CentER Discussion Paper

No. 2005–112

UNBIASED TAIL ESTIMATION BY AN EXTENSION OF THE GENERALIZED PARETO DISTRIBUTION

By Jan Beirlant, Elisabeth Joossens, Johan Segers

October 2005

ISSN 0924-7815

TILBURG UNIVERSITY

Unbiased Tail Estimation by an Extension of the Generalized Pareto Distribution

Jan BEIRLANT, Elisabeth JOOSSENS and Johan SEGERS

Abstract

The generalized Pareto distribution (GPD) is probably the most popular model for inference on the tail of a distribution. The peaks-over-threshold methodology postulates the GPD as the natural model for excesses over a high threshold. However, for the GPD to fit such excesses well, the threshold should often be rather large, thereby restricting the model to only a small upper fraction of the data. In case of heavy-tailed distributions, we propose an extension of the GPD with a single parameter, motivated by a second-order refinement of the underlying Pareto-type model. Not only can the extended model be fitted to a larger fraction of the data, but in addition is the resulting maximum likelihood for the tail index asymptotically unbiased. In practice, sample paths of the new tail index estimator as a function of the chosen threshold exhibit much larger regions of stability around the true value. We apply the method to daily log-returns of the euro-UK pound exchange rate. Some simulation results are presented as well.

1 INTRODUCTION

Analyzing financial and actuarial risks requires statistical models for profit-loss and claim severity distributions over their full range, or at least a large part of that range. Recently, several authors emphasized the need for parametric models that are able to describe well both the central part of the distribution of a risk as well as its tail, so as to be able to calculate accurately, on the one hand, classical risk measures like the variance and, on the other hand, tail-related risk measures like value-at-risk and expected shortfall. To this end, some authors have presented mixture models or models with change points in the densities. See for instance Frigessi *et al.* (2002) and Cooray and Ananda (2005).

This version: October 21, 2005.

Jan Beirlant is full professor in statistics at the University Center of Statistics of the Catholic University Leuven, W. de Croylaan 54b, B-3001 Heverlee, Belgium (email: jan.beirlant@wis.kuleuven.be). Elisabeth Joossens is statistician at the Joint Research Center of the European Commission, TP 361, Via Fermi 1, 21020 Ispra (VA), Italy (email: elisabeth.joossens@jrc.it) and is PhD student at the Department of Mathematics of the Catholic University Leuven. Johan Segers is assistent professor in statistics at the Department of Econometrics and Operations Research of Tilburg University, PO Box 90153, NL-5000 LE Tilburg, the Netherlands (email: jsegers@uvt.nl) and is supported by a VENI grant of the Netherlands Organization for Scientific Research (NWO).

AMS 2000 subject classifications: 62 G 20; 62 G 32. *JEL:* C 13; C 14

Key Words: bias reduction; exchange rate; heavy tails; peaks-over-threshold; regular variation; tail index.

Adequate modelling of the extreme tail is needed in order to be able to estimate frequencies of rare events. The peaks-over-threshold methodology (Davison and Smith, 1990) puts forward the generalized Pareto distribution (GPD) as the appropriate model to be fitted to the excess of a variable X over a high threshold u: for $y \geq 0$,

$$\begin{aligned} \Pr[X - u > y \mid X > u] &\approx (1 + \gamma y/\sigma)_+^{-1/\gamma} \\ &= 1 - G_{\gamma,\sigma}(y). \end{aligned} \tag{1.1}$$

Here, γ is a real shape parameter called the tail index of the distribution of X, while $\sigma > 0$ is scale parameter which can change with u. The motivation for the GPD is a result due to Pickands (1975) claiming the approximation (1.1) to be asymptotically correct for increasingly high thresholds for a large class of distributions, including virtually all continuous text-book distributions.

The requirement that the threshold should be high enough shows up in practical data analysis too: typically, for the GPD to fit threshold excesses well, the threshold needs to be so high that at most 10% of the data can be modelled. It would then be desirable to devise more flexible distribution families capable of modelling excesses over lower thresholds.

To illustrate the abovementioned restriction of the GPD, and in fact every available extreme value method, we make use of the 1691 daily log-returns of the euro-UK pound daily exchange rate from January 4, 1999, till August 8, 2005. The data were collected by the European System of Central Banks and are available at `http://www.bportugal.pt/rates/cambtx`.

Figure 1 shows on the left a histogram of the logarithms of the 805 positive log-returns and on the right a Pareto QQ plot. The latter plot clearly indicates that the distribution of the log-returns has a heavy tail. Further, in Figure 2 the maximum likelihood estimates of γ arising when the GPD is fitted to the threshold excesses are given as a function of the number k of exceedances when the threshold u is chosen as the $(k+1)$-th largest observation. The estimates change considerably with k, raising the question of an appropriate choice of k or u. In Figure 3, we use W-plots (Smith and Shively, 1995) to evaluate the goodness-of-fit of the GPD for excesses over, on the left, a high threshold u equal to 0.0061, corresponding to $k = 150$, and, on the right, a lower threshold of $u = 0.0038$ or $k = 300$. These plots consist of the pairs

$$\left(-\log\left(1 - G_{\hat{\gamma},\hat{\sigma}}(X_{n-j+1,n} - u)\right), -\log(1 - j/(k+1))\right)$$

for $j = 1, \ldots, k$, which should lie along the first diagonal if the GPD fits well. Here, $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ denote the ordered data. The fit at the higher threshold of $k = 150$ is indeed better than at the lower one of $k = 300$.

So, the goal we pursue here is to extend the by now classical GPD in such a way that, at the same time, the tail fits from the GPD are safeguarded and, moreover, the fit is still satisfactory for much lower thresholds and thus for a much larger sample fraction. In the process of fitting this extended GPD (EGPD), we obtain new estimators for the tail index γ that are asymptotically unbiased. In practice this means that the sample paths of the new tail index estimator as a function of the threshold exhibit much larger regions of stability around the target value. In this way, the choice of the threshold becomes less of an issue. Recently, several other bias-reduction methods have been proposed, such as in Feuerverger and Hall (1999), Beirlant *et al.* (1999, 2002), Gomes *et al.* (2000) and Gomes and Martins (2002). These methods are based on refined models for spacings of subsequent order statistics and do not provide an explicit parametric model to be fitted to the data directly.

In section 2, the EGPD is presented and motivated. Maximum likelihood estimators for the EGPD parameters are derived in section 3. Their practical use is illustrated for the exchange rate data and the asymptotic properties of the corresponding tail index estimator are derived. We conclude the paper with some simulations in section 4.

2 THE NEW MODEL

To motivate the EGPD we will use the relative excesses instead of the absolute excesses as used in (1.1). This is rather natural as we confine ourselves to the case of heavy-tailed or Pareto-type distributions, that is, $\gamma > 0$. Replacing y in (1.1) by $ux - u$ with $x \geq 1$ leads to

$$\Pr[X/u > x \mid X > u] \approx [x\{(1+\delta) - \delta x^{-1}\}]^{-1/\gamma}, \tag{2.1}$$

where $\delta = u\gamma/\sigma - 1$. In this sense we propose to extend the GPD model by

$$\begin{aligned} \Pr[X/u > x \mid X > u] &\approx [x\{(1+\delta) - \delta x^{\rho}\}]^{-1/\gamma} \\ &= 1 - G_{\gamma,\delta,\rho}(x), \end{aligned} \tag{2.2}$$

the special case $\rho = -1$ yielding the GPD. The ranges for the parameters are $\gamma > 0$, $\rho < 0$, and $\delta \geq \max(-1, 1/\rho)$. Also, the case $\delta = 0$ in (2.2) leads to the simple strict Pareto model for the relative excesses, as used in Hill (1975).

The motivation for the proposed form of the EGPD is that under general assumptions, the approximation error in (2.2) is an order of magnitude smaller than the one in (2.1), as we will show below. First of all, recall that the assumption that the distribution of a risk X is of Pareto-type with Pareto index $1/\gamma$ for some $\gamma > 0$ is equivalent to the assumption that for high positive thresholds u, the relative excess X/u given $X > u$ is approximately

Pareto distributed. Formally, denoting the distribution and tail functions of X by F and $\bar{F} = 1 - F$, respectively, the assumption is that for $x \geq 1$,

$$\text{(2.3)} \qquad \begin{aligned} \Pr[X/u > x \mid X > u] &= \frac{\bar{F}(ux)}{\bar{F}(u)} \\ &\to x^{-1/\gamma} \qquad \text{as } u \to \infty. \end{aligned}$$

This assumption is equivalent to the case $\gamma > 0$ in Pickands (1975) leading to the GPD approximation in (1.1). Failure of the (generalized) Pareto distribution to fit the excesses over a threshold u is due to a bad approximation in (1.1) and (2.3), arising when u is not yet high enough for the asymptotics to kick in. As a result, estimates of the tail index γ and other tail quantities will be biased.

A second-order refinement of (2.3), capturing the deviation between the left-hand and right-hand sides, is

$$\text{(2.4)} \qquad \frac{\Pr[X/u > x \mid X > u] - x^{-1/\gamma}}{A(u)} \to c x^{-1/\gamma} \frac{x^{\rho} - 1}{\rho}$$

as $u \to \infty$. Here, c and $\rho < 0$ are real numbers and A is a positive function vanishing at infinity; see for example Bingham *et al.* (1987), chapter 3, or Geluk and de Haan (1987). The tail expansion (2.4) is verified by most heavy-tailed distributions, like for instance the ones listed in Table 1. The constant ρ is called the second-order parameter. The closer it is to zero, the slower is the rate of convergence in (2.3) and thus the higher the threshold that is needed in order to make the (generalized) Pareto approximation work.

Rewriting (2.4) gives, as $u \to \infty$,

$$\text{(2.5)} \qquad \begin{aligned} &\Pr[X/u > x \mid X > u] \\ &\quad = \{1 - \rho^{-1} c A(u)\} x^{-1/\gamma} + \rho^{-1} c A(u) x^{-1/\gamma + \rho} + o(A(u)) \\ &\quad = [x\{(1+\delta) - \delta x^{\rho}\}]^{-1/\gamma} + o(A(u)), \end{aligned}$$

with $\delta = \delta(u) = -cA(u)/\rho$. Dropping the remainder term $o(A(u))$ now gives the new model (2.2). The Pareto approximation, which is of order $O(A(u))$, is thus refined to an approximation which is of order $o(A(u))$.

Note that the expansion in the above display can also be used to motivate other extensions of the (generalized) Pareto distribution. For instance, the second line of the previous display provides motivation to model the distribution of relative excess by a mixture of two Pareto distributions.

3 ESTIMATION METHOD

Suppose we have n risks, $X_1, \ldots, X_n$, independent and identically distributed with unknown distribution function F. Maximum likelihood estimation in the Pareto model based on the k relative excesses over the $(k+1)$th

Table 1: *Tail index γ and second-order constant ρ for selected heavy-tailed distributions. Note: Burr($\gamma, -\gamma, \sigma/\gamma$) is the same as GPD($\gamma, \sigma$).*

distribution [parameters]	distribution function	γ	ρ
Burr($\gamma, \tilde{\rho}, \beta$) [$\gamma > 0$, $\tilde{\rho} < 0$, $\beta > 0$]	$1 - \left(1 + x^{-\tilde{\rho}/\gamma}/\beta\right)^{1/\tilde{\rho}}$	γ	$\tilde{\rho}/\gamma$
Student-t_ν [$\nu > 0$]	$C(\nu) \int_{-\infty}^{x} \left(1 + y^2/\nu\right)^{-(\nu+1)/2} \, \mathrm{d}y$	$1/\nu$	-2
Fréchet(α) [$\alpha > 0$]	$\exp(-x^{-\alpha})$	$1/\alpha$	$-\alpha$
Inverse Burr(λ, τ, σ) [$\lambda > 0$, $\tau > 0$, $\sigma > 0$]	$\left(1 + (x/\sigma)^{-\lambda}\right)^{-\tau}$	$1/\lambda$	$-\lambda$

largest observation $X_{n-k,n}$ leads to the Hill (1975) estimator

$$\hat{\gamma}_{k,n}^{H} = \frac{1}{k} \sum_{j=1}^{k} \log \frac{X_{n-j+1,n}}{X_{n-k,n}} \tag{3.1}$$

for the tail index γ. Discrepancies between the true excess distribution and the asymptotic Pareto model may lead to severely biased tail estimates. In such circumstances, the EGPD (2.2) can provide a much more satisfactory fit to the relative excesses.

We propose to fit the EGPD to the relative excesses over a high enough threshold u by first computing an initial estimate of the second-order shape parameter ρ and subsequently estimating γ and δ by maximum likelihood. The estimate for ρ is based on a larger fraction of the sample than just the relative excesses over u, a procedure sometimes referred to as external estimation of ρ. We will follow here the approach in Fraga Alves *et al.* (2003), which is one of the more promising ways to estimate the second-order shape parameter for heavy-tailed distributions, see remark 3.2 below.

The reason we do not estimate the three EGPD parameters (γ, δ, ρ) jointly by maximum likelihood is that for reasonable sample sizes, the likelihood surface is almost flat in ρ, rendering such a procedure rather cumbersome in practice. An additional problem is that for ρ close to zero, the model is hardly identifiable. In a similar situation, the full maximum likelihood method is discussed in Feuerverger and Hall (1999) and Beirlant *et* *al.* (1999, 2002).

We apply the method to the 805 daily log-returns of the euro-UK pound exchange rate introduced in section 1. In Figure 4, we present estimates for the tail index γ for a range of thresholds $u = X_{n-k,n}$ as a function of k. We compare three estimators: the Hill estimator (3.1), the maximum likelihood estimator when fitting the GPD (1.1) to the absolute excesses, and the above described estimator when fitting the EGPD to the relative excesses with initial second-order parameter estimate $\hat{\rho} = -0.84$. Observe the stability of the EGPD-based estimates up to $k = 400$, leading to an estimate for γ close to 0.25. Also for the new model the goodness-of-fit can be assessed by W-plots, this time consisting of the pairs

$$\left(-\log\left(1 - G_{\hat{\gamma},\hat{\delta},\hat{\rho}}(X_{n-j+1,n}/u)\right), -\log(1 - j/(k+1))\right)$$

for $j = 1, \ldots, k$. In Figure 5, W-plots are constructed for the same thresholds as in Figure 3, corresponding to $k = 150$ and 300 exceedances, respectively. For both thresholds, the EGPD provides a clearly better fit, especially in the far tail.

Given an initial estimate $\hat{\rho}$ of the second-order parameter, the maximum likelihood estimates for the EGPD parameters γ and ρ are defined as the solutions to the score equations that arise when the relative excesses

$$E_j = X_{n-j+1,n}/X_{n-k,n}, \quad \text{for } j = 1, \ldots, k.$$

are modelled by the EGPD (2.2). Since $\delta = \delta(u) = -cA(u)/\rho$ is typically close to zero, see (2.5), we can approximately solve the score equations by linearizing them in δ. Writing $\overline{E}_{k,n}(a) = k^{-1}\sum_{j=1}^{k} E_j^a$, we obtain

$$\hat{\gamma}_{k,n}^{EP}(\hat{\rho}) = \hat{\gamma}_{k,n}^{H} + \hat{\delta}_{k,n}^{EP}(\hat{\rho})\left(1 - \overline{E}_{k,n}(\hat{\rho})\right), \tag{3.2}$$

$$\begin{aligned} \hat{\delta}_{k,n}^{EP}(\hat{\rho}) &= \left\{\left(1 - \hat{\rho}\hat{\gamma}_{k,n}^{H}\right)\overline{E}_{k,n}(\hat{\rho}) - 1\right\} \\ &\quad \div \left\{\left(2 + \hat{\rho} - \hat{\rho}\overline{E}_{k,n}(\hat{\rho}) - 2\hat{\rho}\hat{\gamma}_{k,n}^{H}\right)\overline{E}_{k,n}(\hat{\rho}) \right. \\ &\qquad \left. - \left(1 - 2\hat{\gamma}_{k,n}^{H}\hat{\rho}(2+\hat{\rho})\right)\overline{E}_{k,n}(2\hat{\rho}) - 1\right\}. \end{aligned} \tag{3.3}$$

The details of this derivation can be found in Appendix A.1.

The EGPD-based tail index estimator as obtained from (3.2) and (3.3) is asymptotically normal. Theorem 3.1 below shows that for intermediate sequences $k = k_n$ for which $\sqrt{k}A(U(n/k)) \to \lambda \neq 0$, the main bias component of the new estimator is of smaller order than $A(U(n/k))$, the latter being the order of the bias of for instance the Hill or GPD-based estimators. In practice, this leads to more stable sample paths of the estimates as a function of k. The asymptotic variance is the same as the one obtained in Gomes and Martins (2002) for the estimator in (3.4); see also Feuerverger and Hall (1999) and Beirlant *et al.* (2002).

Theorem 3.1. *Let $X_1, \ldots, X_n$ be a random sample from a distribution function F satisfying (2.4). If $k = k_n$ is a sequence of positive integers such that $k \to \infty$, $k/n \to 0$ and $\sqrt{k}A(U(n/k)) \to \lambda \in [0, \infty)$ as $n \to \infty$, then*

$$\sqrt{k}\left(\hat{\gamma}^{EP}_{k,n}(\rho) - \gamma\right) \to_d N\left(0, \frac{(1-\rho\gamma)^2}{\rho^2}\right).$$

This statement holds also true if we replace ρ by a weakly consistent estimator sequence $\hat{\rho}_n$.

The proof of Theorem 3.1 can be found in Appendix A.2.

Remark 3.1. The approach described here should be compared with another bias-reduction technique described in Feuerverger and Hall (1999), Beirlant *et al.* (1999, 2002) and more specifically in Gomes and Martins (2002). These references rely on the fact that the scaled log-spacings

$$Z_j = j(\log X_{n-j+1,n} - \log X_{n-j,n}),$$

for $j = 1, \ldots, k$, are approximately exponentially distributed with mean $\gamma e^{D(j/n)^{-\tilde{\rho}}}$, where $\tilde{\rho} < 0$ and $D \in \mathbb{R}$. A similar linearization of the likelihood then leads to the estimator

$$\hat{\gamma}^{U}_{k,n}(\tilde{\rho}) = \hat{\gamma}^{H}_{k,n} + \bar{Z}_k(\tilde{\rho})\frac{\left(k^{-1}\sum_{j=1}^{k} j^{-\tilde{\rho}}\right)\bar{Z}_k(0) - \bar{Z}_k(\tilde{\rho})}{\left(k^{-1}\sum_{j=1}^{k} j^{-\tilde{\rho}}\right)\bar{Z}_k(\tilde{\rho}) - \bar{Z}_k(2\tilde{\rho})}, \tag{3.4}$$

where $\bar{Z}_k(a) = k^{-1}\sum_{j=1}^{k} j^{-a}Z_j$. Gomes and Martins (2002) studied this estimator in detail under a second-order regular variation assumption formulated for the tail quantile function $U(x) = F^{\leftarrow}(1 - x^{-1})$, with $F^{\leftarrow}$ the quantile or generalized inverse function of F, rather than for the tail function $\bar{F}$ as in our equation (2.4): the assumption is that there exist $\gamma > 0$, $\tilde{\rho} < 0$, $C \in \mathbb{R}$ and a positive function B vanishing at infinity such that

$$\frac{U(tx)/U(t) - x^{\gamma}}{B(t)} \to Cx^{\gamma}\frac{x^{\tilde{\rho}} - 1}{\tilde{\rho}} \tag{3.5}$$

as $t \to \infty$ and for all $x > 0$. It can be shown that (2.4) and (3.5) are equivalent with $B(t) = A(U(t))$, $\tilde{\rho} = \rho\gamma$, and $C = \gamma^2 c$.

Remark 3.2. Computation of the EGPD-parameter estimates in (3.2) and (3.3) requires an initial estimator of the second-order parameter ρ. In the case study and in the simulations in section 4, we have exploited the relation $\rho = \tilde{\rho}/\gamma$ with $\tilde{\rho}$ as in (3.5). The numerator, $\tilde{\rho}$, is estimated by the estimator proposed in Fraga Alves *et al.* (2003), given by

$$\hat{\tilde{\rho}}^{(\tau)}(k_1) = -\left|\frac{3\left(T_n^{(\tau)}(k_1) - 1\right)}{T_n^{(\tau)}(k_1) - 3}\right|, \tag{3.6}$$

where $\tau \geq 0$ is a tuning parameter and

$$\begin{aligned} T_n^{(\tau)}(k_1) &= \frac{\left(M_n^{(1)}(k_1)\right)^{\tau} - \left(M_n^{(2)}(k_1)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k_1)/2\right)^{\tau/2} - \left(M_n^{(3)}(k_1)/6\right)^{\tau/3}}, \\ M_n^{(i)}(k_1) &= \frac{1}{k_1} \sum_{j=1}^{k_1} \left(\log X_{n-j+1,n} - \log X_{n-k_1,n}\right)^i, \end{aligned}$$

with $a^{b\tau}$ standing for $b \log a$ when $\tau = 0$. The value k_1 determines the threshold used to estimate $\tilde{\rho}$. In Fraga Alves *et al.* (2003), it is recommended to take $k_1 = \min(n-1, \lfloor 2n/\log\log n \rfloor)$ and to set the tuning parameter τ equal to 0 whenever one expects $\tilde{\rho}$ to be in the range $[-1, 0)$ and to set τ equal to 1 otherwise. The final estimator for ρ, then, is given by

$$\hat{\rho} = \tilde{\rho}^{(\tau)}(k_1)/\hat{\gamma}, \tag{3.7}$$

with the estimator (3.4) of γ used in the denominator.

Remark 3.3. The actual interest in risk analysis often lies in tail-related risk measures such as value-at-risk (VaR) and expected shortfall. The approximation $\Pr[X > x] \approx \Pr[X > u]\bar{G}_{\gamma,\delta,\rho}(x/u)$ for $x \geq u > 0$ yields estimators for various tail quantities. The tail function $\bar{F}$ can be estimated by

$$\hat{\bar{F}}(x) = \frac{N_u}{n} \bar{G}_{\hat{\gamma},\hat{\delta},\hat{\rho}}(x/u) \tag{3.8}$$

for $x \geq u > 0$. Invert (3.8) to obtain an estimator for $\mathrm{VaR}_p = F^{\leftarrow}(1-p)$:

$$\widehat{\mathrm{VaR}}_p = G^{\leftarrow}_{\hat{\gamma},\hat{\delta},\hat{\rho}}\left(\frac{np}{N_u}\right) \tag{3.9}$$

for $0 < p < N_u/n$, where $G^{\leftarrow}_{\hat{\gamma},\hat{\delta},\hat{\rho}}$ denotes the quantile function of the fitted EGPD. Finally, the expected shortfall at tail probability p,

$$\mathrm{ESF}_p = \mathrm{E}(X - \mathrm{VaR}_p)_+ = \int_{\mathrm{VaR}_p}^{\infty} \bar{F}(x)\,\mathrm{d}x,$$

can be estimated by substitution of (3.8) and (3.9) into the previous display.

4 SIMULATIONS

We assessed the performance of the maximum likelihood estimator for the tail index γ that arises when fitting the EGPD to the relative excesses over a large threshold. The second-order parameter ρ was estimated as in (3.7), with estimator (3.4) of γ used in the denominator. The EGPD-based estimate for the tail index was then compared to the Hill estimator (3.1), the

Table 2: *Distributions in simulation study.*

Burr$(\gamma, \tilde\rho, \beta)$	•	$(\gamma, \tilde\rho, \beta) = (0.5, -1, 1)$
	•	$(\gamma, \tilde\rho, \beta) = (0.5, -0.5, 1)$
	•	$(\gamma, \tilde\rho, \beta) = (0.5, -0.25, 1)$
Student-t_ν	•	$\nu = 4$, so $\gamma = 1/\nu = 0.25$ and $\rho = -2$
Fréchet(α)	•	$\alpha = 2$, so $\gamma = 0.5$

maximum likelihood estimator based on the GPD-approximation for absolute excesses, and the estimator in equation (3.4) with initial estimator for $\tilde\rho$ as in (3.6). The distributions considered in the simulation study are listed in Table 2.

For each distribution, 100 random samples of size 1000 were generated. The plots in Figures 6 and 7 contain, as a function of k, the medians (left) as well as the root mean squared errors (right) of the estimates.

The new method leads to competitive results in comparison with the existing methods. Note that in case of the Burr(0.5,-0.5,1) distribution, which coincides with the GPD(0.5,0.5) distribution, the root MSE performance of the EGPD-based estimator is comparable to the one of the GPD-based estimator.

When simpler estimators than (3.4) such as the Hill estimator are used for the denominator of (3.7), then the overall picture as given by the simulations still stands, except for t-distributions in which case the method proposed in Gomes and Martins (2002) works best.

A APPENDIX

Recall $\overline{E}_{k,n}(a) = k^{-1} \sum_{j=1}^k E_j^a$ with $E_j = X_{n-j+1,n}/X_{n-k,n}$.

A.1 Score Equations

The score equations that arise when modelling the relative excesses by the EGPD family are

$$-\frac{k}{\hat\gamma} + \frac{1}{\hat\gamma^2} \sum_{j=1}^{k} \log E_j + \frac{1}{\hat\gamma 2} \sum_{j=1}^{k} \log(1 + \hat\delta - \hat\delta E_j^{\hat\rho}) = 0, \tag{A.1}$$

and

$$\left(\frac{1}{\hat\gamma} + 1\right) \sum_{j=1}^{k} \frac{1 - E_j^{\hat\rho}}{1 + \hat\delta - \hat\delta E_j^{\hat\rho}} = \sum_{j=1}^{k} \frac{1 - (1+\hat\rho) E_j^{\hat\rho}}{1 + \hat\delta - \hat\delta(1+\hat\rho) E_j^{\hat\rho}} \tag{A.2}$$

respectively. By (A.1) we can express $\hat{\gamma}$ explicitly as a function of $\hat{\delta}$:

$$\hat{\gamma} = \frac{1}{k}\sum_{j=1}^{k} \log E_j + \frac{1}{k}\sum_{j=1}^{k} \log(1+\hat{\delta}-\hat{\delta}E_j^{\hat{\rho}}). \tag{A.3}$$

Linearize equation (A.2) in $\hat{\delta}$ to find

$$\hat{\delta} = \frac{\left(\frac{1}{\hat{\gamma}}+1\right)\sum_{j=1}^{k}\left(1-E_j^{\hat{\rho}}\right) - \sum_{j=1}^{k}\left(1-(\hat{\rho}+1)E_j^{\hat{\rho}}\right)}{\left(\frac{1}{\hat{\gamma}}+1\right)\sum_{j=1}^{k}\left(1-E_j^{\hat{\rho}}\right)^2 - \sum_{j=1}^{k}\left(1-(\hat{\rho}+1)E_j^{\hat{\rho}}\right)^2}. \tag{A.4}$$

The first term on the right-hand side in (A.3) is equal to the Hill estimator. Hence, after linearization in $\hat{\delta}$, the maximum likelihood estimator for γ can be written as in (3.2). Substitute (3.2) into (A.4) and linearize once more to get arrive at expression (3.3) for the estimator of δ.

A.2 Proof of Theorem 3.1

Using the methods of proof for Theorem 3.1 in Dekkers *et al.* (1989) and Theorem 1 in de Haan and Peng (1998), we find

$$\sqrt{k}\left\{\left(\hat{\gamma}_{k,n}^{H}, \overline{E}_{k,n}(\rho), \overline{E}_{k,n}(2\rho)\right) - \left(\gamma, \frac{1}{1-\rho\gamma}, \frac{1}{1-2\rho\gamma}\right)\right\}' \to_d N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where

$$\boldsymbol{\mu} = \left(\frac{c\gamma^2\lambda}{1-\rho\gamma}, \frac{c\gamma^2\lambda}{(1-\rho\gamma)(1-2\rho\gamma)}, \frac{2c\gamma^2\lambda}{(1-2\rho\gamma)(1-3\rho\gamma)}\right)'$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \gamma^2 & \frac{\rho\gamma^2}{(1-\rho\gamma)^2} & \frac{2\rho\gamma^2}{(1-2\rho\gamma)^2} \\ \frac{\rho\gamma^2}{(1-\rho\gamma)^2} & \frac{\rho^2\gamma^2}{(1-2\rho\gamma)(1-\rho\gamma)^2} & \frac{2\rho^2\gamma^2}{(1-\rho\gamma)(1-2\rho\gamma)(1-3\rho\gamma)} \\ \frac{2\rho\gamma^2}{(1-2\rho\gamma)^2} & \frac{2\rho^2\gamma^2}{(1-\rho\gamma)(1-2\rho\gamma)(1-3\rho\gamma)} & \frac{4\rho^2\gamma^2}{(1-4\rho\gamma)(1-2\rho\gamma)^2} \end{pmatrix}.$$

From equation (3.2) it follows that $\hat{\gamma}_{k,n}^{EP}$ can be written as

$$\begin{aligned} \hat{\gamma}_{k,n}^{EP}(\rho) &= g\left(\hat{\gamma}_{k,n}^{H}, \overline{E}_{k,n}(\rho), \overline{E}_{k,n}(2\rho)\right) \\ &= \hat{\gamma}_{k,n}^{H} + \left(1-\overline{E}_{k,n}(\rho)\right) f\left(\hat{\gamma}_{k,n}^{H}, \overline{E}_{k,n}(\rho), \overline{E}_{k,n}(2\rho)\right), \end{aligned}$$

where

$$f(x,y,z) = \frac{1+(\rho x-1)y}{1+(\rho y-\rho+2\rho x-2)y+(1-(1+\rho)\rho x)z}.$$

Since, moreover,

$$g\left(\gamma, \frac{1}{1-\rho\gamma}, \frac{1}{1-2\rho\gamma}\right) = \gamma,$$

we find by the delta-method,

$$\sqrt{k}\left(\hat{\gamma}_{k,n}^{EP}(\rho)-\gamma\right) \to_d N\left(\nabla g'\boldsymbol{\mu}, \nabla g'\boldsymbol{\Sigma}\nabla g\right),$$

where ∇g is the gradient of g evaluated in $\big(\gamma,(1-\rho\gamma)^{-1},(1-2\rho\gamma)^{-1}\big)'$. Some tedious algebra now leads from the previous display to the expressions for the limiting mean and variance as stated in the theorem.

The final statement in the theorem can be proved in the same way as Corollary 2.1 in Gomes and Martins (2002). This concludes the proof of Theorem 3.1.

References

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

Beirlant, J., Dierckx, G., Guillou, A., and Stărică, C. (2002), "On exponential representations of log-spacings of extreme order statistics," *Extremes*, 5, 157–180.

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

Cooray, K., and Ananda, M. A. (2005), "Modeling actuarial data with a composite lognormal-Pareto model, " *Scandinavian Actuarial Journal*, 5, 321–334.

Dekkers, A. L. M, Einmahl, J. H. J., and de Haan, L. (1989), "A moment estimator for the index of the extreme-value distribution", *The Annals of Statistics*, 17, 1833–1855.

Davison, A. C. and Smith, R. L. (1990), "Models for exceedances over high thresholds," *Journal of the Royal Statistical Society*, Ser. B, 52, 393–442.

Feuerverger, A. and Hall, P. (1999), "Estimating a tail exponent by modelling departure from a Pareto distribution," *The Annals of Statistics*, 27, 760–781.

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, 193–214.

Frigessi, A., Haug, O., and Rue, H. (2002), "A dynamic mixture model for unsupervised tail estimation without threshold selection," *Extremes*, 5, 219–235.

Geluk, J. and de Haan, L. (1987), *Regular Variation, Extensions and Tauberian Theorems*, Amsterdam: CWI tract 40.

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

Gomes, M. I. and Martins, M. (2002), "Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter," *Extremes*, 5, 5–31.

de Haan, L. and Peng, L. (1998), "Comparison of tail index estimators," *Statistica Neerlandica*, 52, 60–70.

Hill, B. M. (1975), "A simple approach to inference about the tail of a distribution," *The Annals of Statistics*, 3, 1163–1174.

Pickands, J. (1975), "Statistical inference using extreme order statistic," *The Annals of Statistics*, 3, 119–131.

Smith, R. L. and Shively, T. S. (1995), "A point process approach to modeling trends in tropospheric ozone," *Atmospheric Environment*, 29, 3489–3499.

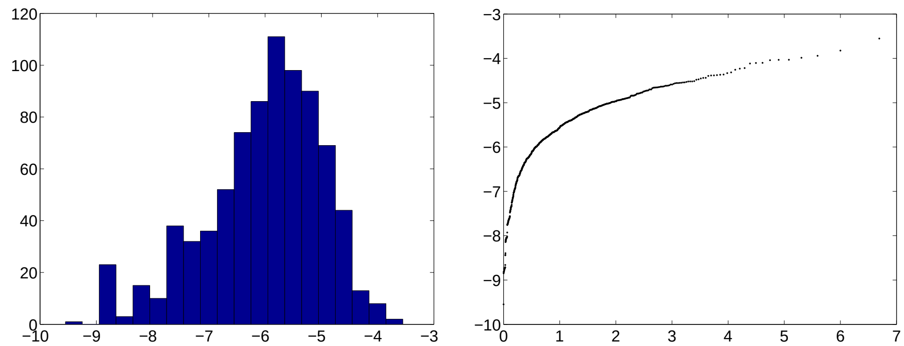

Figure 1: Left: *Histogram of the logarithms of the positive daily log-returns of the euro-UK pound exchange rate.* Right: *Pareto QQ plot of the daily log-returns*

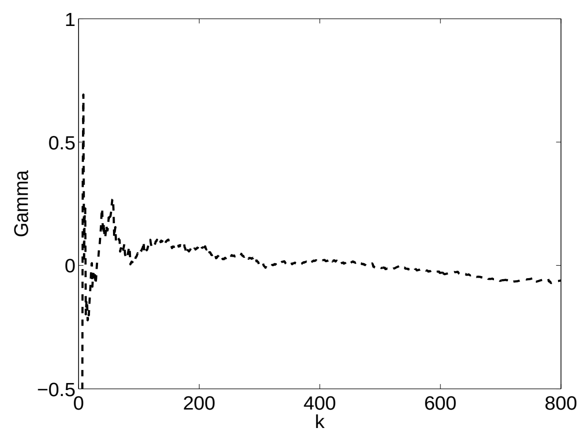

Figure 2: *Maximum likelihood estimates of γ for the daily log-returns of the euro-UK pound exchange rate that result from fitting the GPD to excesses over thresholds with k exceedances*

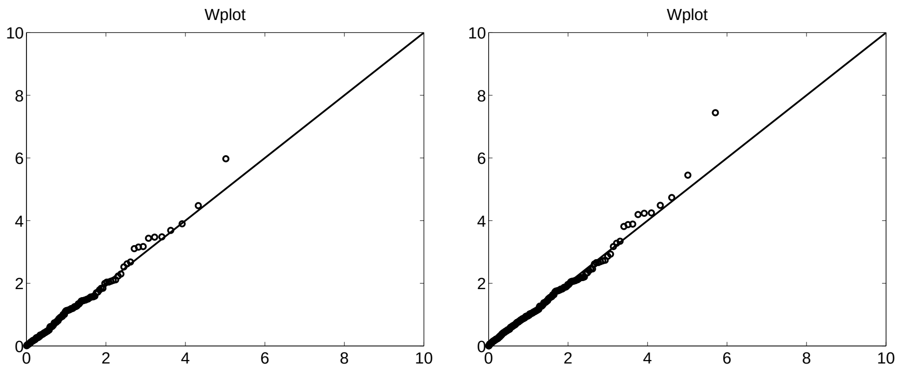

Figure 3: *W-plots assessing the goodness-of-fit of the GPD to excesses over thresholds with* $k = 150$ *(left) and* $k = 300$ *(right) exceedances for the daily log-returns of the euro-UK pound exchange rate*

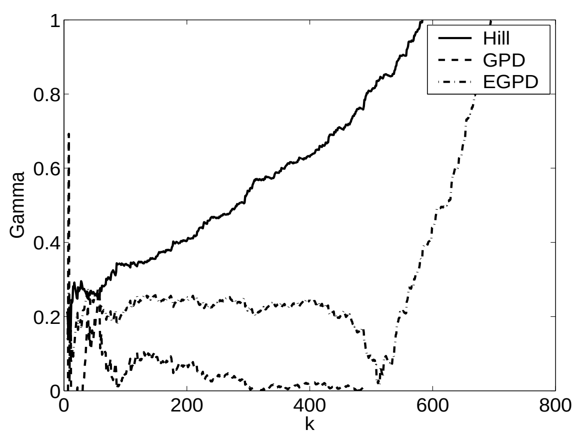

Figure 4: *Maximum likelihood estimates of* γ *for the daily log-returns of the euro-UK pound exchange rate that result from fitting the Pareto, GPD, and EGPD to excesses over thresholds with* k *exceedances*

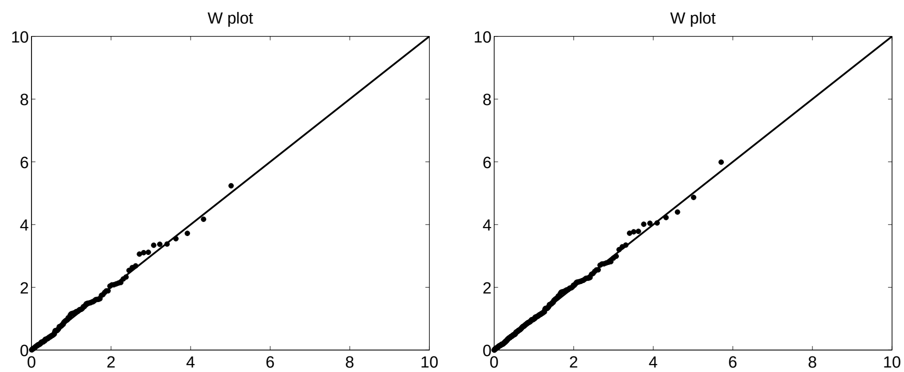

Figure 5: *W-plots assessing the goodness-of-fit of the EGPD to excesses over thresholds with* $k = 150$ *(left) and* $k = 300$ *(right) exceedances for the daily log-returns of the euro-UK pound exchange rate*

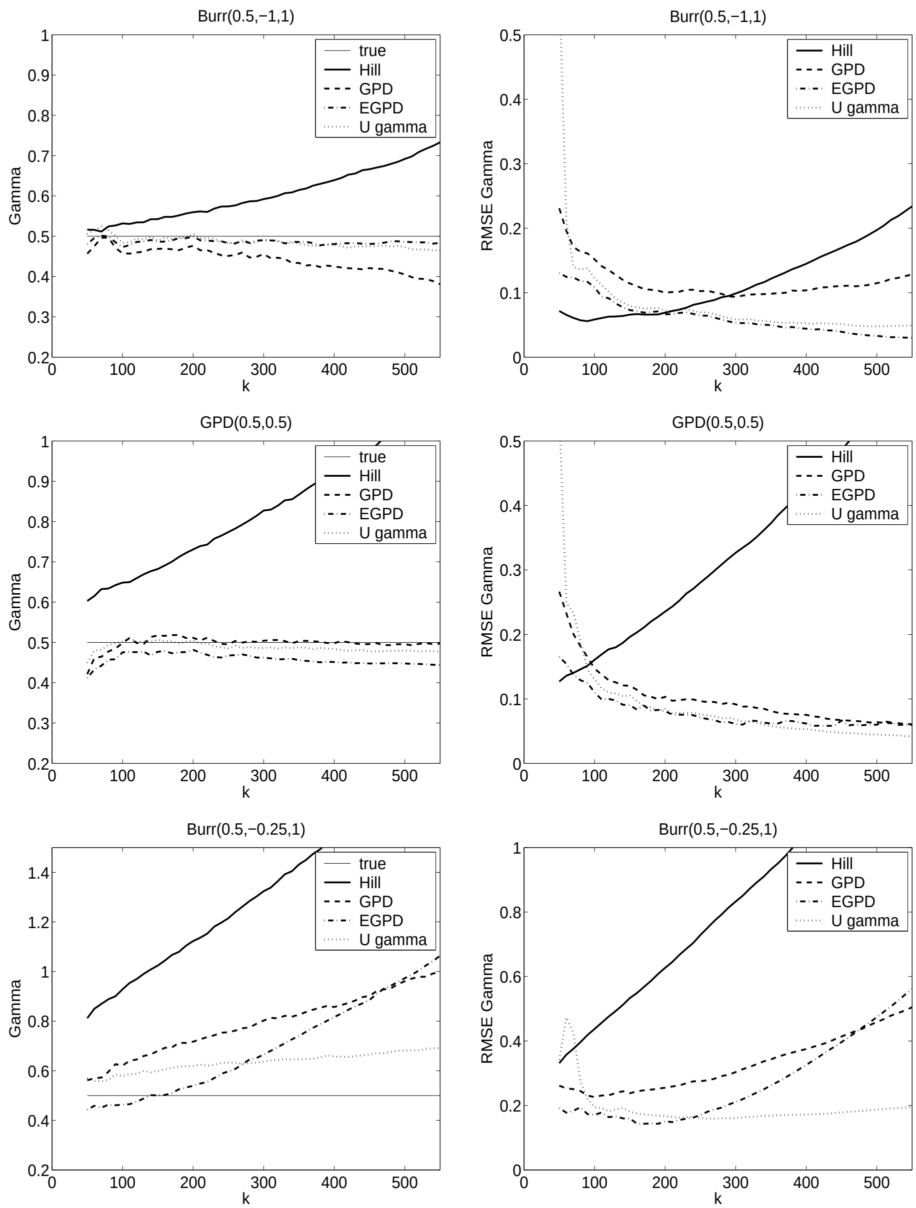

Figure 6: *Medians and RMSEs of $\hat{\gamma}$ as function of k, based on 100 samples of size 1000 from Burr(0.5,-1,1), Burr(0.5,-0.5,1) and Burr(0.5,-0.25,1) distributions*

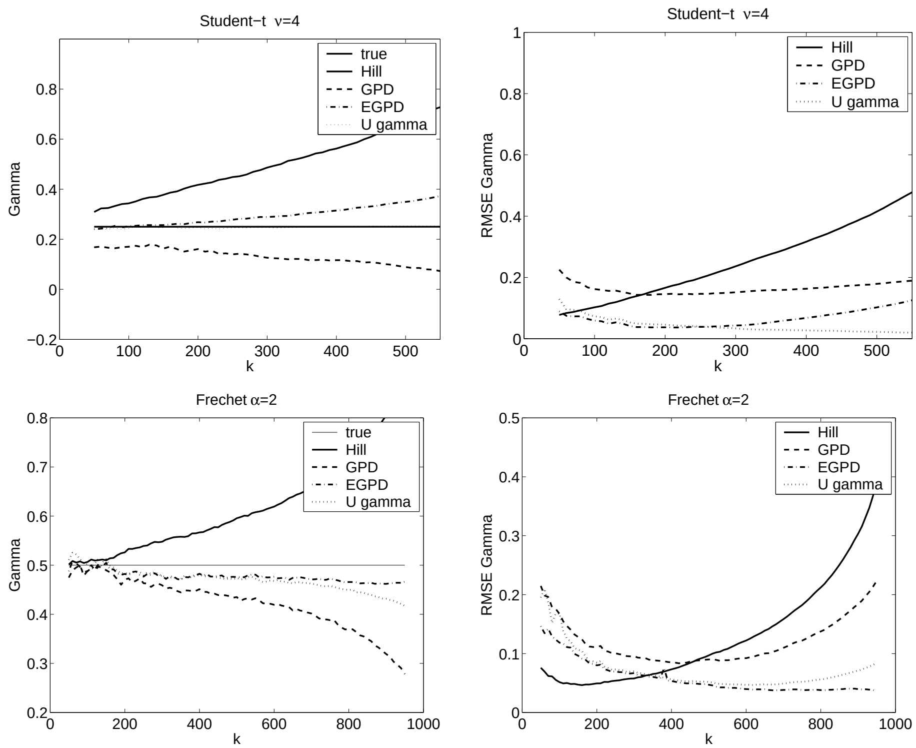

Figure 7: *Medians and RMSEs of* $\hat{\gamma}$ *as function of* k*, based on 100 samples of size 1000 from Student-t* ($\nu = 4$) *and Fréchet* ($\alpha = 2$) *distributions*