résultats de la recherche sur le secteur bancaire suisse

2003.08 César BELTRAN
The face simplex method in the cutting plane framework

2003.09 Philippe GAUD
Choix de financement des Firmes Européennes

2003.10 Olivier RENAULT, Olivier SCAILLET
On the Way to Recovery : A Nonparametric Bias Free Estimation of Recovery Rate
Densities

2003.11 Jean-Emile DENIS, Stéphane GRABER, Sandor CZELLAR
Orientation vers le marché (OVM), qualité des services et satisfaction des clients dans le
secteur du private banking (PB)

2003.12 Steven C. BOURASSA, Martin HOESLI, Vincent S. PENG
Do Housing Submarkets Really Matter ?

2003.13 Steven C. BOURASSA, Martin HOESLI, Jian SUN
What’s in a view ?

2003.14 Peter ENGLUND, Ake GUNNELIN, Martin HOESLI, Bo SÖDERBERG
Implicit Forward Rents as Predictors of Future Rents
2003.15  John GUNSON, Jean-Paul DE BLASIS, Mary NEARY
Leadership in Real Time: A Model of Five Levels of Attributes Needed by a Project Manager in ERP Implementations

2003.16  Kpate ADJAOUTE, Jean-Pierre DANTHINE, Dusan ISAKOV
Portfolio Diversification in Europe

2003.17  Elion JANI
Les déterminants des liquidités des entreprises suisses

2003.18  Jean-David FERMANIAN, Olivier SCAILLET
Sensitivity Analysis of Var and Expected Shortfall for Portfolios under Netting Agreements

2003.19  Debbie J. DUPUIS, Maria-Pia VICTORIA-FESER
A Prediction Error Criterion for Choosing the Lower Quantile in Pareto Index Estimation

La liste complète des cahiers de recherche se trouve sur notre site web
http://www.hec.unige.ch/recherches_publications/recherche_publications.htm

Pour commander des documents, veuillez vous adresser à :
Caroline Sahlé – HEC Uni-Mail – 40 bd du Pont-d’Arve – 1211 Genève 4
par tél. +41 22 705 81 09 ou par e-mail : caroline.sahle@hec.unige.ch