Modelling Excesses Over High Thresholds by Perturbed Generalized Pareto Distributions

Jan Beirlant[†]
Elisabeth Joossens[‡]
Johan Segers[§]

July 22, 2002

Abstract. We present a second-order expansion of the classical approximation of the distribution of excesses over a high threshold by a generalized Pareto distribution. Perturbing the generalized Pareto distribution as to absorb the second-order term of this expansion improves the rate of convergence of the approximation. In case the original approximation is poor, tail-related quantities are more accurately estimated in the refined model, especially when the tail is heavy, as shown by simulations. A case study from reinsurance illustrates the new method.

[†]Catholic University Leuven, University Center for Statistics, W. de Croylaan 52b, B-3001 Leuven, Belgium. E-mail: jan.beirlant@wis.kuleuven.ac.be.

[‡]Catholic University Leuven, University Center for Statistics, W. de Croylaan 52b, B-3001 Leuven, Belgium. E-mail: bettie.joossens@wis.kuleuven.ac.be.

[§]EURANDOM, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands. E-mail: segers@eurandom.tue.nl.

i
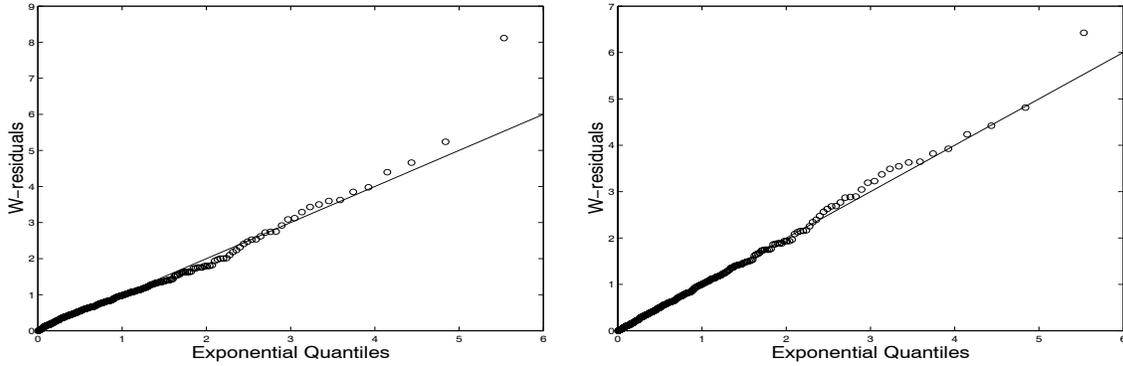
Figure 1: Goodness of fit to large claim data by GPD (left) and perturbed GPD (right).

1   Introduction

A popular procedure in extreme value statistics is to model the excesses of a high threshold in a data-set by a generalized Pareto distribution (GPD), a method commonly referred to as the Peaks-Over-Threshold (POT) method (Smith, 1987; Davison and Smith, 1990). The model enables estimation of large quantiles, small tail probabilities, and in fact any quantity related to the tail of the distribution. Crucial in the whole approach is the accuracy of the GPD approximation to the tail of the population distribution. This concern is far from academic, as textbook distributions show.

We assess the GPD approximation accuracy by a second-order expansion under a second-order refinement of the max-domain of attraction condition. Based on this expansion, we propose an extended family of distributions, coined perturbed generalized Pareto distributions (PGPD). The new family incorporates the second-order term in the expansion, and hence improves the GPD approximation by an order of magnitude.

For the 252 claims in a certain insurance portfolio exceeding 1.1 million euro, a so-called W-plot (Figure 1) assesses the goodness of fit of the GPD and the PGPD (more details in Sections 2 and 3). The GPD (left panel) fits rather unsatisfactorily, especially in the far tail, the main region of interest. The PGPD (right panel) does a much better job.

The improved approximation to the excess-over-threshold distribution leads in many cases to better estimates of the tail of the distribution, especially when the tail is heavy. Moreover, the PGPD can cope with smaller thresholds, that is, it can handle a larger fraction of the data than the GPD is capable of; the threshold choice becomes of much lesser concern. On the other hand, thresholds for which both models correspond closely can be considered appropriate for the classical GPD approach.

The POT method is reviewed in Section 2. A second-order expansion of the GPD approximation serves as motivation for the new PGPD model in Section 3. The simulation study of Section 4 compares the estimators of tail quantities that are generated by the GPD and the PGPD. The appendix contains some computations concerning the PGPD. All figures are gathered at the back of the paper.

2 Peaks-Over-Threshold method

The general framework governing the POT approach and in fact all existing extreme value methods is the assumption that the distribution F is in the domain of attraction of an extreme value distribution. More specifically, let X be a random variable (rv) with distribution function (df) F and tail function $\overline{F} = 1 - F$. Put

$$
\begin{aligned}
\omega &= \sup\{x : F(x) < 1\}, \\
U(t) &= \inf\{x : F(x) \geq 1 - 1/t\}, \quad \text{for } t > 1,
\end{aligned}
$$

the right endpoint and the tail quantile function of F, respectively. In terms of U, the domain of attraction condition reads as follows:

Assumption 2.1 (de Haan, 1970) *There exists a positive function a, regularly varying at infinity with index $\gamma \in \mathbb{R}$, such that*

$$
\lim_{t \to \infty} \frac{U(ty) - U(t)}{a(t)} = \int_1^y v^{\gamma - 1} \, \mathrm{d}v =: h_\gamma(y), \quad \text{for } y > 0. \tag{1}
$$

The index γ is called the extreme value index.

The POT procedure is based on a result by Pickands (1975) which rephrases the extreme-value condition (1) in terms of the distribution of excesses over a high threshold u. This result guarantees that the excess $X - u$, conditionally on $X > u$, is approximately generalized Pareto distributed as $u \uparrow \omega$. Let G_γ be the df of the GPD with shape parameter γ, that is, for $x \geq 0$,

$$
\overline{G}_\gamma(x) = 1 - G_\gamma(x) = 1/h_\gamma^{\leftarrow}(x) = \begin{cases} (1 + \gamma x)^{-1/\gamma} & \text{if } \gamma \neq 0,\ 1 + \gamma x > 0; \\ 0 & \text{if } \gamma < 0,\ x \geq 1/|\gamma|; \\ e^{-x} & \text{if } \gamma = 0, \end{cases} \tag{2}
$$

where $h_\gamma^{\leftarrow}$ is the inverse function of h_γ in $x \geq 0$, i.e. $h_\gamma^{\leftarrow}(x) = \inf\{y \geq 1 : h_\gamma(y) \geq x\}$.

Proposition 2.2 (Pickands, 1975) *Put $\sigma(u) = a(1/\overline{F}(u))$ for $u < \omega$, with a as in Assumption 2.1. Equation (1) is equivalent to*

$$
\lim_{u \uparrow \omega} \Pr\left(\frac{X - u}{\sigma(u)} \leq x \,\middle|\, X > u \right) = G_\gamma(x), \quad \text{for } x \geq 0. \tag{3}
$$

By Proposition 2.2, the excesses in a sample over a sufficiently high threshold can be approximately described by a scaled GPD with df $G_\gamma(\cdot/\sigma)$, for some real γ and positive σ. Maximum likelihood or other estimation methods lead to estimates $\hat{\gamma}$ and $\hat{\sigma}$ of the GPD parameters. Furthermore, the approximation $P(X > y) \approx P(X > u)\overline{G}_\gamma(\frac{y-u}{\sigma})$ yields estimators for various tail quantities, as we will describe next.

Let N_u be the number of observations in a sample of size n that exceed the threshold u. The tail function is estimated by

$$\widehat{\overline{F}}(x) = \frac{N_u}{n} \overline{G}_{\hat{\gamma}} \left(\frac{x-u}{\hat{\sigma}} \right), \quad \text{for } x > u. \tag{4}$$

Invert equation (4) to obtain an estimator for the $(1-p)$-quantile $x_p = U(1/p)$:

$$\hat{x}_p = u + \hat{\sigma} G_{\hat{\gamma}}^{\leftarrow} \left(1 - \frac{np}{N_u} \right), \quad \text{for } 0 < p < N_u/n, \tag{5}$$

where $G_{\hat{\gamma}}^{\leftarrow}$ denotes the quantile function of the GPD.

In an excess-loss reinsurance treaty, the reinsurer takes responsibility for the overshoot $(X - R)_+$ of a claim X over a fixed priority R. The net premium for such a contract is

$$EXL_R = E(X - R)_+ = \int_R^{\infty} \overline{F}(x) \, \mathrm{d}x.$$

Since the priority R is typically in the upper range of the already observed claims, the POT method is a natural candidate to estimate the net premium (McNeil, 1997; Beirlant *et al.*, 2001): when $R > u$

$$\widehat{EXL}_R = \int_R^{\infty} \widehat{\overline{F}}(x) \, \mathrm{d}x = \frac{N_u}{n} \int_R^{\infty} \overline{G}_{\hat{\gamma}} \left(\frac{x-u}{\hat{\sigma}} \right) \mathrm{d}x, \tag{6}$$

provided $\hat{\gamma} < 1$. The mean excess function m is a conditional version of EXL_R:

$$m_R = E(X - R \mid X > R) = \int_R^{\infty} \frac{\overline{F}(x)}{\overline{F}(R)} \, \mathrm{d}x,$$

If $R > u$ and $\hat{\gamma} < 1$, the POT estimator is

$$\hat{m}_R = \left(\overline{G}_{\hat{\gamma}} \left(\frac{R-u}{\hat{\sigma}} \right) \right)^{-1} \int_R^{\infty} \overline{G}_{\hat{\gamma}} \left(\frac{x-u}{\hat{\sigma}} \right) \mathrm{d}x. \tag{7}$$

Case study. As a practical example from reinsurance, we consider a set of 252 claims from a given line of business. The available data concern claims of at least 1.1 million euro. The portfolio is protected by an excess-loss reinsurance treaty with priority R set at 5 million euro. The reinsurance company hopes to combine a good tail fit with a good global fit, that is, taking the threshold u as low as possible, ideally down to 1.1 million euro. The left column of Figure 2 shows the estimates $\hat{\gamma}$, $\hat{x}_{0.001}$, and $\widehat{EXL}_R$ as a function of $N_u = k$, where the threshold u is taken as the $(k+1)$th largest observation $x_{n-k,n}$, $k = 1, ..., 251$. The estimates show a clear downward trend in k, raising the question which threshold $u = x_{n-k,n}$ to choose. To evaluate the goodness-of-fit of the GPD for the threshold u equal to 1.1 million euro, we constructed in the left panel of Figure 1 the so-called W-plot (Smith and Shively, 1995), consisting of the pairs

$$\left(-\log \left(1 - G_{\hat{\gamma}}((X_{j,n} - u)/\hat{\sigma}) \right), -\log(1 - j/253) \right), \quad \text{for } j = 1, \dots, 252.$$

The deviation of these pairs from the first diagonal indicates the poor fit by the GPD.

3

3 The perturbed GPD model

If, as in the previous example, the GPD model turns out too restrictive, then it seems natural to look for a wider class of models (paragraph 3.2). The new model is justified by a second-order expansion of the GPD-approximation (paragraph 3.1).

3.1 Second-order expansion of the GPD-approximation

In order to assess the accuracy of the GPD-approximation in Proposition 2.2, we need to refine the extreme value condition of Assumption 2.1 as in Pereira (1994).

Assumption 3.1 *There exists a positive function A such that $\lim_{t\to\infty} A(t) = 0$ and a function $H : (0,\infty) \to \mathbb{R}$ such that*

$$\lim_{t\to\infty} \frac{1}{A(t)} \left(\frac{U(ty) - U(t)}{a(t)} - h_\gamma(y) \right) = H(y), \quad for \; y > 0. \tag{8}$$

For $\gamma \in \mathbb{R}$ and $\rho \leq 0$, define

$$H_{\gamma,\rho}(x) = \int_1^x v^{\gamma-1} h_\rho(v) \, \mathrm{d}v, \quad \text{for } x > 0.$$

The theory of generalized regular variation of second order provides a characterization of the limit function H, see Theorem 1 of de Haan and Stadtmüller (1996).

Proposition 3.2 *Under Assumption 3.1, there exist constants $c_1, c_2 \in \mathbb{R}$ and $\rho \leq 0$ such that*

$$H(y) = c_1 H_{\gamma,\rho}(y) + c_2 h_{\gamma+\rho}(y), \quad for \; y > 0.$$

If $\rho < 0$, then an appropriate choice of the auxiliary function a results in a simplification of the limit function H.

Proposition 3.3 *Under Assumption 3.1 and with c_1, c_2, and ρ as in Proposition 3.2, if $\rho < 0$, then*

$$\lim_{t\to\infty} \frac{1}{A(t)} \left(\frac{U(ty) - U(t)}{\widetilde{a}(t)} - h_\gamma(y) \right) = c h_{\gamma+\rho}(y),$$

where $c = \rho^{-1} c_1 + c_2$ and $\widetilde{a}(t) = a(t) \left(1 - \rho^{-1} c_1 A(t) \right)$ for all t such that $A(t) < |\rho/c_1|$.

Proof. Since $\rho < 0$, we have for all $y > 0$

$$H_{\gamma,\rho}(y) = \int_1^y v^{\gamma-1} \frac{v^\rho - 1}{\rho} \, \mathrm{d}v = \rho^{-1} \left(h_{\gamma+\rho}(y) - h_\gamma(y) \right).$$

Hence by Proposition 3.2, for all $y > 0$,

$$H(y) = (\rho^{-1} c_1 + c_2) h_{\gamma+\rho}(y) - \rho^{-1} c_1 h_\gamma(y).$$

But then by Assumption 3.1, we have for all $y > 0$

$$\frac{1}{A(t)} \left(\frac{U(ty) - U(t)}{\widetilde{a}(t)} - h_\gamma(y) \right)$$

$$= \frac{1}{A(t)(1 - \rho^{-1}c_1 A(t))} \left(\frac{U(ty) - U(t)}{a(t)} - h_\gamma(y) \right) + \frac{1}{1 - \rho^{-1}c_1 A(t)} \rho^{-1} c_1 h_\gamma(y)$$

$$\rightarrow H(y) + \rho^{-1} c_1 h_\gamma(y) = c h_{\gamma + \rho}(y),$$

as $t \rightarrow \infty$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The tail quantile approximation in Assumption 2.1 gives rise to the GPD approximation in Proposition 2.2. In the same way, the second-order tail quantile approximation in Assumption 3.1 gives rise to a second-order GPD approximation in Proposition 3.4 below.

We need some notation. The support of the GPD G_γ is the interval $I_\gamma = \{x \geq 0 : 1 + \gamma x > 0\}$. The probability density function g_γ of G_γ in $x \in I_\gamma$ is given by

$$g_\gamma(x) = [(1 + \gamma x) h_\gamma^\leftarrow(x)]^{-1} = \begin{cases} (1 + \gamma x)^{-1/\gamma - 1} & \text{if } \gamma \neq 0, \\ e^{-x} & \text{if } \gamma = 0. \end{cases}$$

We also define $g_\gamma(x) = 0$ for x outside I_γ.

Proposition 3.4 *Under Assumption 3.1, we have for $x \in I_\gamma$*

$$\lim_{u \uparrow \omega} \frac{1}{B(u)} \left(\Pr \left(\frac{X - u}{\sigma(u)} > x \,\middle|\, X > u \right) - \overline{G}_\gamma(x) \right) = g_\gamma(x) H(h_\gamma^\leftarrow(x)),$$

where $B(u) = A(1/\overline{F}(u))$ and $\sigma(u) = a(1/\overline{F}(u))$.

Proof. Since $1/\overline{F}$ is the right-continuous inverse of U, we have by Theorem 3 of de Haan and Stadtmüller (1996)

$$\lim_{u \uparrow \omega} \frac{1}{B(u)} \left(\frac{1/\overline{F}(u + x\sigma(u))}{1/\overline{F}(u)} - h_\gamma^\leftarrow(x) \right) = -(1 + \gamma x)^{-1} h_\gamma^\leftarrow(x) H(h_\gamma^\leftarrow(x)).$$

Since $s^{-1} - t^{-1} = (t - s)/(st)$, the Proposition follows. $\qquad\qquad\qquad\qquad\qquad\square$

Corollary 3.5 *Under the assumptions of Proposition 3.3, we have*

$$\lim_{u \uparrow \omega} \frac{1}{B(u)} \left(\Pr \left(\frac{X - u}{\widetilde{\sigma}(u)} > x \,\middle|\, X > u \right) - \overline{G}_\gamma(x) \right) = c g_\gamma(x) h_{\gamma + \rho}(h_\gamma^\leftarrow(x)), \quad \text{for } x \in I_\gamma, \quad (9)$$

where $\widetilde{\sigma}(u) = \widetilde{a}(1/\overline{F}(u))$.

Under Assumption 3.1, the GPD approximates the distribution of the excess over the threshold u up to error $O(B(u))$, see Proposition 3.4. If the rate at which B vanishes is slow, this error might be appreciably large. Especially the modelling of the extreme tail of the excess distribution by a GPD might be poor.

5

3.2 Perturbed GPD

First we introduce some notation. For $\tau < 0$, we extend the definition of h_τ to $h_\tau(\infty) = \int_1^\infty v^{\tau-1}\,dv = 1/|\tau|$. For $\gamma \in \mathbb{R}$ and $\rho < 0$, define the function $\phi_{\gamma,\rho} : [0,\infty) \to [0,\infty)$ by

$$\phi_{\gamma,\rho}(x) = h_{\gamma+\rho}(h_\gamma^\leftarrow(x)) \tag{10}$$

$$= \begin{cases} \dfrac{(1+\gamma x)^{1+\rho/\gamma}-1}{\gamma+\rho} & \text{if } \gamma \neq 0,\ \rho+\gamma \neq 0,\ 1+\gamma x > 0, \\[2mm] \gamma^{-1}\log(1+\gamma x) & \text{if } \gamma+\rho = 0, \\[2mm] \dfrac{e^{\rho x}-1}{\rho} & \text{if } \gamma = 0, \\[2mm] 1/|\gamma+\rho| & \text{if } \gamma < 0,\ x \geq 1/|\gamma|. \end{cases}$$

The left derivative of $\phi_{\gamma,\rho}$ is $\phi'_{\gamma,\rho}(x) = (h_\gamma^\leftarrow(x))^\rho$, for $x \geq 0$, where we set $\infty^\rho := 0$. In particular, $0 \leq \phi'_{\gamma,\rho}(x) \leq 1$ for $x \geq 0$.

Definition 3.6 *The perturbed generalized Pareto distribution (PGPD) with parameters $\gamma \in \mathbb{R}$, $\rho < 0$, and $\delta > -1$, is given by its distribution function*

$$G_{\gamma,\rho,\delta}(x) = G_\gamma(x + \delta\phi_{\gamma,\rho}(x)), \quad \text{for } x \geq 0.$$

By the properties of $\phi_{\gamma,\rho}$ listed above, the function

$$[0,\infty) \to [0,\infty) : x \mapsto x + \delta\phi_{\gamma,\rho}(x)$$

is strictly increasing, one-to-one and onto. Hence $G_{\gamma,\rho,\delta}$ is indeed a distribution function. We can include an additional scale parameter $\sigma > 0$ in the PGPD by setting

$$G_{\gamma,\sigma,\rho,\delta}(x) = G_{\gamma,\rho,\delta}(x/\sigma), \quad \text{for } x \geq 0.$$

Computational details for the PGPD are provided in the appendix.

To appreciate the improvement of the PGPD over the GPD as an approximation to the excess-over-threshold distribution, compare Proposition 3.7 with Corollary 3.5.

Proposition 3.7 *Under the assumptions of Proposition 3.3, we have*

$$\lim_{u\uparrow\omega} \frac{1}{B(u)}\left(\Pr\left(\left.\frac{X-u}{\widetilde{\sigma}(u)} > x \;\right|\; X > u\right) - \overline{G}_{\gamma,\rho,-cB(u)}(x)\right) = 0, \quad \text{for } x \in I_\gamma,$$

where $\widetilde{\sigma}(u) = \widetilde{a}(1/\overline{F}(u))$.

Proof. Take $x \in I_\gamma$. For δ close enough to zero, we have $x + \delta\phi_{\gamma,\rho}(x) \in I_\gamma$. Hence

$$\left.\frac{\partial}{\partial\delta}G_{\gamma,\rho,\delta}(x)\right|_{\delta=0} = g_\gamma(x)\phi_{\gamma,\rho}(x). \tag{11}$$

Hence

$$
\frac{1}{B(u)} \left(\Pr\left(\frac{X-u}{\widetilde{\sigma}(u)} > x \,\middle|\, X > u \right) - \overline{G}_{\gamma,\rho,-cB(u)}(x) \right)
$$

$$
= \frac{1}{B(u)} \left(\Pr\left(\frac{X-u}{\widetilde{\sigma}(u)} > x \,\middle|\, X > u \right) - \overline{G}_{\gamma}(x) \right) - \frac{1}{B(u)} \left(\overline{G}_{\gamma,\rho,-cB(u)}(x) - \overline{G}_{\gamma}(x) \right)
$$

$$
\rightarrow \ c g_{\gamma}(x)\phi_{\gamma,\rho}(x) - c g_{\gamma}(x)\phi_{\gamma,\rho}(x) = 0,
$$

by Corollary 3.5 and equation (11). $\qquad\qquad\square$

Parallel to the classical POT method based on the GPD, we propose to model the excesses over a large threshold by a PGPD with df $G_{\gamma,\sigma,\rho,\delta}$. Maximum likelihood provides a natural way to estimate the parameters. Estimators of tail quantities as in the introduction are obtained by replacing the GPD $G_{\hat{\gamma}}$ in (4), (5), (6), and (7) by a PGPD $G_{\hat{\gamma},\hat{\rho},\hat{\delta}}$.

Some caution is needed for the estimation of ρ: for $\delta = 0$, the parameter ρ is not identifiable, and for δ close to zero, the likelihood surface is typically rather flat along ρ, reflecting the lack of information on ρ in a sample. To avoid numerical instabilities in the estimation procedure, the search interval for ρ needs to be restricted in advance, for instance to $[-2, -0.2]$. Alternatively, we consider the option of fixing the unknwon parameter ρ to a value $\bar{\rho}$, say, typically -0.5 or -1.

Case study (continued). We continue our analysis of Section 2 of the large insurance claims, this time replacing the GPD by the PGPD. The modified W-plot (right panel of Figure 1), consisting of the pairs

$$
\left(-\log\left(1 - G_{\hat{\gamma},\hat{\sigma},\hat{\rho},\hat{\delta}}(X_{j,n} - u) \right), -\log(1 - j/253) \right), \quad \text{for } j = 1, \ldots, 252,
$$

illustrates the good PGPD fit to the conditional excess distribution at threshold u equal to 1.1 million euro. At the same u, a likelihood ratio test of the GPD versus the PGPD gives a p-value of 0.037. The extra parameters in the new model cause the PGPD estimates of γ, $x_{0.001}$, and EXL_R to oscillate more than the GPD-based ones (Figure 2), but the threshold-dependent bias of the latter, reflected by a downward trend, has been removed. Overall, the PGPD point estimates at the lowest threshold (largest k) seem acceptable.

4 Simulations

We performed a simulation study to compare the finite sample behaviour of the novel PGPD-based and the classical GPD-based estimators for the extreme value index ($\hat{\gamma}$), extreme quantiles ($\hat{x}_p$) with $p = 1/n$, Expected Excess Losses ($\widehat{EXL_R}$) and mean excesses ($\hat{m}_R$) with R fixed at $Q(1 - 1/n) = U(n)$. We restricted the search range for ρ to the interval $[-2, -0.2]$, and we also investigated the effect of fixing the parameter ρ to a value $\bar{\rho} = -0.5$ or -1, which may or may not be equal to the true value. We generated samples

$\gamma > 0$	• Burr(γ,ρ,β) distribution $1 - F(x) = \left(1 + x^{-\rho/\gamma}/\beta\right)^{1/\rho}$ $(x > 0)$ with $(\gamma, \rho, \beta) = (0.5, -1, 1), (0.5, -0.5, 1),$ and $(0.5, -0.25, 1)$
	• Student-t distribution with $\nu = 4$ degrees of freedom, so $\gamma = 1/\nu = 0.25$ and $\rho = -2/\nu = -0.5$
$\gamma = 0$	• standard normal distribution
	• standard log-normal distribution
$\gamma < 0$	• Beta distribution with density $B(\alpha, \beta)^{-1}x^{\alpha-1}(1-x)^{\beta-1}$ $(0 < x < 1)$ with $\alpha = 5$ and $\beta = 4$, so $\gamma = -1/\beta = -0.25$ and $\rho = -(\alpha-1)/\beta = -1$

Table 1: Distributions involved in the simulation study.

from the distributions listed in Table 1. The threshold u was set at the $(k + 1)$th largest order statistic $x_{n-k,n}$ for various k.

Figures 3 and 4 show the results for 100 simulated Burr($0.5, -0.25, 1$) samples of size 500. Histograms (Figure 3) contrast the sampling distributions of $\hat{\gamma}$ at $k = 200$ for the GPD model and the PGPD model. The improvement over the GPD approach is striking, especially with respect to bias. The left panel of Figure 4 depicts the construction of the 95% profile likelihood confidence interval for γ for one randomly chosen Burr(0.5,-0.25,1) sample. Observe that the PGPD interval, although somewhat wider due to the extra parameters to be estimated, is better located around the true value. The right panel shows the increased empirical coverage probability of the PGPD confidence intervals.

In the subsequent figures, we present the results for all distributions mentioned above. The left-hand panels show the medians and the right-hand panels show the empirical root mean squared errors (RMSE) of the estimators as a function of $k = 50, 60, \ldots, 990$, computed from 100 samples of size 1000. In most cases, the PGPD based estimators have the smaller bias, and the same is true for the minimal RMSE. A notable exception is the Burr(0.5,-0.5,1) distribution: if $\gamma + \rho = 0$, then the GPD based estimators possess zero asymptotic bias (Smith, 1987).

References

Beirlant, J., Matthys, G. and Dierckx, G. (2001). Heavy-tailed distributions and rating, *Astin Bulletin* **31**, 41–62.

Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds, *J. R. Stat. Soc., Ser. B* **52**, 393–442.

de Haan, L. (1970). *On regular variation and its applications to the weak convergence of sample extremes.* Math. Centre Tract **72**, Amsterdam.

de Haan, L. and Stadmüller, U. (1996). Generalized regular variation of second order, *J. Austral.*

Math. Soc. (Series A) **61**, 381–395.

McNeil, A.J. (1997). Estimating the tails of loss severity distributions using extreme value theory, *ASTIN Bulletin* **27**, 117–137.

Pereira, T.T. (1994). Second order behavior of domains of attraction and the bias of generalized Pickands' estimator, in: *Extreme Value Theory and Applications* (eds: J. Galambos, J. Lechner and E. Simiu), NIST Special Publication **860**, 165–178.

Pickands, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.* **3**, 119–131.

Smith, R.L. (1987). Estimating tails of probability distributions, *Ann. Statist.* **15**, 1174–1207.

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

A   Computational aspects

For the PGPD with parameters $\gamma \in \mathbb{R}$, $\sigma > 0$, $\rho < 0$, and $\delta > -1$, we compute the distribution function, the density and the quantile function in paragraph A.1, and the log-likelihood with its derivatives in paragraph A.2.

A.1   Distribution function, density, and quantile function

For $\gamma \in \mathbb{R}$ and $x \in I_\gamma = \{y \geq 0 : 1 + \gamma y > 0\}$, define

$$K_\gamma(x) = \int_0^x \frac{\mathrm{d}v}{1 + \gamma v} = \begin{cases} \gamma^{-1} \log(1 + \gamma x) & \text{if } \gamma \neq 0, \\ x & \text{if } \gamma = 0. \end{cases} \tag{12}$$

Recall the function $\phi_{\gamma,\rho}$ of equation (10). For $x \geq 0$, let

$$z = x/\sigma, \quad \text{and} \quad t = t(z) = z + \delta\phi_{\gamma,\rho}(z). \tag{13}$$

The support of the PGPD is the interval $S = \{x \geq 0 : 1 + \gamma t > 0\}$. The function t is strictly increasing, concave if $\delta > 0$, and convex if $\delta < 0$. The inverse function of t is denoted by $t^\leftarrow$ (to be computed numerically).

- *Distribution function in $x \in S$*

$$G_{\gamma,\sigma,\rho,\delta}(x) = 1 - \exp\{-K_\gamma(t)\}$$

- *Density function in $x \in S$*

$$g_{\gamma,\sigma,\rho,\delta}(x) = \begin{cases} \sigma^{-1}\left(1 + \delta \exp(\rho K_\gamma(z))\right) \exp\left\{-(1+\gamma)K_\gamma(t)\right\} & \text{if } 1 + \gamma z > 0 \\ \sigma^{-1} \exp\left\{-(1+\gamma)K_\gamma(t)\right\} & \text{if } \gamma < 0, \, z \geq 1/|\gamma| \end{cases}$$

- *Quantile function in $p \in (0,1)$*

$$Q(p) = \sigma t^\leftarrow(h_\gamma(1/(1-p)))$$

9

A.2 Log-likelihood and derivatives

For $\gamma \in \mathbb{R}$ and $x \geq 0$, define

$$H_\gamma(x) \;=\; \int_0^x e^{\gamma v}\,\mathrm{d}v = \begin{cases} \gamma^{-1}\left(e^{\gamma x} - 1\right) & \text{if } \gamma \neq 0, \\ x & \text{if } \gamma = 0; \end{cases}$$

$$H_\gamma^*(x) = \frac{\partial}{\partial \gamma} H_\gamma(x) \;=\; \int_0^x e^{\gamma v} v\,\mathrm{d}v = \begin{cases} \gamma^{-2}\left(e^{\gamma x}(\gamma x - 1) + 1\right) & \text{if } \gamma \neq 0, \\ x^2/2 & \text{if } \gamma = 0. \end{cases}$$

Recall K_γ from equation (12), and for $\gamma \in \mathbb{R}$ and $x \in I_\gamma$, put

$$K_\gamma^*(x) = \frac{\partial}{\partial \gamma} K_\gamma(x) = -\int_0^x \frac{v\,\mathrm{d}v}{(1 + \gamma v)^2} = \begin{cases} \gamma^{-2}\left(\dfrac{\gamma x}{1 + \gamma x} - \log(1 + \gamma x)\right) & \text{if } \gamma \neq 0, \\ -x^2/2 & \text{if } \gamma = 0. \end{cases}$$

The derivatives of $\phi_{\gamma,\rho}(x)$ w.r.t. γ and ρ are

$$\frac{\partial}{\partial \gamma} \phi_{\gamma,\rho}(x) \;=\; \begin{cases} H_{\gamma+\rho}^*(K_\gamma(x)) + K_\gamma^*(x)\exp\left((\gamma + \rho)K_\gamma(x)\right) & \text{if } 1 + \gamma x > 0, \\ (\gamma + \rho)^{-2} & \text{if } \gamma < 0,\ x \geq 1/|\gamma|; \end{cases}$$

$$\frac{\partial}{\partial \rho} \phi_{\gamma,\rho}(x) \;=\; \begin{cases} H_{\gamma+\rho}^*(K_\gamma(x)) & \text{if } 1 + \gamma x > 0, \\ (\gamma + \rho)^{-2} & \text{if } \gamma < 0,\ x \geq 1/|\gamma|. \end{cases}$$

Recall z and t from equation (13), and let

$$v = \frac{1 + \gamma}{1 + \gamma t} \quad \text{and} \quad s = \begin{cases} \exp\{\rho K_\gamma(z)\} & \text{if } 1 + \gamma z > 0, \\ 0 & \text{if } \gamma \leq 0,\ z \geq 1/|\gamma|. \end{cases}$$

For $x \in S$, we have the following formulas.

- *Log-likelihood*

$$\ell \equiv \ell_{\gamma,\sigma,\rho,\delta}(x) = -\log(\sigma) + \log(1 + \delta s) - (1 + \gamma)K_\gamma(t)$$

- *Derivative log-likelihood w.r.t. γ*

$$\frac{\partial}{\partial \gamma}\ell = \begin{cases} \dfrac{\delta \rho s K_\gamma^*(z)}{1 + \delta s} - K_\gamma(t) - (1 + \gamma)K_\gamma^*(t) - \delta v \dfrac{\partial}{\partial \gamma}\phi_{\gamma,\rho}(z) & \text{if } 1 + \gamma z > 0 \\ -K_\gamma(t) - (1 + \gamma)K_\gamma^*(t) - \delta v(\gamma + \rho)^{-2} & \text{if } \gamma < 0,\ z \geq 1/|\gamma| \end{cases}$$

- *Derivative log-likelihood w.r.t. σ*

$$\frac{\partial}{\partial \sigma}\ell = \begin{cases} -\dfrac{1}{\sigma} - \dfrac{\delta \rho s z}{\sigma(1 + \delta s)(1 + \gamma z)} + \dfrac{z}{\sigma} v(1 + \delta s) & \text{if } 1 + \gamma z > 0 \\ -\dfrac{1}{\sigma} + \dfrac{z}{\sigma} v & \text{if } \gamma < 0,\ z \geq 1/|\gamma| \end{cases}$$

10

- *Derivative log-likelihood w.r.t. ρ*

$$\frac{\partial}{\partial \rho}\ell = \begin{cases} \dfrac{\delta s K_\gamma(z)}{1 + \delta s} - \delta v H^*_{\gamma+\rho}(K_\gamma(z)) & \text{if } 1 + \gamma z > 0 \\ -\delta v (\gamma + \rho)^{-2} & \text{if } \gamma < 0,\ z \geq 1/|\gamma| \end{cases}$$

- *Derivative log-likelihood w.r.t. δ*

$$\frac{\partial}{\partial \delta}\ell = \frac{s}{1 + \delta s} - v\phi_{\gamma,\rho}(z)$$
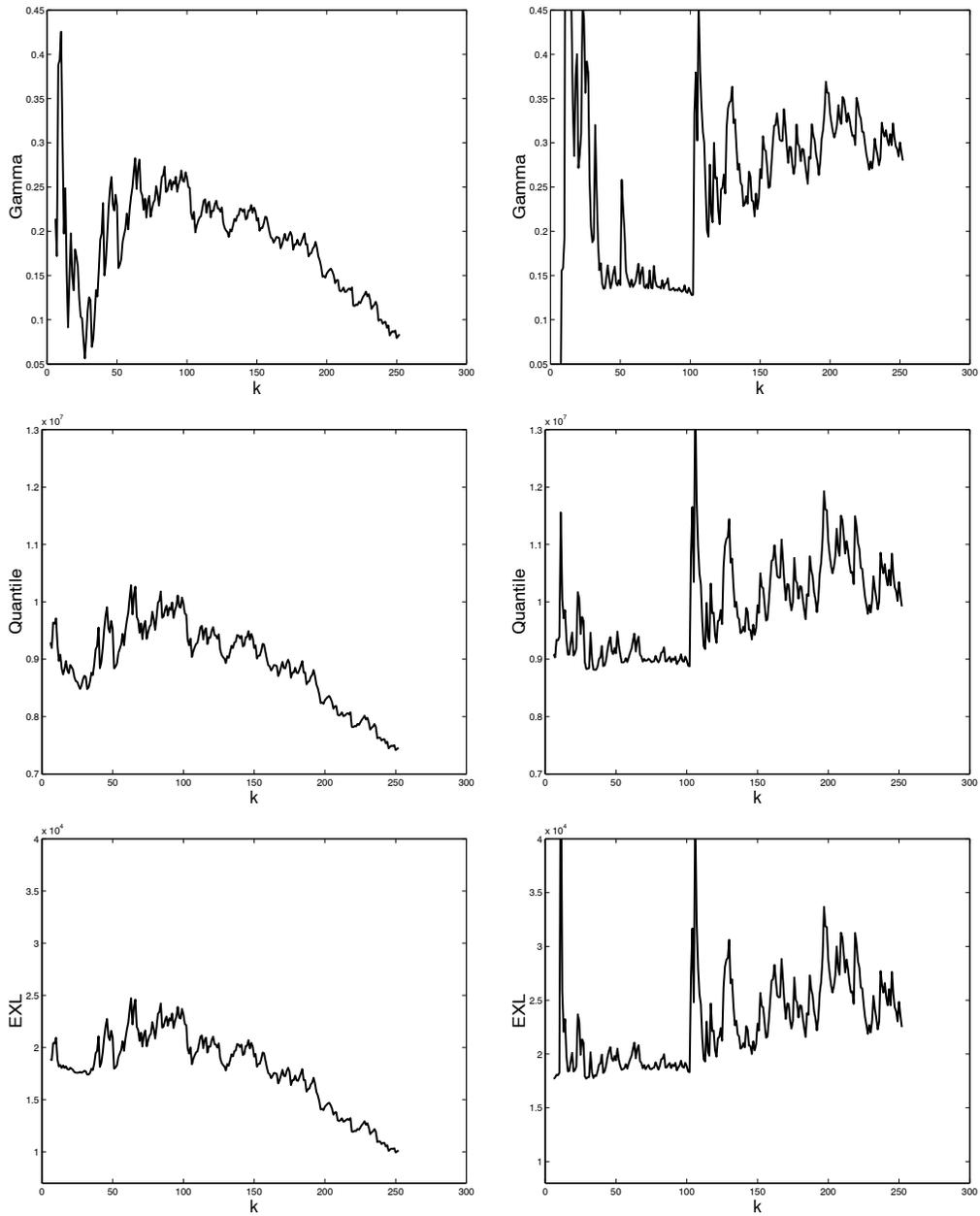
Figure 2: Large claim data: estimates of γ (top), $x_{0.001}$ (middle), and EXL_R (bottom) based on GPD (left) and PGPD (right).
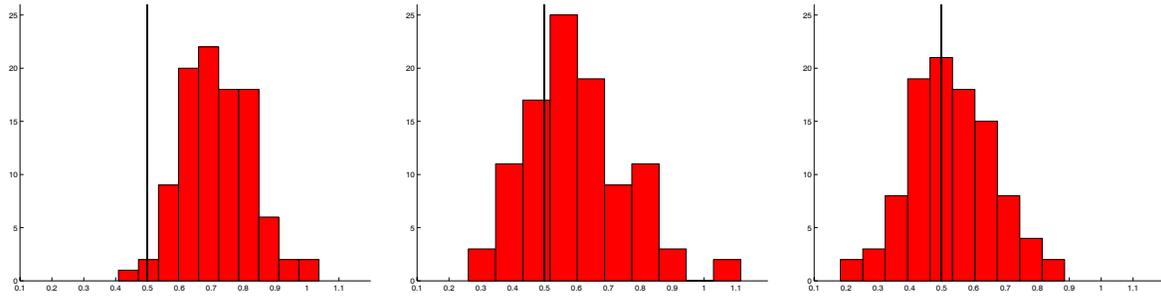
Figure 3: Histogram of γ-estimates for 100 Burr (0.5,-0.25,1) samples of size 500 with k=200: GPD (left), PGPD with $\overline{\rho}$=-0.5 (middle), and PGPD with $\overline{\rho} \in$[-2, -0.2] (right).
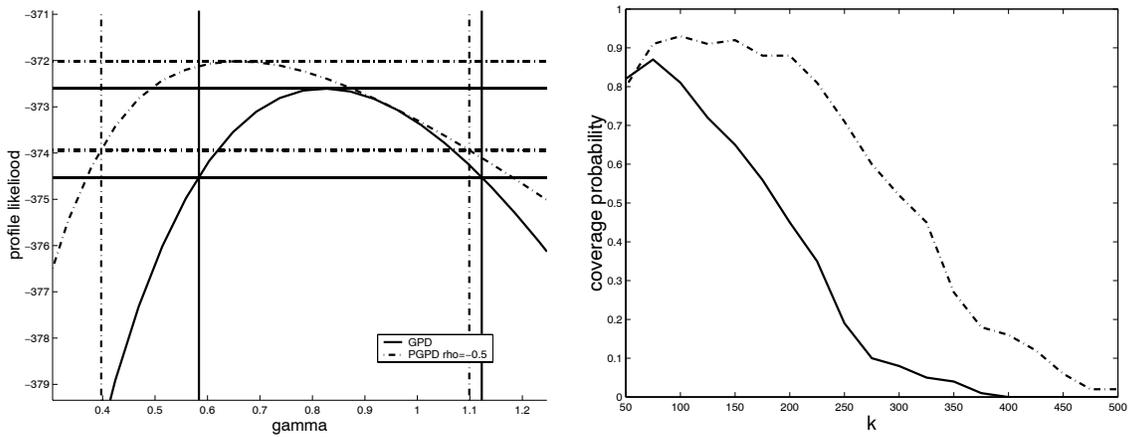
Figure 4: *Left panel:* construction of 95% profile likelihood confidence interval for $\gamma = 0.5$ from a Burr(0.5,-0.25,1) sample of size $n = \ldots$, with k=200, based on GPD (solid) and PGPD with $\overline{\rho}$=-0.5 (dash-dotted). *Right panel:* empirical coverage percentages for 100 such intervals.
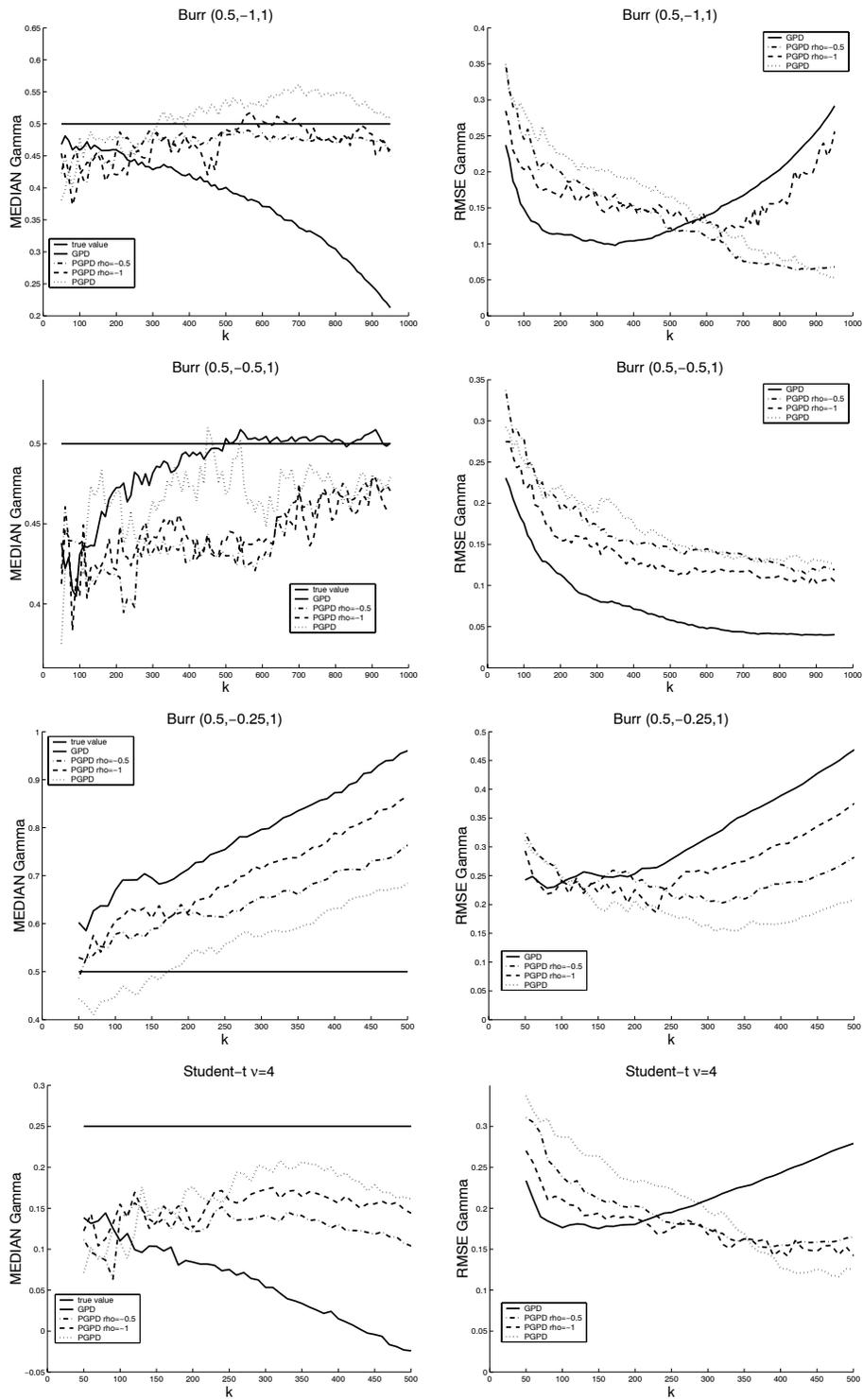
Figure 5: Medians and RMSEs of $\hat{\gamma}$ as function of k, based on 100 samples of size 1000 from Burr(0.5,-1,1), Burr(0.5,-0.5,1), Burr(0.5,-0.25,1) and Student-t (ν=4) distribution.

14
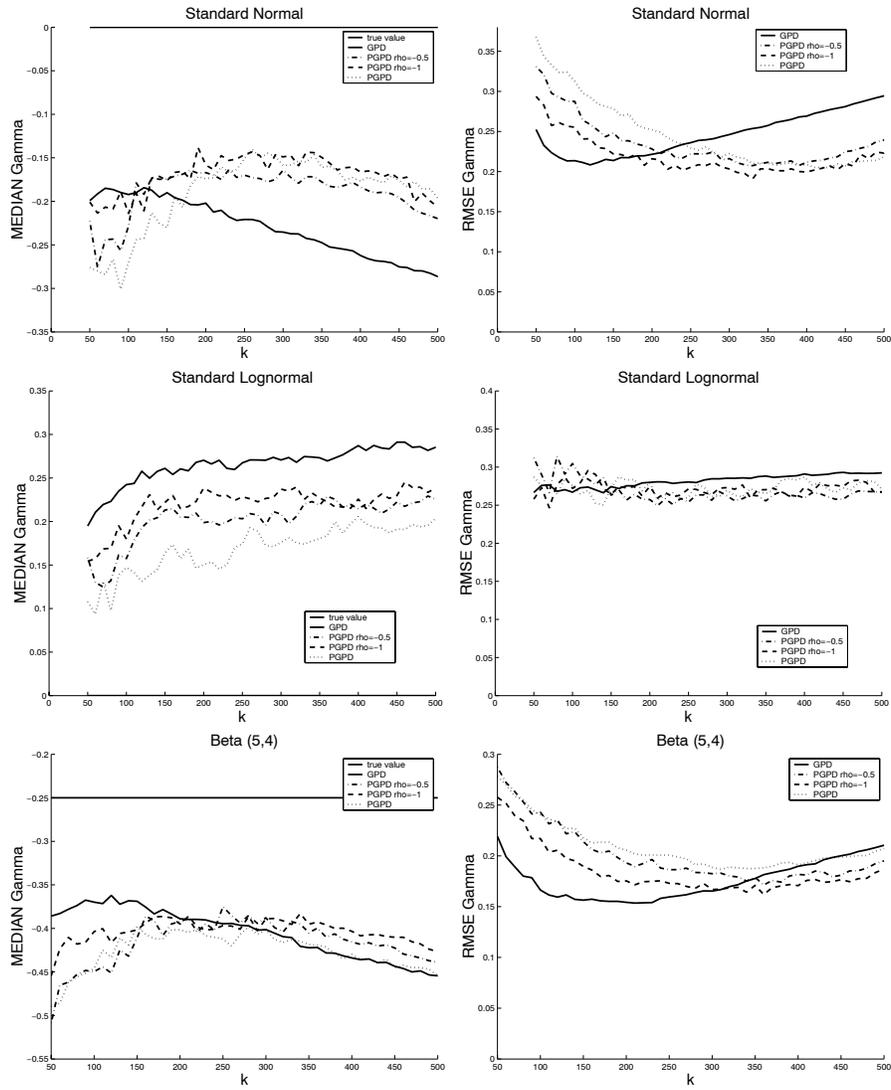
Figure 6: Medians and RMSEs of $\hat{\gamma}$ as function of k, based on 100 samples of size 1000 from standard normal, standard lognormal and Beta(5,4) distribution.
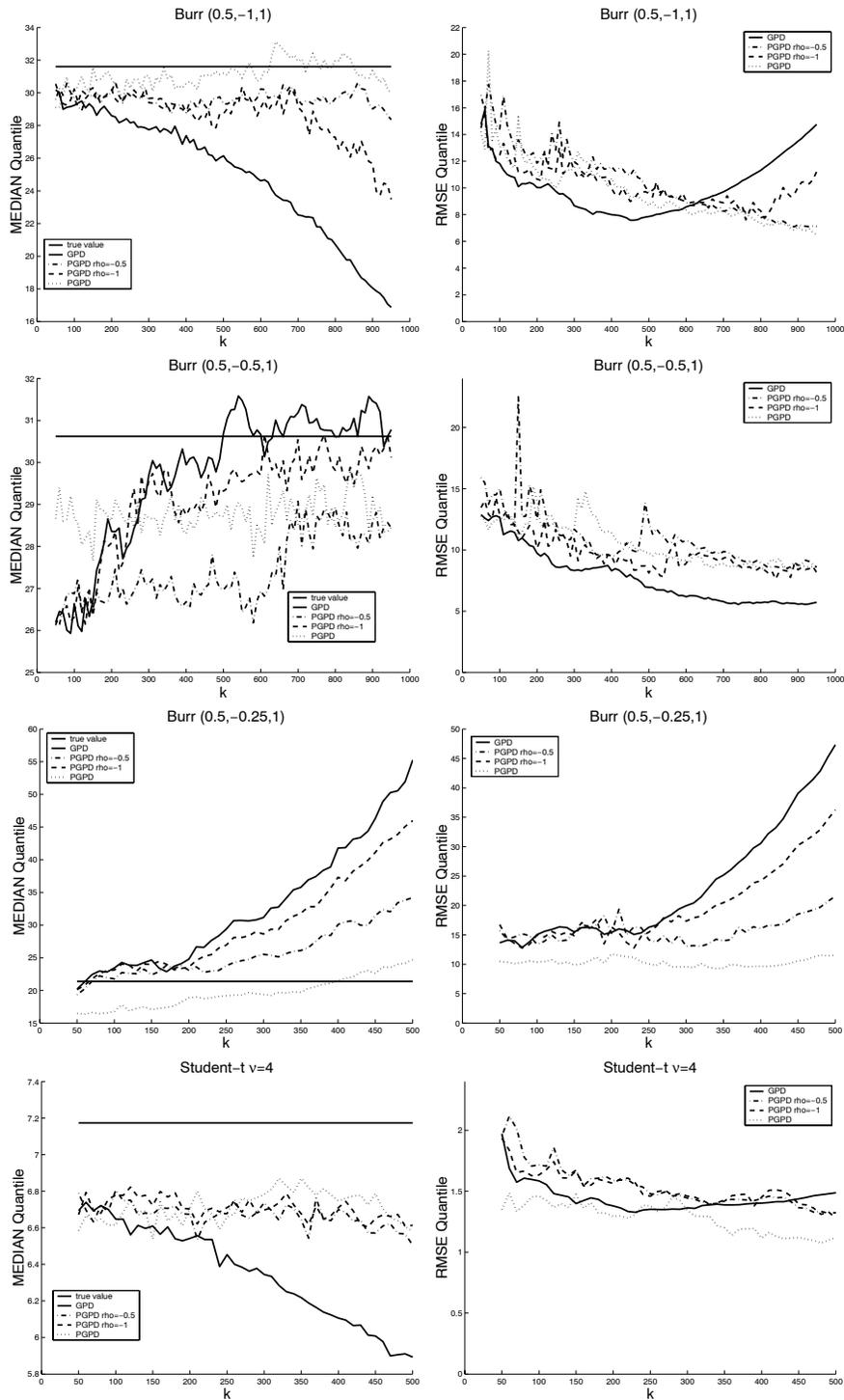
Figure 7: Medians and RMSEs of $\hat{x}_p$ as a function of k, based on 100 samples of size 1000 from a Burr(0.5,-1,1), Burr(0.5,-0.5,1), Burr(0.5,-0.25,1) and Student-t ($\nu = 4$) distribution.
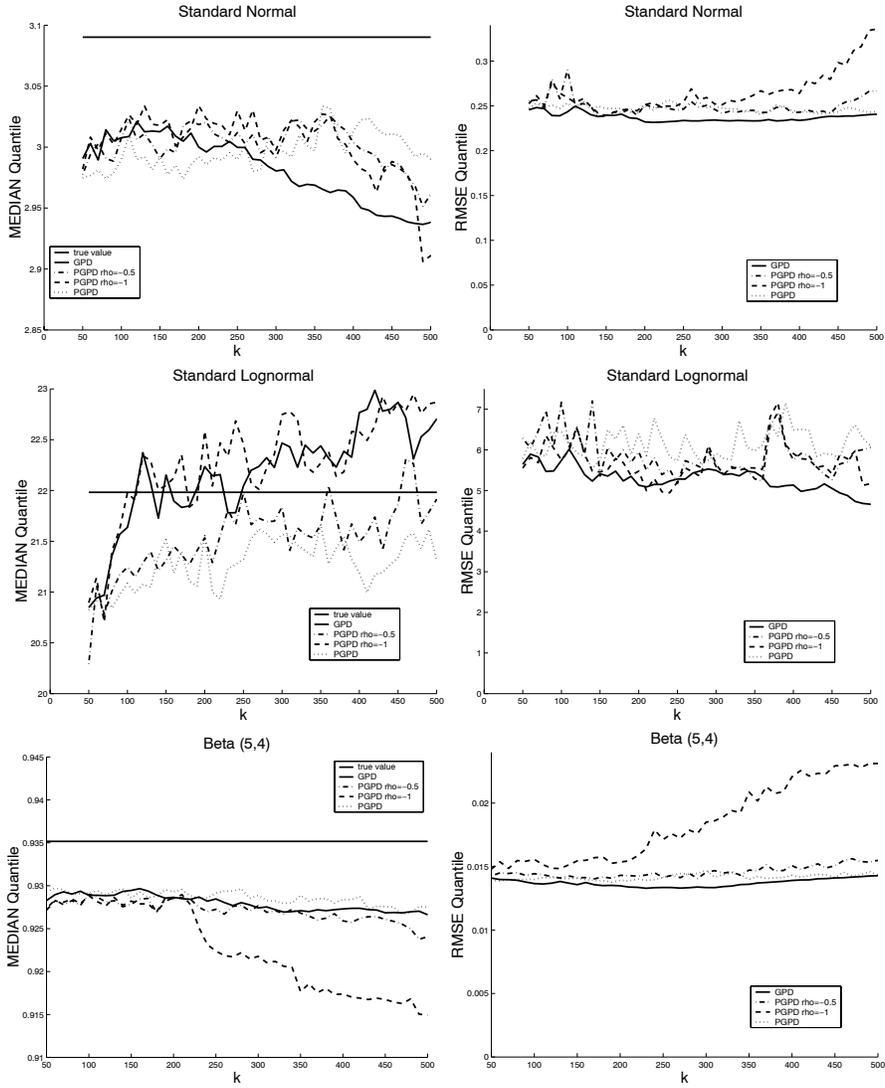
Figure 8: Medians and RMSEs of $\hat{x}_p$ as function of k, based on 100 samples of size 1000 from standard normal, standard lognormal and Beta(5,4) distribution.
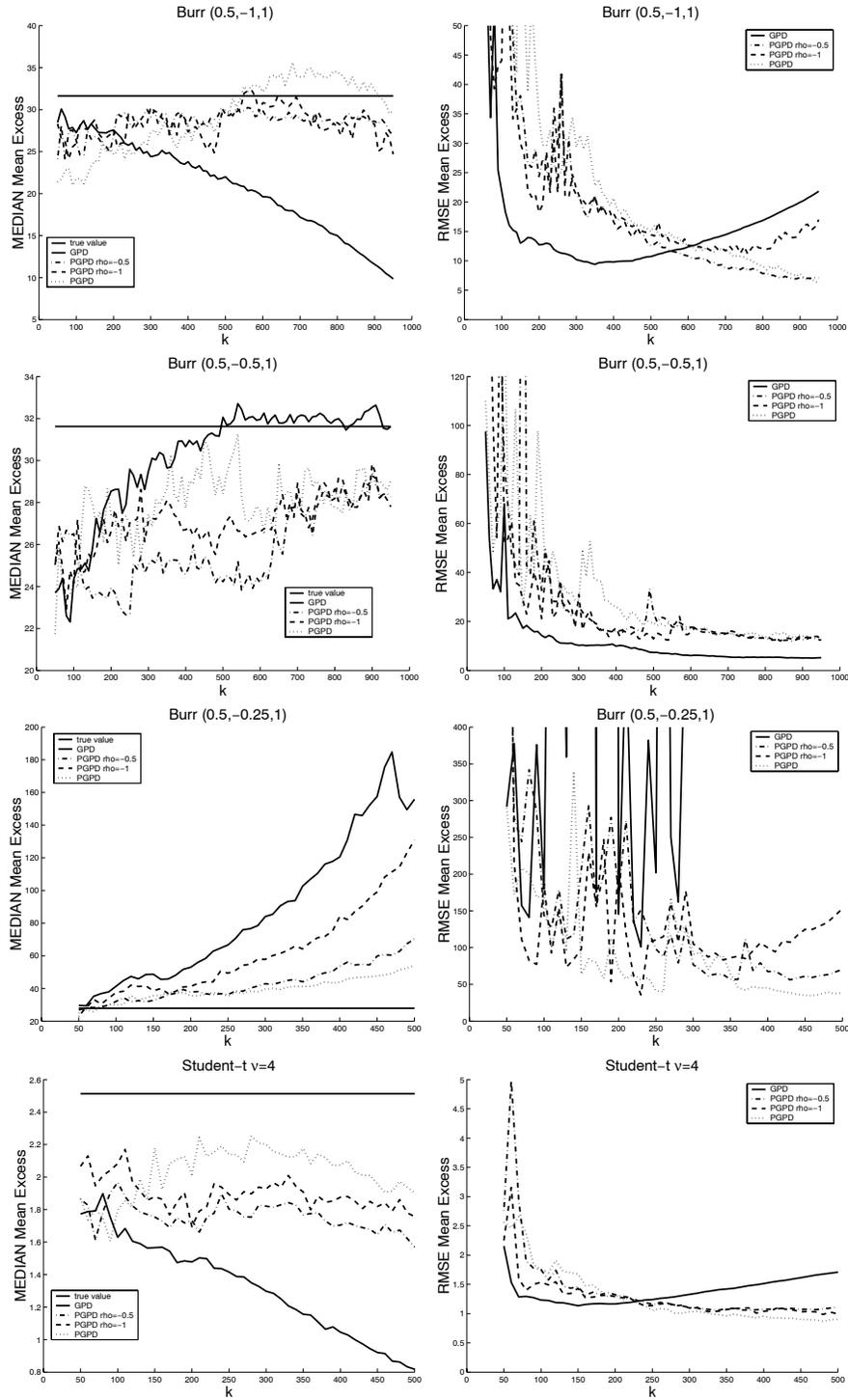
Figure 9: Medians and RMSEs of $\hat{m}_R$ as a function of k, based on 100 samples of size 1000 from a Burr(0.5,-1,1), Burr(0.5,-0.5,1), Burr(0.5,-0.25,1) and Student-t ($\nu = 4$) distribution.
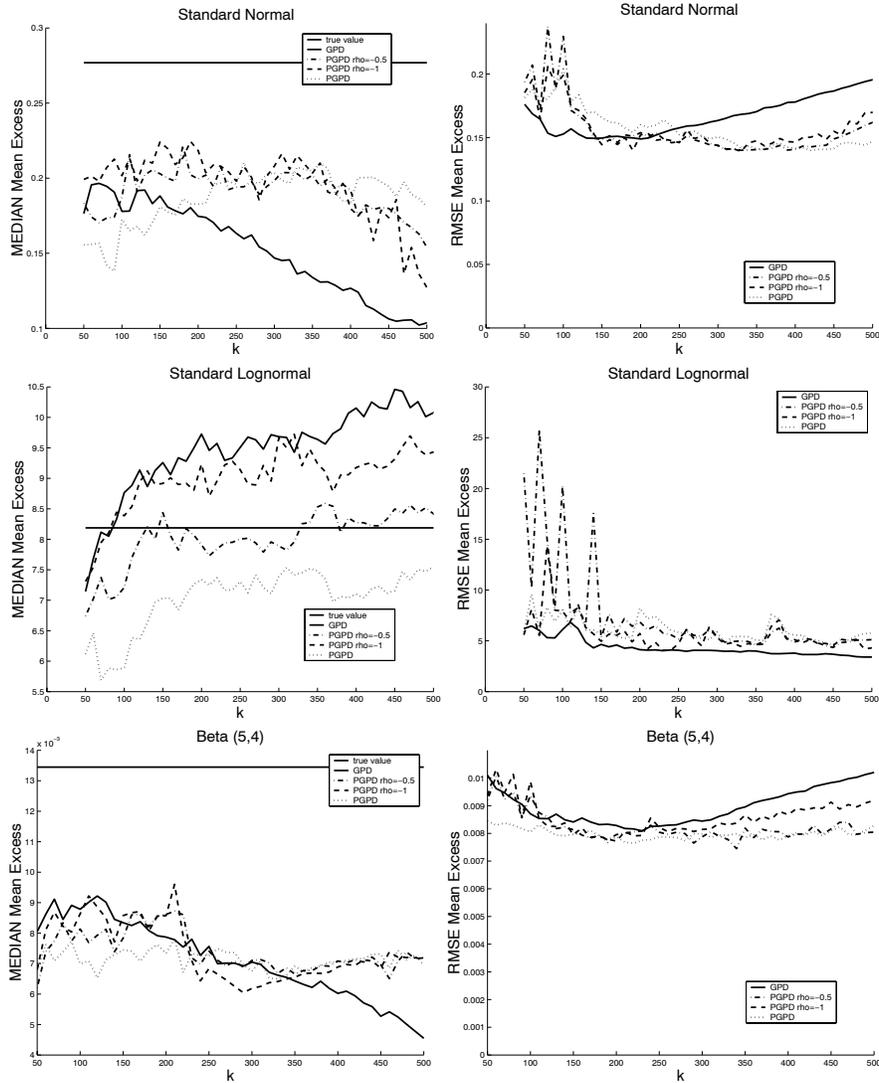
Figure 10: Medians and RMSEs of $\hat{m}_R$ as function of k, based on 100 samples of size 1000 from standard normal, standard lognormal and Beta(5,4) distribution.
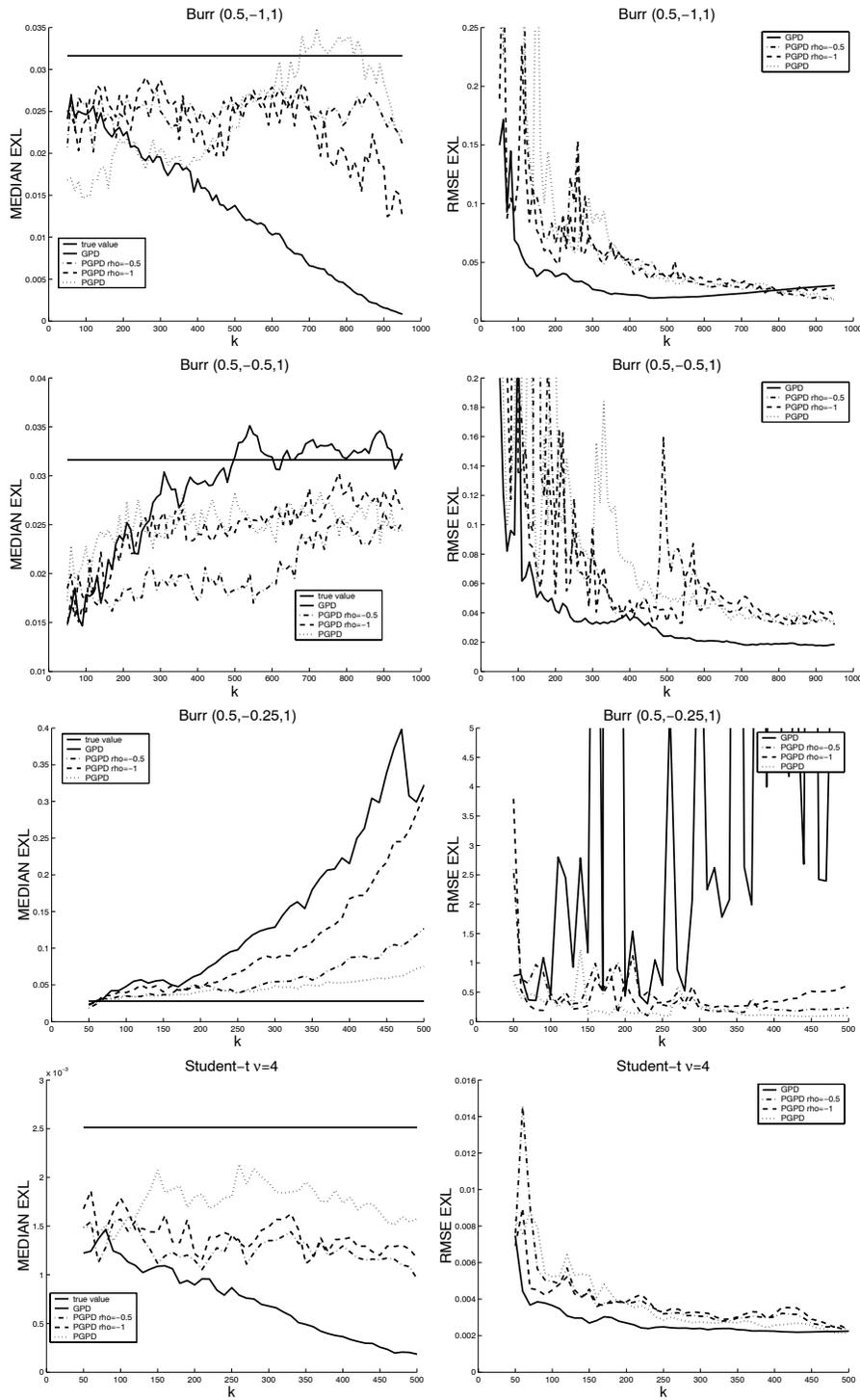
Figure 11: Medians and RMSEs of $\widehat{EXL}_R$ as a function of k, based on 100 samples of size 1000 from a Burr(0.5,-1,1), Burr(0.5,-0.5,1), Burr(0.5,-0.25,1) and Student-t ($\nu = 4$) distribution.
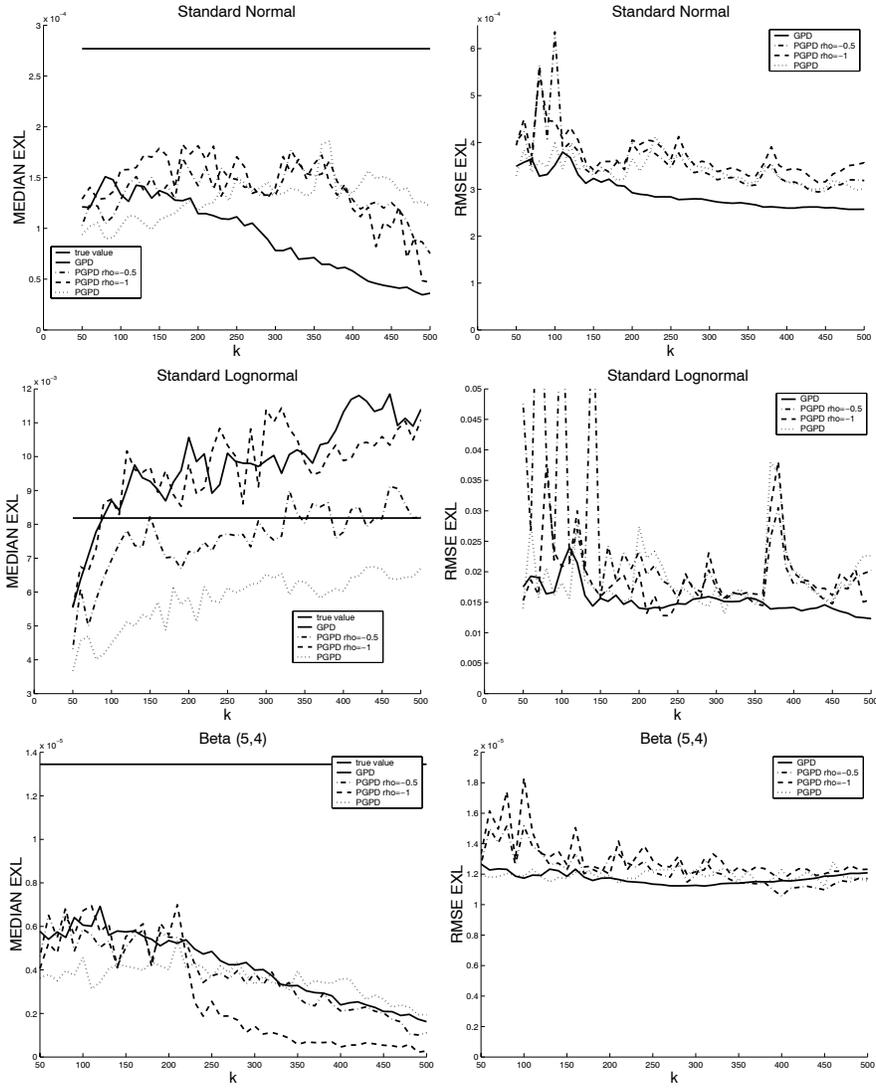
Figure 12: Medians and RMSEs of $\widehat{EXL}_R$ as function of k, based on 100 samples of size 1000 from standard normal, standard lognormal and Beta(5,4) distribution.