Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view

Article *in* Insurance Mathematics and Economics · December 2009

DOI: 10.1016/j.insmatheco.2009.09.002 · Source: RePEc

Robust and Efficient Fitting of the Generalized Pareto Distribution with Actuarial Applications in View

Vytaras Brazauskas[1]
University of Wisconsin-Milwaukee

Andreas Kleefeld[2]
University of Wisconsin-Milwaukee

Abstract

Due to advances in extreme value theory, the generalized Pareto distribution (GPD) emerged as a natural family for modeling exceedances over a high threshold. Its importance in applications (e.g., insurance, finance, economics, engineering and numerous other fields) can hardly be overstated and is widely documented. However, despite the sound theoretical basis and wide applicability, fitting of this distribution in practice is not a trivial exercise. Traditional methods such as maximum likelihood and method-of-moments are undefined in some regions of the parameter space. Alternative approaches exist but they lack either robustness (e.g., probability-weighted moments) or efficiency (e.g., method-of-medians), or present significant numerical problems (e.g., minimum-divergence procedures). In this article, we propose a computationally tractable method for fitting the GPD, which is applicable for all parameter values and offers competitive trade-offs between robustness and efficiency. The method is based on 'trimmed moments'. Large-sample properties of the new estimators are provided, and their small-sample behavior under several scenarios of data contamination is investigated through simulations. We also study the effect of our methodology on actuarial applications. In particular, using the new approach, we fit the GPD to the Danish insurance data and apply the fitted model to a few risk measurement and ratemaking exercises.

JEL Codes: C13, C14, C16, C46. *Subject Categories*: IM10, IM11, IM41, IM54.
Insurance Branch Categories: IB30.

Keywords: Pure Premium; Robust Statistics; Simulations; Trimmed Moments; Value-at-Risk.

[1]Corresponding author: Vytaras Brazauskas is Associate Professor, Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, U.S.A. E-mail address: `vytaras@uwm.edu`

[2]Andreas Kleefeld is Ph.D. Candidate, Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, U.S.A. E-mail address: `kleefeld@uwm.edu`

1 Introduction

Due to advances in extreme value theory, the generalized Pareto distribution (GPD) emerged as a natural family for modeling exceedances over a high threshold. Its importance in applications such as insurance, finance, economics, engineering and numerous other fields, can hardly be overstated and is widely documented. For a general introduction to this distribution, the reader can be referred to, for example, Johnson *et al*. (1994, Chapter 20) and McNeil *et al*. (2005, Chapter 7).

There is also a substantial literature on various specialized topics involving the GPD. For example, the problem of parameter estimation has been addressed by Hosking and Wallis (1987), Castillo and Hadi (1997), Peng and Welsh (2001), and Juárez and Schucany (2004), among others. The work of Davison and Smith (1990) presents extensions of the GPD to data with covariates and time series models. It also contains insightful discussions about the validity of the GPD assumption in real-world applications. An interesting approach toward threshold selection has been proposed by Dupuis (1999); it is based on robust procedures. Finally, the papers by McNeil (1997) and Cebrian *et al*. (2003) are excellent illustrations of the GPD's role in actuarial applications.

However, despite the sound theoretical basis and wide applicability, fitting of this distribution in practice is not a trivial exercise. Traditional methods such as maximum likelihood and method-of-moments are undefined in some regions of the parameter space. Alternative approaches exist and include estimators that are based on: probability-weighted moments (Hosking and Wallis, 1987), method-of-medians (Peng and Welsh, 2001), quantiles (Castillo and Hadi, 1997), and minimum-divergence procedures (Juárez and Schucany, 2004). While each methodology has its own advantages, it should also be pointed out some of their caveats. For example, the probability-weighted estimators lack robustness (see the influence curve plots of Davison and Smith, 1990), the method-of-medians estimators are not very efficient (see He and Fung, 1999, who introduced this method of parameter estimation), and the minimum-divergence estimators can present significant computational problems (see the simulation studies of Juárez and Schucany, 2004).

In this article, we propose a computationally tractable method for fitting the GPD, which is applicable for all parameter values and, in addition, offers competitive trade-offs between robustness and efficiency. The method is based on 'trimmed moments', abbreviated as MTM, and the resulting estimators are found by following a general methodology introduced by Brazauskas *et al*. (2009). Note that the idea of coupling robust-statistics methods and extreme-value data is not a contradiction. This has been argued by some of the aforementioned authors, but perhaps the most eloquent discussion on this topic has been provided by Dell'Aquila and Embrechts (2006). Interestingly, the estimators of Castillo and Hadi (1997) also possess similar qualities as MTMs, though they have not been presented or viewed from the robustness and efficiency perspective. As we will show later, the quantile-based estimators are a special/limiting case of MTMs, which occurs when one trims all the available data except for a few observations, i.e., the chosen quantiles. It is also worthwhile mentioning that in the actuarial literature, the quantile-based estimators are better known as estimators that are derived by employing the *percentile-matching* approach (see, Klugman *et al*., 2004, Section 12.1).

The rest of the article is organized as follows. In Section 2, we first provide key distributional properties of the GPD and graphically examine shape changes of its density function. Later, we study various issues related to model-fitting. Specifically, a number of methods for the estimation of parameters (e.g., method of maximum likelihood, method of moments, percentile-matching method, and method of trimmed moments) are presented, and large- and small-sample robustness properties of these estimators are investigated in detail. In Section 3, we fit the GPD to the Danish insurance data. The fitted models are then employed in a few ratemaking and quantitative risk management examples. In particular, point estimates for several value-at-risk measures and net premiums are calculated. Results are summarized and conclusions are drawn in Section 4.

2 Fitting the GPD

In this section, we study essential issues related to model-fitting. The key facts and formulas of the GPD are presented, illustrated and discussed in subsection 2.1. A number of existing and new methods for estimation of the GPD parameters are provided in subsection 2.2. Finally, subsection 2.3 is devoted to small-sample properties of the (theoretically and computationally) most favorable estimators, which are investigated using simulations.

2.1 The Model

The cumulative distribution function (cdf) of the GPD is given by

$$F(x) = \begin{cases} 1 - (1 - \gamma(x - x_0)/\sigma)^{1/\gamma}, & \gamma \neq 0 \\ 1 - \exp\left(-(x - x_0)/\sigma\right), & \gamma = 0, \end{cases} \tag{2.1}$$

and the probability density function (pdf) by

$$f(x) = \begin{cases} \sigma^{-1} (1 - \gamma(x - x_0)/\sigma)^{1/\gamma - 1}, & \gamma \neq 0 \\ \sigma^{-1} \exp\left(-(x - x_0)/\sigma\right), & \gamma = 0, \end{cases} \tag{2.2}$$

where the pdf is positive for $x \geq x_0$, when $\gamma \leq 0$, or for $x_0 \leq x \leq \sigma/\gamma$, when $\gamma > 0$. The parameters $-\infty < x_0 < \infty$, $\sigma > 0$, and $-\infty < \gamma < \infty$ control the location, scale, and shape of the distribution, respectively. In insurance applications, the location parameter x_0 is typically known and can be interpreted as a deductible, retention level, or attachment point. But σ and γ are both unknown and have to be estimated from the data. Note that, when $\gamma = 0$ and $\gamma = 1$, the GPD reduces to the exponential distribution (with location x_0 and scale σ) and the uniform distribution on $[x_0, \sigma]$, respectively. If $\gamma < 0$, then the Pareto distributions are obtained. In Figure 1, we illustrate shape changes of $f(x)$ for different choices of the shape parameter γ.

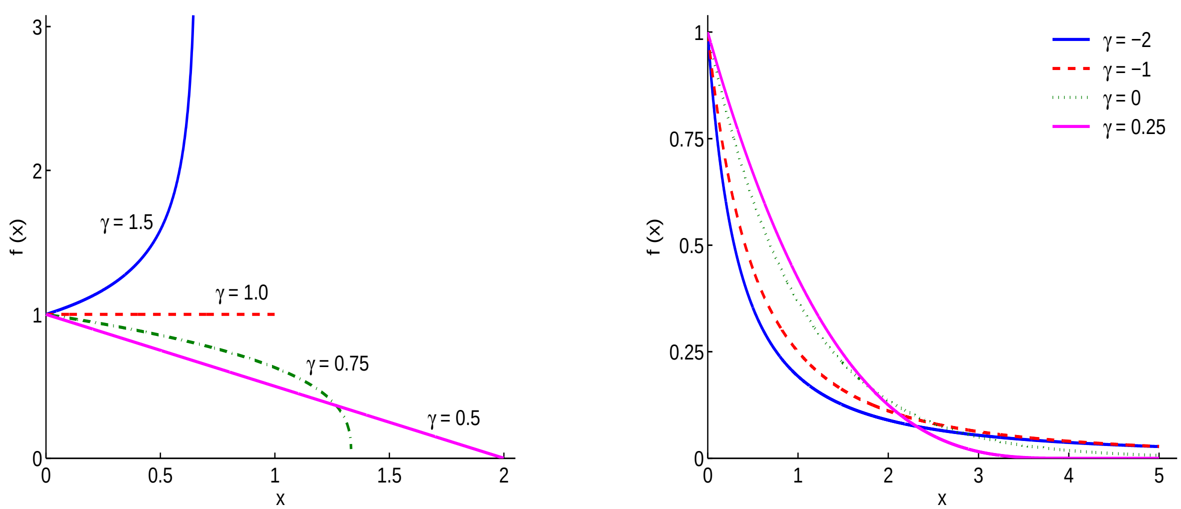

FIGURE 1: Probability density functions of the GPD for $x_0 = 0$, $\sigma = 1$, and various values of γ.

Further, besides functional simplicity of its cdf and pdf, another attractive feature of the GPD is that its quantile function (qf) has an explicit formula. This is especially useful for estimation of the parameters via MTM and percentile-matching methods, and for portfolio risk evaluations by using value-at-risk measures. Specifically, the qf is given by

$$F^{-1}(u) = \begin{cases} x_0 + (\sigma/\gamma)\left(1 - (1-u)^{\gamma}\right), & \gamma \neq 0 \\ x_0 - \sigma \log(1-u), & \gamma = 0. \end{cases} \tag{2.3}$$

Also, the mean and variance of the GPD are given by

$$\mathbf{E}[X] = x_0 + \frac{\sigma}{1+\gamma}, \quad \gamma > -1 \tag{2.4}$$

$$\mathbf{Var}[X] = \frac{\sigma^2}{(2\gamma+1)(\gamma+1)^2}, \quad \gamma > -1/2. \tag{2.5}$$

Finally, as discussed by Hosking and Wallis (1987), GPDs with $\gamma > 1/2$ have finite endpoints with $f(x) > 0$ at each endpoint (see Figure 1), and such shapes rarely occur in statistical applications. GPDs with $\gamma \leq -1/2$ have infinite variance (see expression (2.5)), and this too is unusual in statistical applications. We certainly agree with this assessment of the statistical applications but in actuarial work things are slightly different. For example, when pricing an insurance layer, it is not unreasonable to employ a probability distribution with both endpoints finite. What is even more common in actuarial applications is the heavy-tailed distributions, with no finite moments at all, that often appear in modeling of catastrophic claims. This discussion, therefore, suggests that parameter-estimation methods that work over the entire parameter space of the GPD are indeed needed in actuarial applications.

2.2 Parameter Estimation

In subsection 2.2.1, standard estimators, based on the maximum likelihood and method-of-moments approaches, are presented and their large-sample properties are examined. Then, in subsection 2.2.2, we briefly review quantile-type (percentile-matching) estimators and specify their large-sample distribution. Finally, in subsection 2.2.3, we consider a recently introduced robust estimation technique, the method of trimmed moments, and study asymptotic behavior of estimators based on it.

Also, throughout this section, we will consider a sample of n independent and identically distributed random variables, $X_1, \ldots, X_n$, from a GPD family with its cdf, pdf, and qf given by (2.1), (2.2) and (2.3), respectively, and denote $X_{1:n} \leq \cdots \leq X_{n:n}$ the order statistics of $X_1, \ldots, X_n$.

2.2.1 Standard Methods

A maximum likelihood estimator of (σ, γ) of the GPD, denoted by $(\widehat{\sigma}_{\text{MLE}}, \widehat{\gamma}_{\text{MLE}})$, is found by numerically maximizing the log-likelihood function:

$$\log \mathcal{L}(\sigma, \gamma \,|\, X_1, \ldots, X_n) = -n \log \sigma + \frac{1-\gamma}{\gamma} \sum_{i=1}^{n} \log \Big(1 - \frac{\gamma}{\sigma}(X_i - x_0) \Big) .$$

Note that this function can be made arbitrarily large by choosing $\gamma > 1$ and σ/γ close to the maximum of X_i's, i.e., close to $X_{n:n}$. Hence, the obtained result will be a *local* maximum, not global. Further, as presented by Hosking and Wallis (1987), the estimator $(\widehat{\sigma}_{\text{MLE}}, \widehat{\gamma}_{\text{MLE}})$ is consistent, asymptotically normal, and asymptotically efficient when $\gamma < 1/2$. Specifically, we shall write that, provided $\gamma < 1/2$,

$$\big(\widehat{\sigma}_{\text{MLE}}, \widehat{\gamma}_{\text{MLE}}\big) \;\sim\; \mathcal{AN}\big((\sigma,\gamma),\, n^{-1}\mathbf{\Sigma_0}\big) \quad \text{with} \quad \mathbf{\Sigma_0} = (1-\gamma) \begin{bmatrix} 2\sigma^2 & \sigma \\ \sigma & (1-\gamma) \end{bmatrix}, \tag{2.6}$$

where $\mathcal{AN}$ stands for 'asymptotically normal'. On the other hand, if $\gamma \geq 1/2$, then we have the nonregular situation. The latter case has been treated by Smith (1985) who demonstrated that: for $\gamma = 1/2$, MLE's asymptotic distribution is still normal but with a different convergence rate; for $1/2 < \gamma \leq 1$, it is not normal (and very messy); and for $\gamma > 1$, the MLE does not exist. We will not consider the nonregular case in this paper.

Next, a method-of-moments estimator of (σ, γ) can be found by matching the GPD mean and variance, given by (2.4) and (2.5), with the sample mean $\overline{X}$ and variance S^2, and then solving the system of equations with respect to σ and γ. This leads to

$$\widehat{\sigma}_{\text{MM}} = 0.5 \left(\overline{X} - x_0\right) \left(\left(\overline{X} - x_0\right)^2 / S^2 + 1 \right) \quad \text{and} \quad \widehat{\gamma}_{\text{MM}} = 0.5 \left(\left(\overline{X} - x_0\right)^2 / S^2 - 1 \right) .$$

If $\gamma > -1/4$, then $\left(\widehat{\sigma}_{\text{MM}}, \widehat{\gamma}_{\text{MM}}\right) \;\sim\; \mathcal{AN}\big((\sigma,\gamma),\, n^{-1}\mathbf{\Sigma_1}\big)$ with

$$\mathbf{\Sigma_1} = \frac{(1+\gamma)^2}{(1+3\gamma)(1+4\gamma)} \begin{bmatrix} 2\sigma^2(1+6\gamma+12\gamma^2)/(1+2\gamma) & \sigma(1+4\gamma+12\gamma^2) \\ \sigma(1+4\gamma+12\gamma^2) & (1+2\gamma)(1+\gamma+6\gamma^2) \end{bmatrix}. \tag{2.7}$$

In the special case $\gamma = 0$, the MLE and MM are the same, which implies that the asymptotic normality results (2.6) and (2.7) are identical. Hence, the MM estimator is asymptotically fully efficient. In general, the asymptotic relative efficiency (ARE) of one estimator with respect to another is defined as the ratio of their asymptotic variances. In the multivariate case, the two variances are replaced by the corresponding generalized variances, which are the determinants of the asymptotic variance-covariance matrices of the k-variate parameter estimators, and then the ratio is raised to the power $1/k$. (For further details on these issues, we refer, for example, to Serfling, 1980, Section 4.1). Thus, for the MLE and MM estimators of the GPD parameters, we have that

$$\text{ARE(MM, MLE)} = \left(|\boldsymbol{\Sigma_0}|/|\boldsymbol{\Sigma_1}|\right)^{1/2} = \frac{(1-\gamma)(1+3\gamma)}{(1+\gamma)^2}\sqrt{\frac{(1-2\gamma)(1+4\gamma)}{1+2\gamma}}, \tag{2.8}$$

where $|\boldsymbol{\Sigma_i}|$ denotes the determinant of matrix $\boldsymbol{\Sigma_i}$, $i = 0, 1$. In Table 1, we provide numerical illustrations of expression (2.8) for selected values of the shape parameter γ. Note that while the MLE dominates MM for $\gamma < 1/2$, the MM continues to maintain the same (simple) asymptotic behavior for $\gamma \geq 1/2$, and hence it is a more practical method than the MLE in this range of γ.

Table 1: ARE(MM, MLE) for selected values of $\gamma < 0.50$.

γ	$(-\infty; -0.25]$	-0.20	-0.10	-0.05	0	0.05	0.10	0.20	0.30	0.40	0.49
ARE	0	.512	.902	.978	1	.982	.934	.781	.584	.362	.098

Finally, notice that the definition of the MLE and MM estimators as well as the asymptotic results (2.6)–(2.8) are not valid over the entire parameter space. This prompts us to search for parameter-estimation methods that work everywhere, i.e., for $\sigma > 0$ and $-\infty < \gamma < \infty$.

2.2.2 Percentile-Matching

The quantile, or percentile-matching, estimators of the GPD parameters have been proposed by Castillo and Hadi (1997). In this section, we shall provide the computational formulas of these estimators and specify their large-sample distribution.

A percentile-matching estimator of (σ, γ) of the GPD is found by matching two theoretical quantiles $F^{-1}(p_1)$ and $F^{-1}(p_2)$ with the corresponding empirical quantiles $X_{n:[np_1]}$ and $X_{n:[np_2]}$, and then solving the resulting system of equations with respect to σ and γ. Here the percentile levels p_1 and p_2 $(0 < p_1 < p_2 < 1)$ are chosen by the researcher, the quantile function $F^{-1}(\cdot)$ is given by (2.3), and $[\cdot]$ denotes 'greatest integer part'. This leads to

$$\widehat{\sigma}_{\text{PM}} = \begin{cases} \widehat{\gamma}_{\text{PM}}(X_{n:[np_1]} - x_0)/(1-(1-p_1)^{\widehat{\gamma}_{\text{PM}}}), & \text{if } \widehat{\gamma}_{\text{PM}} \neq 0, \\ -(X_{n:[np_1]} - x_0)/\log(1-p_1), & \text{if } \widehat{\gamma}_{\text{PM}} = 0, \end{cases}$$

where $\widehat{\gamma}_{\mathrm{PM}}$ is found by numerically solving the following equation with respect to γ

$$(X_{n:[np_2]} - x_0)\big/(X_{n:[np_1]} - x_0) - (F^{-1}(p_2) - x_0)\big/(F^{-1}(p_1) - x_0) = 0.$$

Note that the ratio $(F^{-1}(p_2) - x_0)\big/(F^{-1}(p_1) - x_0)$ depends only on γ and is a *continuous* function.

As discussed by Castillo and Hadi (1997, Section 3), the percentile-matching estimators are consistent and asymptotically normal. More specifically,

$$\big(\widehat{\sigma}_{\mathrm{PM}}, \widehat{\gamma}_{\mathrm{PM}}\big) \sim \mathcal{AN}\big((\sigma, \gamma), n^{-1}\mathbf{\Sigma_2}\big) \quad \text{with} \quad \mathbf{\Sigma_2} = \mathbf{C}\mathbf{\Sigma}_*\mathbf{C}', \tag{2.9}$$

where

$$\mathbf{\Sigma}_* = \sigma^2 \begin{bmatrix} p_1(1-p_1)^{2\gamma-1} & p_1(1-p_1)^{\gamma-1}(1-p_2)^{\gamma} \\ p_1(1-p_1)^{\gamma-1}(1-p_2)^{\gamma} & p_2(1-p_2)^{2\gamma-1} \end{bmatrix}$$

and

$$\mathbf{C} = \frac{1}{c_*(p_1, p_2, \gamma)} \begin{bmatrix} -F^{-1}(p_2) - \sigma(1-p_2)^{\gamma}\log(1-p_2) & F^{-1}(p_1) + \sigma(1-p_1)^{\gamma}\log(1-p_1) \\ (1-p_2)^{\gamma} - 1 & 1 - (1-p_1)^{\gamma} \end{bmatrix}$$

with $c_*(p_1, p_2, \gamma) = F^{-1}(p_2)(1-p_1)^{\gamma}\log(1-p_1) - F^{-1}(p_1)(1-p_2)^{\gamma}\log(1-p_2)$.

Remark 1: *The* $\gamma = 0$ *case.*

The asymptotic normality result (2.9) is valid for $-\infty < \gamma < \infty$. When $\gamma = 0$, however, the entries of $\mathbf{C}$ have to be calculated by taking the limit $\gamma \to 0$. (The elements of $\mathbf{\Sigma}_*$ are computed directly by substituting $\gamma = 0$.) This leads to the following simplified formula of $\mathbf{C}$:

$$\mathbf{C} = \frac{1}{\log(1-p_1)\log(1-p_2)\log\big((1-p_2)/(1-p_1)\big)} \begin{bmatrix} -\log^2(1-p_2) & \log^2(1-p_1) \\ 2\log(1-p_2)/\sigma & -2\log(1-p_1)/\sigma \end{bmatrix}.$$

□

2.2.3 MTM Estimation

The underlying idea of the MTM—method of trimmed moments—is identical to that of the percentile-matching, except now we match theoretical and empirical *trimmed moments* instead of percentiles (quantiles). A general methodology for finding MTM estimators has been introduced and developed by Brazauskas *et al.* (2009). Here we adapt their proposal to the GPD case.

We first calculate two sample trimmed moments (in this case, trimmed means):

$$\widehat{\mu}_j = \frac{1}{n - m_n(j) - m_n^*(j)} \sum_{i=m_n(j)+1}^{n-m_n^*(j)} \big(X_{i:n} - x_0\big), \quad \text{for} \;\; j = 1, 2, \tag{2.10}$$

where $m_n(j)$ and $m_n^*(j)$ are integers such that $0 \le m_n(j) < n - m_n^*(j) \le n$, and $m_n(j)/n \to a_j$ and $m_n^*(j)/n \to b_j$ as $n \to \infty$. The trimming proportions a_j and b_j must be chosen by the researcher. If one selects $a_j > 0$ and $b_j > 0$ $(0 < a_j + b_j < 1)$, for $j = 1, 2$, then the resulting estimators will be resistant against outliers, i.e., they will be robust with the lower and upper breakdown points given by $\text{LBP} = \min\{a_1, a_2\}$ and $\text{UBP} = \min\{b_1, b_2\}$, respectively. The robustness of such estimators against extremely small or large outliers comes from the fact that the order statistics with the index less than $n \times \text{LBP}$ or more than $n \times (1 - \text{UBP})$ are simply not included in the computation of estimates.

Next, we derive the corresponding theoretical trimmed moments:

$$\begin{aligned} \mu_j &= \frac{1}{1 - a_j - b_j} \int_{a_j}^{1-b_j} \big(F^{-1}(u) - x_0\big)\, \mathrm{d}u \\ &= \sigma \times \begin{cases} -1 + \dfrac{1}{1 - a_j - b_j} \log\left(\dfrac{1 - a_j}{b_j}\right), & \text{if } \; \gamma = -1, \\[2ex] 1 + \dfrac{b_j \log(b_j) - (1 - a_j)\log(1 - a_j)}{1 - a_j - b_j}, & \text{if } \; \gamma = 0, \\[2ex] (1/\gamma)\left(1 - \dfrac{(1 - a_j)^{\gamma+1} - b_j^{\gamma+1}}{(\gamma + 1)(1 - a_j - b_j)}\right), & \text{otherwise.} \end{cases} \end{aligned} \tag{2.11}$$

Now, we match $\widehat{\mu}_j$, given by (2.10), with μ_j, given by (2.11), which results in a system of two equations. After some straightforward simplifications, we find that

$$\widehat{\sigma}_{\text{MTM}} = \widehat{\mu}_1 \times \begin{cases} -\left(1 - \dfrac{\log(1 - a_1) - \log(b_1)}{1 - a_1 - b_1}\right)^{-1}, & \text{if } \; \widehat{\gamma}_{\text{MTM}} = -1, \\[2ex] \left(1 - \dfrac{(1 - a_1)\log(1 - a_1) - b_1 \log(b_1)}{1 - a_1 - b_1}\right)^{-1}, & \text{if } \; \widehat{\gamma}_{\text{MTM}} = 0, \\[2ex] \widehat{\gamma}_{\text{MTM}} \left(1 - \dfrac{(1 - a_1)^{\widehat{\gamma}_{\text{MTM}}+1} - b_1^{\widehat{\gamma}_{\text{MTM}}+1}}{(\widehat{\gamma}_{\text{MTM}} + 1)(1 - a_1 - b_1)}\right)^{-1}, & \text{otherwise,} \end{cases}$$

where $\widehat{\gamma}_{\text{MTM}}$ is found by numerically solving the following equation with respect to γ

$$\widehat{\mu}_1/\widehat{\mu}_2 - \mu_1/\mu_2 = 0.$$

Note that the ratio μ_1/μ_2 depends only on γ and is a continuous function. Also, the estimators $\widehat{\sigma}_{\text{MTM}}$ and $\widehat{\gamma}_{\text{MTM}}$ are functions of $\widehat{\mu}_1$ and $\widehat{\mu}_2$, which we will denote $g_1(\widehat{\mu}_1, \widehat{\mu}_2)$ and $g_2(\widehat{\mu}_1, \widehat{\mu}_2)$, respectively.

Further, as demonstrated by Brazauskas *et al.* (2009), the MTM estimators are consistent and asymptotically normal. In particular, for the GPD case, we have that

$$\big(\widehat{\sigma}_{\mathrm{MTM}},\,\widehat{\gamma}_{\mathrm{MTM}}\big)\ \sim\ \mathcal{AN}\big((\sigma,\gamma),\,n^{-1}\mathbf{\Sigma_3}\big)\quad\text{ with }\quad \mathbf{\Sigma_3}=\mathbf{D}\mathbf{\Sigma}_{**}\mathbf{D}', \tag{2.12}$$

where $\mathbf{\Sigma}_{**} := \big[\sigma_{ij}^2\big]_{i,j=1}^2$ with the entries

$$\sigma_{ij}^2\ =\ \frac{1}{(1-a_i-b_i)(1-a_j-b_j)}\int_{a_i}^{1-b_i}\int_{a_j}^{1-b_j}\big(\min\{u,v\}-uv\big)\,\mathrm{d}F^{-1}(v)\,\mathrm{d}F^{-1}(u), \tag{2.13}$$

and $\mathbf{D}=[d_{ij}]_{i,j=1}^2$ with the entries $d_{ij}=\partial g_i/\partial\widehat{\mu}_j\big|_{(\mu_1,\mu_2)}$:

$$d_{11}=\sigma(d_*-\mu_1'\mu_2)/(\mu_1 d_*),\qquad d_{12}=\sigma\mu_1'/d_*,\qquad d_{21}=\mu_2/d_*,\qquad d_{22}=-\mu_1/d_*.$$

Here $d_*=\mu_1'\mu_2-\mu_2'\mu_1$, the trimmed moment μ_j is given by (2.11), and μ_j' is the derivative of μ_j with respect to γ, which is provided in the appendix. Of course, the entries σ_{ij}^2 can be derived analytically by performing straightforward (but tedious!) integration of (2.13), where F^{-1} is given by (2.3). We, however, used the bivariate trapezoidal rule to approximate the double integral in (2.13), and found it to be a much more effective approach than the theoretical one. It also is sufficiently accurate for all practical purposes. The details of this approximation are presented in the appendix.

Remark 2: *MTM and PM estimators.*

Suppose that the percentile p_j used in the PM estimation is between the a_j and $1-b_j$ of the MTM approach, i.e., $a_j<p_j<1-b_j$. Let $a_j\uparrow p_j$ and $1-b_j\downarrow p_j$. Then, it is easy to see that $\mu_j\to F^{-1}(p_j)-x_0$ and $\widehat{\mu}_j\to X_{n:[np_j]}-x_0$. Hence, the MTM estimators can be reduced to the PMs. Similar computations involving the matrices in (2.12) show that $\mathbf{D}\to\mathbf{C}$, $\mathbf{\Sigma}_{**}\to\mathbf{\Sigma}_*$, and thus $\mathbf{\Sigma_3}\to\mathbf{\Sigma_2}$ as a_j and $1-b_j$ approach p_j. In summary, for estimation of the GPD parameters, the PM approach is a limiting case of the MTM. Its robustness properties also directly follow from those of the MTM. That is, the lower and upper breakdown points of the PM estimator are: $\textsc{lbp}=p_1$ and $\textsc{ubp}=1-p_2$. $\square$

We complete this section with an investigation of the efficiency properties of the MTM and PM procedures with respect to the standard methods. Since the MLE is asymptotically normal (at the n^{-1} rate) and fully efficient for $\gamma<1/2$, we will use its generalized variance as a benchmark for that range of γ. For the case $\gamma\geq 1/2$, the MTM estimators will be compared to the MM. We will choose

PM and MTM estimators so that they will form pairs with respect to robustness properties. That is, for each PM estimator with percentile levels p_1 and p_2, we will choose an MTM estimator with $\text{LBP} = p_1$ and $\text{UBP} = 1 - p_2$. That is, the trimming proportions of the MTM estimator shall satisfy:

$$0 < a_1 < 1 - b_1 \leq a_2 < 1 - b_2 < 1 \quad \text{and} \quad (p_1, p_2) = (a_1, 1 - b_2).$$

The other two proportions, a_2 and $1 - b_1$, will be chosen to maximize the efficiency of the MTM estimator for most of the selected γ values. Note that there are other ways to arrange the proportions $(a_1, 1 - b_1)$ and $(a_2, 1 - b_2)$ while still maintaining $\text{LBP} = p_1$ and $\text{UBP} = 1 - p_2$. For example, we can choose: $0 < a_1 \leq a_2 < 1-b_2 \leq 1-b_1 < 1$ with $(p_1, p_2) = (a_1, 1-b_1)$, or $0 < a_1 \leq a_2 < 1-b_1 \leq 1-b_2 < 1$ with $(p_1, p_2) = (a_1, 1 - b_2)$. The advantage of the initial choice is that it directly yields PM estimators while the other two arrangements do not. The findings of our study are summarized in Table 2.

Table 2: ARE of MTM estimators with respect to: MLE (for $\gamma < 0.50$) and MM (for $\gamma \geq 0.50$). The PM estimators correspond to MTMs with $a_j \approx p_j \approx 1 - b_j$, $j = 1, 2$, and are marked with *.

Trimming Proportions		$\gamma < 0.50$							
$(a_1,\ 1-b_1)$	$(a_2,\ 1-b_2)$	−4	−2	−1	−0.40	−0.20	0	0.20	0.40
0.05, 0.30	0.70, 0.95	0.803	0.839	0.749	0.585	0.502	0.402	0.284	0.141
0.05*	0.95*	0.474	0.405	0.351	0.294	0.265	0.227	0.174	0.095
0.10, 0.30	0.60, 0.90	0.829	0.802	0.658	0.482	0.403	0.315	0.217	0.105
0.10*	0.90*	0.648	0.562	0.472	0.373	0.326	0.268	0.196	0.102
0.15, 0.35	0.80, 0.90	0.789	0.802	0.705	0.557	0.483	0.393	0.283	0.143
0.15*	0.90*	0.705	0.643	0.553	0.443	0.389	0.321	0.236	0.122
0.30, 0.50	0.70, 0.85	0.396	0.591	0.604	0.495	0.429	0.345	0.244	0.120
0.30*	0.85*	0.679	0.693	0.615	0.490	0.426	0.348	0.251	0.128
0.50, 0.60	0.70, 0.75	0.295	0.391	0.374	0.299	0.258	0.208	0.148	0.073
0.50*	0.75*	0.404	0.462	0.424	0.337	0.292	0.235	0.168	0.084

		$\gamma \geq 0.50$							
		0.50	0.75	1	1.50	2	2.50	3	4
0.05, 0.30	0.70, 0.95	0.412	0.495	0.614	0.981	1.607	2.674	4.501	13.125
0.05*	0.95*	0.295	0.420	0.634	1.643	4.832	15.569	53.534	718.650
0.10, 0.30	0.60, 0.90	0.302	0.351	0.419	0.617	0.922	1.385	2.088	4.783
0.10*	0.90*	0.305	0.397	0.540	1.088	2.384	5.572	13.704	92.599
0.15, 0.35	0.80, 0.90	0.425	0.530	0.687	1.221	2.277	4.394	8.712	36.429
0.15*	0.90*	0.369	0.481	0.656	1.328	2.919	6.834	16.822	113.740
0.30, 0.50	0.70, 0.85	0.349	0.413	0.497	0.729	1.055	1.503	2.111	4.069
0.30*	0.85*	0.380	0.478	0.625	1.132	2.162	4.301	8.854	40.752
0.50, 0.60	0.70, 0.75	0.215	0.259	0.321	0.508	0.813	1.301	2.079	5.261
0.50*	0.75*	0.247	0.302	0.380	0.626	1.057	1.810	3.133	9.644

Several conclusions emerge from the table. First, note that the PM estimator with $p_1 = 0.50$ and $p_2 = 0.75$ is the well-known Pickands' estimator which is highly robust but lacks efficiency. The ARE of PM estimators can be improved by choosing p_1 and p_2 further apart, and hence by sacrificing robustness. Second, for a practically relevant robustness properties, e.g., LBP ≤ 0.15 and UBP ≤ 0.10, efficiency of the PMs can be improved by an equally robust MTM, though the improvements are not uniform over all values of γ. Third, the least favorable range of γ for PMs and MTMs seems to be around the exponential distribution, i.e., for γ between -0.20 and 0.40. In that range of γ, their AREs can fall well below 0.50. Fourth, for $\gamma \geq 1.50$, however, the PMs and MTMs perform spectacularly with their ARE reaching even hundreds.

Remark 3: *A modification of the PMs.*

Castillo and Hadi (1997) also noticed that the PM estimators lack efficiency because they are based on only two data points. Thus the information contained in other observations is not utilized. These authors, therefore, called such estimators "initial estimates" and proposed to improve their efficiency properties by using the following algorithm. For a sample of size n, compute PM estimates of σ and γ for all possible pairs of percentile levels $p_1^{(i)} = i/n$ and $p_2^{(j)} = j/n$, $1 \leq i < j \leq n$. Such an approach produces $\binom{n}{2} = n(n-1)/2$ estimates of σ and γ. Then, the "final estimates" are:

$$\widetilde{\sigma} = \text{median}\left\{\widehat{\sigma}_1, \ldots, \widehat{\sigma}_{n(n-1)/2}\right\} \quad \text{and} \quad \widetilde{\gamma} = \text{median}\left\{\widehat{\gamma}_1, \ldots, \widehat{\gamma}_{n(n-1)/2}\right\}.$$

This modification indeed improves efficiency of the PMs, which Castillo and Hadi (1997) successfully demonstrated using Monte Carlo simulations. The improvement, however, comes at the price of rather inflexible robustness properties of the estimators. Also, to specify their asymptotic distribution, one has to deal with very messy analytical derivations. We finally note that these final estimates belong to a broad class of 'generalized median' estimators which, for a single-parameter Pareto model, were extensively studied by Brazauskas and Serfling (2000). The asymptotic breakdown points of the generalized median procedures are: UBP $= 1 -$ LBP $\rightarrow 1 - 1/\sqrt{2} \approx 0.293$ as $n \rightarrow \infty$. $\square$

2.3 Simulations

Here we supplement the theoretical large-sample results of Section 2.2 with finite-sample investigations. The objective of simulations is two-fold:

(a) to see how large the sample size n should be for the MLE, MM, PM, and MTM estimators to achieve (asymptotic) unbiasedness and for their finite-sample relative efficiency (RE) to reach the corresponding ARE level, and

(b) to see how robustness or non-robustness of an estimator manifests itself in computations.

To make the calculations of MM, MLE, and AREs (with respect to the MLE) possible, we will confine our study to the values of γ in the range $(-0.25; 0.50)$. The RE definition is similar to that of the ARE except that we now want to account for possible bias, which we do by replacing all entries in the variance-covariance matrix by the corresponding mean-squared errors. Also, for the objective (b), we will evaluate the bias and RE of estimators when the underlying GPD model is contaminated. In particular, we will employ the following ε-contamination neighborhoods:

$$F_{\varepsilon} = (1-\varepsilon)\, F_0 + \varepsilon\, G, \qquad (2.14)$$

where F_0 is the assumed GPD(x_0, σ, γ) model, G is a contaminating distribution which generates observations that violate the distributional assumptions, and the level of contamination ε represents the probability that a sample observation comes from the distribution G instead of F_0. Note that the choice $\varepsilon = 0$ results in a "clean" scenario which allows us to answer the questions raised in (a).

The study design is as follows: From a specified model (2.14) we generate 10,000 samples of size n using Monte Carlo. For each sample, we estimate the scale parameter σ and the shape parameter γ using the estimators of Section 2.2. Then we compute the average mean and RE of those 10,000 estimates. This process is repeated 10 times and the 10 average means and the 10 REs are again averaged and their standard deviations are reported. (Such repetitions are useful for assessing standard errors of the estimated means and REs. Hence, our findings are essentially based on 100,000 samples.) The standardized mean that we report is defined as the average of 100,000 estimates divided by the true value of the parameter that we are estimating. The standard error is standardized in a similar fashion. The study was performed for the following choices of simulation parameters:

- Parameters of F_0: $x_0 = 0$, $\sigma = 1$, and $\gamma = -0.20, -0.10, 0.15, 0.40$.
- Distribution G: GPD with $x_0 = 0$, $\sigma = 1$, and $\gamma = -5$.
- Level of contamination: $\varepsilon = 0$, 0.01, 0.05, 0.10, 0.15.
- Sample size: $n = 25$, 50, 100, 500.
- Estimators of (σ, γ):
  - MLE and MM.
  - PM with (p_1, p_2): $(0.05, 0.95)$, denoted PM1; $(0.15, 0.90)$, denoted PM2; $(0.30, 0.85)$, denoted PM3.
  - MTM with: $(a_1, b_1) = (0.05, 0.70)$ and $(a_2, b_2) = (0.70, 0.05)$, denoted MTM1; $(a_1, b_1) = (0.15, 0.65)$ and $(a_2, b_2) = (0.80, 0.10)$, denoted MTM2; $(a_1, b_1) = (0.30, 0.50)$ and $(a_2, b_2) = (0.70, 0.15)$, denoted MTM3.

Findings of the simulation study are summarized in Tables 3, 4 and 5. Note that the entries of the column $n \to \infty$ in Tables 3–4 are included as target quantities and follow from the theoretical results of Section 2.2, not from simulations.

Let us start with Tables 3 and 4 which relate to the objective (a). Several expected, as well as some surprising, conclusions emerge from these tables. First of all, we observe that it takes quite a large sample ($n = 500$) to get the bias of σ and γ within a reasonable range (e.g., within 5%) of the target. Among the estimators under consideration and for all choices of γ, the MTMs are best behaved. But for samples of size $n \leq 50$, the bias is indeed substantial for all estimators, especially for the shape parameter γ. Further, the MTMs' advantage with respect to the bias criterion is even more evident when we compare all estimators with respect to the RE (see Table 4). In many cases, MTMs nearly attain their ARE values for $n = 100$. What is interesting is that REs of the MLE and MM converge to the corresponding AREs *extremely* slowly when γ is near the theoretical boundaries. (Recall that the MLE's asymptotic normality result is valid for $\gamma < 1/2$, and the MM's for $\gamma > -1/4$.) To get a better idea about this issue, we performed additional simulations and discovered that, for $\gamma = -0.20$, RE of the MM is: 0.81 (for $n = 1,000$), 0.74 (for $n = 2,500$), 0.70 (for $n = 5,000$), 0.65 (for $n = 10,000$).

TABLE 3: Standardized mean of MLE, MM, PM and MTM estimators for selected values of γ. The entries are mean values based on 100,000 simulated samples of size n.

γ	*Estimator*	$n=25$		$n=50$		$n=100$		$n=500$		$n\to\infty$	
		σ	γ	σ	γ	σ	γ	σ	γ	σ	γ
−0.20	MLE	1.14	0.41	1.06	0.74	1.03	0.87	1.01	0.98	1	1
	MM	1.16	0.28	1.10	0.56	1.06	0.73	1.02	0.91	1	1
	PM1	1.63	−0.49	1.20	0.74	1.21	0.43	1.04	0.89	1	1
	PM2	1.07	1.02	1.08	0.87	1.08	0.78	1.02	0.95	1	1
	PM3	1.09	0.98	1.10	0.63	1.05	0.69	1.01	0.95	1	1
	MTM1	1.18	0.40	1.03	1.13	1.03	0.90	1.01	0.98	1	1
	MTM2	1.04	1.22	1.06	0.82	1.03	0.93	1.01	0.99	1	1
	MTM3	1.09	0.93	1.04	1.03	1.03	0.92	1.00	0.98	1	1
−0.10	MLE	1.15	−0.26	1.06	0.44	1.03	0.73	1.01	0.95	1	1
	MM	1.12	−0.10	1.07	0.37	1.04	0.66	1.01	0.91	1	1
	PM1	1.63	−2.10	1.21	0.40	1.21	−0.18	1.04	0.77	1	1
	PM2	1.07	1.00	1.07	0.73	1.08	0.52	1.01	0.91	1	1
	PM3	1.09	0.93	1.10	0.22	1.05	0.39	1.01	0.90	1	1
	MTM1	1.18	−0.33	1.03	1.24	1.03	0.78	1.01	0.96	1	1
	MTM2	1.05	1.35	1.05	0.65	1.03	0.83	1.01	0.97	1	1
	MTM3	1.09	0.80	1.04	1.02	1.02	0.83	1.00	0.97	1	1
0.15	MLE	1.17	2.02	1.07	1.44	1.03	1.22	1.01	1.05	1	1
	MM	1.06	1.43	1.03	1.21	1.02	1.10	1.00	1.02	1	1
	PM1	1.61	3.35	1.20	1.48	1.21	1.86	1.04	1.17	1	1
	PM2	1.07	1.04	1.07	1.22	1.07	1.35	1.01	1.07	1	1
	PM3	1.08	1.14	1.09	1.56	1.05	1.40	1.01	1.07	1	1
	MTM1	1.17	2.00	1.03	0.91	1.03	1.16	1.01	1.03	1	1
	MTM2	1.04	0.84	1.05	1.26	1.02	1.12	1.01	1.03	1	1
	MTM3	1.08	1.22	1.04	1.04	1.02	1.12	1.00	1.02	1	1
0.40	MLE	1.21	1.49	1.08	1.21	1.04	1.11	1.01	1.03	1	1
	MM	1.05	1.13	1.02	1.06	1.01	1.03	1.00	1.01	1	1
	PM1	1.61	2.06	1.20	1.25	1.21	1.38	1.04	1.07	1	1
	PM2	1.06	1.03	1.07	1.10	1.07	1.15	1.01	1.03	1	1
	PM3	1.08	1.08	1.09	1.23	1.05	1.15	1.01	1.02	1	1
	MTM1	1.17	1.43	1.03	0.99	1.03	1.07	1.01	1.01	1	1
	MTM2	1.03	0.96	1.05	1.10	1.02	1.05	1.00	1.01	1	1
	MTM3	1.07	1.10	1.03	1.04	1.02	1.05	1.00	1.01	1	1

NOTE: The ranges of standard errors for the simulated entries of σ and γ, respectively, are: 0.0001–0.0032 and 0.0007–0.0106 (for $\gamma=-0.20$); 0.0001–0.0042 and 0.0012–0.0217 (for $\gamma=-0.10$); 0.0002–0.0034 and 0.0008–0.0106 (for $\gamma=0.15$); 0.0001–0.0052 and 0.0003–0.0092 (for $\gamma=0.40$).

Likewise, for $\gamma = 0.40$, RE of the MLE is: 0.69 (for $n = 1,000$), 0.75 (for $n = 2,500$), 0.79 (for $n = 5,000$), 0.82 (for $n = 10,000$). Note that similar observations, with no specific numbers though, were also made by Hosking and Wallis (1987). Finally, what is a bit surprising is that the

MM estimator performs quite well around $\gamma = -0.20$ and approaches its ARE from *above*. In typical cases, finite-sample REs approach corresponding AREs from below (see, e.g., MLE for $\gamma = 0.40$).

TABLE 4: Relative efficiency of MLE, MM, PM and MTM estimators for selected values of γ. The entries are mean values based on 100,000 simulated samples of size n.

γ	*Estimator*	$n = 25$	$n = 50$	$n = 100$	$n = 500$	$n \to \infty$
-0.20	MLE	0.61	0.79	0.89	0.98	1
	MM	0.84	0.93	0.96	0.87	0.512
	PM1	0.13	0.18	0.19	0.24	0.265
	PM2	0.32	0.34	0.35	0.38	0.389
	PM3	0.36	0.36	0.39	0.42	0.426
	MTM1	0.37	0.46	0.47	0.50	0.502
	MTM2	0.43	0.44	0.46	0.48	0.483
	MTM3	0.36	0.39	0.40	0.41	0.429
-0.10	MLE	0.56	0.76	0.87	0.97	1
	MM	0.86	0.96	1.00	0.97	0.902
	PM1	0.11	0.16	0.18	0.23	0.247
	PM2	0.29	0.31	0.32	0.35	0.357
	PM3	0.33	0.33	0.36	0.38	0.389
	MTM1	0.33	0.42	0.43	0.45	0.454
	MTM2	0.39	0.40	0.42	0.44	0.440
	MTM3	0.33	0.35	0.36	0.38	0.389
0.15	MLE	0.41	0.63	0.75	0.91	1
	MM	0.66	0.76	0.81	0.85	0.865
	PM1	0.08	0.11	0.12	0.17	0.189
	PM2	0.20	0.22	0.23	0.25	0.259
	PM3	0.23	0.24	0.25	0.27	0.277
	MTM1	0.23	0.29	0.29	0.31	0.315
	MTM2	0.28	0.28	0.30	0.31	0.312
	MTM3	0.23	0.24	0.25	0.26	0.271
0.40	MLE	0.18	0.33	0.44	0.62	1
	MM	0.28	0.32	0.35	0.36	0.362
	PM1	0.03	0.05	0.06	0.08	0.095
	PM2	0.09	0.10	0.10	0.12	0.122
	PM3	0.11	0.11	0.11	0.13	0.128
	MTM1	0.10	0.13	0.13	0.14	0.141
	MTM2	0.13	0.13	0.14	0.14	0.143
	MTM3	0.11	0.11	0.11	0.12	0.120

NOTE: The range of standard errors for the simulated entries is: 0.0007–0.0036 (for $\gamma = -0.20$); 0.0007–0.0042 (for $\gamma = -0.10$); 0.0005–0.0028 (for $\gamma = 0.15$); 0.0002–0.0025 (for $\gamma = 0.40$).

In Table 5, we illustrate the behavior of estimators under several data-contamination scenarios. We choose GPD($x_0 = 0$, $\sigma = 1, \gamma = 0.15$) as the "clean" model because for $\gamma = 0.15$ the MLE and MM estimators are much more efficient than the PM and MTM estimators (see Table 4). As one can see from Table 5, however, just 1% of "bad" observations can completely erase the huge advantage of standard (non-robust) procedures over the robust PMs and MTMs. Indeed, for $\varepsilon > 0$, the MLE estimates become totally uninformative, and the MM procedure simply collapses. On the other hand, the robust estimators stay on target when estimating σ and exhibit a gradually deteriorating performance as ε increases. The deterioration, in this case, is not unexpected for as the level of data-contamination reaches or exceeds PMs' and MTMs' UBP, they also become uninformative. Finally, note that for all estimators the primary source of impaired performance is the bias in estimation of γ.

TABLE 5: Mean and relative efficiency of MLE, MM, PM and MTM estimators under several data-contamination models $F_\varepsilon = (1-\varepsilon)\,\mathrm{GPD}(x_0 = 0,\, \sigma = 1, \gamma = 0.15) + \varepsilon\,\mathrm{GPD}(\sigma = 1, \gamma = -5)$. The entries are mean values based on $100,000$ simulated samples of size $n = 500$.

Statistic	*Estimator*	$\varepsilon = 0$		$\varepsilon = 0.01$		$\varepsilon = 0.05$		$\varepsilon = 0.10$		$\varepsilon = 0.15$	
		$\widehat{\sigma}$	$\widehat{\gamma}$	$\widehat{\sigma}$	$\widehat{\gamma}$	$\widehat{\sigma}$	$\widehat{\gamma}$	$\widehat{\sigma}$	$\widehat{\gamma}$	$\widehat{\sigma}$	$\widehat{\gamma}$
MEAN	MLE	1.01	0.16	0.76	−0.29	0.65	−0.81	0.61	−1.19	0.60	−1.51
	MM*	1.00	0.15	∞	−0.41	∞	−0.50	∞	−0.50	∞	−0.50
	PM1	1.04	0.18	1.04	0.15	1.04	0.05	1.04	−0.46	1.02	−1.36
	PM2	1.01	0.16	1.02	0.14	1.02	0.07	1.03	−0.05	1.03	−0.25
	PM3	1.01	0.16	1.01	0.15	1.02	0.08	1.03	−0.01	1.03	−0.12
	MTM1	1.01	0.15	1.01	0.14	1.01	0.07	1.01	−0.12	0.96	−0.74
	MTM2	1.01	0.15	1.01	0.14	1.01	0.07	1.02	−0.02	1.02	−0.15
	MTM3	1.01	0.15	1.01	0.14	1.02	0.08	1.03	0.00	1.03	−0.10
RE	MLE	0.91		0.07		0.03		0.02		0.02	
	MM*	0.85		0		0		0		0	
	PM1	0.17		0.14		0.05		0.01		0.00	
	PM2	0.25		0.23		0.10		0.05		0.02	
	PM3	0.27		0.26		0.14		0.07		0.04	
	MTM1	0.31		0.28		0.13		0.04		0.02	
	MTM2	0.31		0.28		0.14		0.07		0.04	
	MTM3	0.26		0.25		0.15		0.08		0.05	

* For the MM estimator, the entries ∞ and 0 correspond to numbers of the order 10^{21} and 10^{-20}, respectively.

3 Actuarial Applications

In this section, we fit the GPD model to the Danish insurance data which has been extensively studied in the actuarial literature (see, e.g., McNeil, 1997). We also investigate the implications of a model fit on risk evaluations and ratemaking. In particular, we use empirical, parametric and robust parametric approaches to compute point estimates of several value-at-risk measures and net premiums of a few insurance contracts.

3.1 Fitting Insurance Data

The Danish insurance data were collected at Copenhagen Re and comprise 2167 fire losses over the period 1980 to 1990. They have been adjusted for inflation to reflect 1985 values and are expressed in millions of Danish krones (DKK). All losses are one million DKK or larger. A thorough diagnostic analysis of this data set was performed by McNeil (1997) who concluded that the GPD assumption is a reasonable one for the Danish insurance data. He then used the MLE approach to fit the GPD to the whole data set as well as to data above various thresholds. The important issues of 'model uncertainty', 'parameter uncertainty', and 'data uncertainty' were also discussed by McNeil (1997). Some of those discussions are taken as motivation for the foregoing investigations.

In this section, we fit the GPD to the entire data set and to data above several thresholds, and examine the stability of fits under a few data-perturbation scenarios. Our main objectives are to see: (a) how important is the choice of a parameter-estimation method on the model fit, and (b) what influence it has on subsequent pricing and risk measurement exercises. For visual and quantitative assessments of the quality of model fit, we employ the percentile-residual (PR) plot and a *trimmed* mean absolute deviation (tMAD), respectively. These tools are taken from Brazauskas (2009) and are defined as follows. The PR plots are constructed by plotting the empirical percentile levels, $(j/n)100\%$, versus the standardized residuals

$$R_{j,n} \;=\; \frac{X_{j:n} - \widehat{F}^{-1}\left(\frac{j-0.5}{n}\right)}{\text{standard deviation of } \widehat{F}^{-1}\left(\frac{j-0.5}{n}\right)} \quad \text{for } j = 1, \ldots, n, \tag{3.1}$$

where $X_{j:n}$ is the observed (j/n)th quantile and the qf F^{-1}, given by (2.3), is estimated by replacing parameters σ and γ with their respective estimates $\widehat{\sigma}$ and $\widehat{\gamma}$. For finding $\widehat{\sigma}$ and $\widehat{\gamma}$ we use the following estimators: MLE, PM3 with $(p_1, p_2) = (0.30, 0.85)$, MTM3 with $(a_1, b_1) = (0.30, 0.50)$, $(a_2, b_2) = (0.70, 0.15)$, and MTM4 with $(a_1, b_1) = (0.10, 0.55)$, $(a_2, b_2) = (0.70, 0.05)$. The MM approach is excluded from further consideration because its asymptotic properties are not valid for the Danish insurance data. The denominator of (3.1) will be estimated by using the delta method (see, e.g., Serfling, 1980, Section 3.3) in conjunction with the corresponding variance-covariance matrix $\mathbf{\Sigma_0}$, $\mathbf{\Sigma_2}$, or $\mathbf{\Sigma_3}$; the matrices are defined by (2.6), (2.9), (2.12), respectively. In the PR-plot, the horizontal line at 0 represents the estimated quantiles, and the ± 2.5 lines are the tolerance limits. A good fit would be the one for which the majority of points (ideally, all points) are scattered between the tolerance limits. The PR-plots for MLE, PM, and MTM fits are presented in Figure 2. The plots are based on the whole data set, i.e., on 2156 losses in excess of one million DKK.

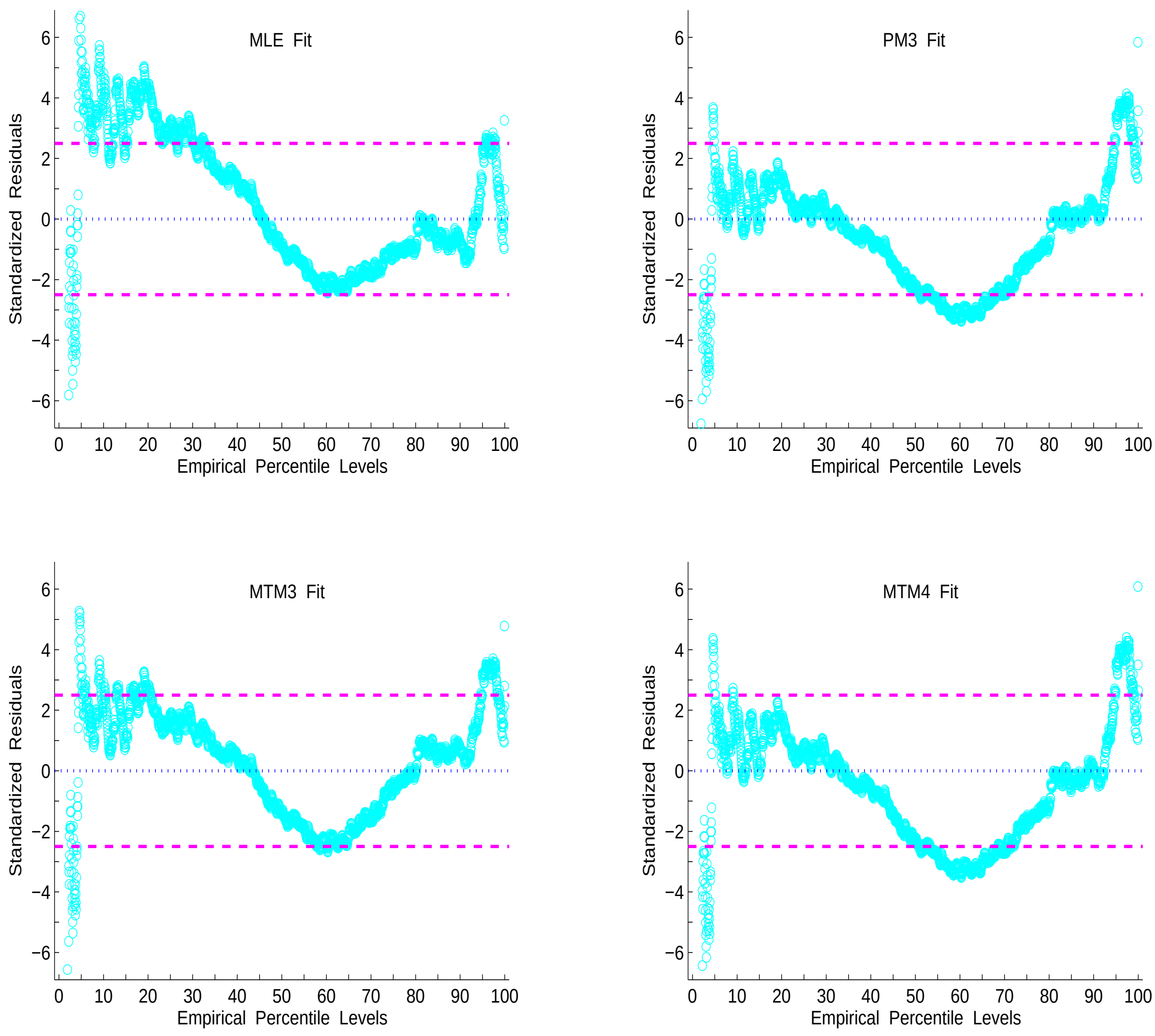

FIGURE 2: PR-plots for the GPD model fitted by the MLE, PM, and MTM methods.

As one can see from Figure 2, all parameter-estimation methods do a mediocre job at fitting the "small" losses, i.e., those slightly above the threshold of one million, but they perform reasonably well for the "medium" and "large" section of the losses. Overall, among the four PR-plots, the MLE fit looks worst. One should keep in mind, however, that the vertical deviations in these plots depend on the efficiency of the estimator. Thus, the same value of absolute residual will appear as significantly larger on the MLE's plot than it will on, for example, the MTM4's because the latter estimator is less efficient. Indeed, its ARE is: 0.832 (for $\gamma = -1$), 0.711 (for $\gamma = -0.50$), 0.609 (for $\gamma = -0.25$). Therefore, since for practical decision-making the actual (not statistical!) discrepancies matter more, it is important to monitor the non-standardized residuals as well.

Next, the tMAD measure evaluates the absolute distance between the fitted GPD quantiles and the observed data. It is defined by $\Delta_\delta = \frac{1}{[n\delta]} \sum_{i=1}^{[n\delta]} b_{i:n}$, where $b_{i:n}$ denotes the ith smallest distance among $\left| X_{j:n} - \widehat{F}^{-1}\left((j-0.5)/n\right) \right|$, $j = 1, \ldots, n$. We use the following values of δ: 0.50, 0.75, 0.90, 0.95, 1. The choice $\delta = 0.90$, for instance, indicates how far, on the average, are the 90% closest observations from their corresponding fitted quantiles. In Table 6, we report parameter estimates and the goodness-of-fit measurements Δ_δ for various data thresholds x_0.

Table 6: Parameter estimates and goodness-of-fit measurements of the GPD model for selected model-fitting procedures and several data thresholds x_0.

Threshold	Fitting	Parameter Estimates		Model Fit (Δ_δ)				
(Excesses)	Procedure	$\widehat{\sigma}$	$\widehat{\gamma}$	$\delta = 0.50$	$\delta = 0.75$	$\delta = 0.90$	$\delta = 0.95$	$\delta = 1$
$x_0 = 1$ (2156)	MLE	0.946	−0.604	0.02	0.04	0.05	0.06	0.19
	PM3	1.036	−0.501	0.01	0.03	0.05	0.07	0.43
	MTM3	0.989	−0.520	0.02	0.03	0.04	0.07	0.41
	MTM4	1.035	−0.515	0.01	0.04	0.05	0.07	0.39
$x_0 = 3$ (532)	MLE	2.189	−0.668	0.03	0.06	0.16	0.23	0.74
	PM3	2.171	−0.788	0.03	0.10	0.17	0.29	2.34
	MTM3	2.079	−0.794	0.03	0.08	0.12	0.22	2.19
	MTM4	2.209	−0.720	0.03	0.10	0.16	0.21	1.26
$x_0 = 10$ (109)	MLE	6.975	−0.497	0.12	0.26	0.46	0.61	2.21
	PM3	7.101	−0.345	0.13	0.28	0.41	0.64	3.53
	MTM3	7.819	−0.290	0.08	0.15	0.24	0.47	3.51
	MTM4	7.546	−0.377	0.08	0.18	0.33	0.43	2.85
$x_0 = 20$ (36)	MLE	9.635	−0.684	0.28	0.52	0.91	1.34	3.32
	PM3	11.751	−0.476	0.29	0.70	1.12	2.51	6.27
	MTM3	9.920	−0.686	0.31	0.64	1.13	1.80	3.59
	MTM4	10.524	−0.813	0.37	1.33	2.73	4.06	9.13

After examining Table 6 we make the following observations. Clearly, the MLE fits are best for all data-thresholds under consideration if we measure the fit by Δ_δ with $\delta = 1$. This should not be surprising since likelihood-based procedures attempt, and are designed, to fit *all* data points. But if we look at the other values of δ (which reflect the fit for most—not all—observations), we see that the MLE and robust fits are similar and fairly close to the actual data, for all data-thresholds. We also note a strong performance by MTMs for $x_0 = 10$. Aside from the last point, however, so far the robust procedures have not offered any significant improvements over the MLE. But this changes substantially when we perform a sensitivity analysis under a few data-perturbation scenarios.

TABLE 7: Parameter estimates and goodness-of-fit measurements of the GPD model for selected model-fitting procedures, $x_0 = 10$, and under several data-perturbation scenarios.

Scenario	Fitting Procedure	Parameter Estimates $\widehat{\sigma}$	$\widehat{\gamma}$	Model Fit (Δ_δ) $\delta = 0.50$	$\delta = 0.75$	$\delta = 0.90$	$\delta = 0.95$	$\delta = 1$
Remove $x = 263$	MLE	7.230	−0.390	0.09	0.19	0.30	0.40	1.14
	PM3	7.132	−0.321	0.11	0.23	0.35	0.47	1.66
	MTM3	7.709	−0.267	0.08	0.15	0.24	0.38	1.73
	MTM4	7.420	−0.336	0.08	0.16	0.26	0.35	1.32
Add $x = 350$	MLE	6.778	−0.598	0.16	0.32	0.68	0.90	3.70
	PM3	7.422	−0.304	0.11	0.29	0.49	0.90	6.70
	MTM3	7.897	−0.316	0.09	0.16	0.26	0.56	6.09
	MTM4	7.620	−0.421	0.09	0.21	0.40	0.53	5.14
Replace $x = 263$ with $x = 350$	MLE	6.892	−0.517	0.13	0.26	0.51	0.68	3.02
	PM3	7.101	−0.345	0.13	0.28	0.41	0.64	4.42
	MTM3	7.819	−0.290	0.08	0.15	0.24	0.47	4.40
	MTM4	7.546	−0.377	0.08	0.18	0.33	0.43	3.74

In Table 7, we report parameter estimates and the goodness-of-fit measurements Δ_δ for $x_0 = 10$ under the following three scenarios. The first two scenarios are taken from McNeil (1997) and the third one is a combination of the other two. In the first scenario (labeled "Remove $x = 263$") we remove the largest loss from the original sample. In the second scenario (labeled "Add $x = 350$") we introduce a new largest observation of 350 to the data set. And in the third scenario (labeled "Replace $x = 263$ with $x = 350$") we replace the current largest point of 263 with a new loss of 350. If we compare parameter estimates and Δ_δ evaluations with the original results in Table 6 (for $x_0 = 10$), we see that robust estimates and their fits are significantly less affected by the removal or addition of the largest loss than those of the MLE. Moreover, for the third scenario all robust estimates and their fit measurements for $\delta < 1$ are absolutely identical to the original ones whilst the results for MLE are distorted. In the next two sections, we further explore the data-perturbation effects in insurance applications by studying their influence on the estimates of risk measures and net premiums.

3.2 Risk Measurement

To see how the quality of model fit affects insurance risk evaluations, we compute empirical, parametric and robust parametric point estimates for several value-at-risk (VaR) measures. The computations are performed for the whole data set (i.e., for $x_0 = 1$) and for data above the threshold $x_0 = 10$ including some data-perturbation scenarios of Section 3.1. Mathematically, the VaR measure is the $(1-\beta)$-level quantile of the distribution function F, that is, $\mathrm{VaR}(\beta) = F^{-1}(1-\beta)$. For empirical estimation, we replace F with the empirical cdf $\widehat{F}_n$ and arrive at

$$\widehat{\mathrm{VaR}}_{\mathrm{EMP}}(\beta) = X_{n:n-[n\beta]}.$$

For parametric (MLE) and robust parametric (PM, MTM) estimation, $\widehat{F}^{-1}$ is found by simply replacing parameters σ and γ with their respective MLE, PM, and MTM estimates in (2.3). Note that for parametric quantile-estimation based on upper tail of the data (i.e., for the thresholds $x_0 > 1$), we apply the results of McNeil (1997, Section 3.5). In particular, we use σ, γ and location x_0 estimates which are calculated according to the following formulas:

$$\widetilde{\gamma} = \widehat{\gamma}, \quad \widetilde{\sigma} = \widehat{\sigma}(1-\widehat{F}_n(x_0))^{-\widetilde{\gamma}}, \quad \widetilde{x}_0 = x_0 + (\widetilde{\sigma}/\widetilde{\gamma})\left((1-\widehat{F}_n(x_0))^{\widetilde{\gamma}} - 1\right),$$

where $\widehat{\sigma}$ and $\widehat{\gamma}$ are the parameter estimates based only on data above x_0. Table 8 presents empirical, parametric, and robust parametric point estimates of $\mathrm{VaR}(\beta)$ for several levels of β.

A number of conclusions emerge from the table. First, for the whole data set ($x_0 = 1$), GPD-based risk evaluations of not-too-extreme significance levels ($\beta \geq 0.01$) are fairly close to their empirical counterparts. Second, for very extreme significance levels ($\beta < 0.01$), the empirical and parametric estimates of VaR diverge. Of course, one can argue that the empirical approach underestimates $\mathrm{VaR}(\beta = 0.0001)$ because there is simply no observed data at that level. On the other hand, the MLE's estimate seems like an exaggeration of risk. Third, the last point gains even more credibility if we look at $\mathrm{VaR}(\beta < 0.01)$ estimates which are based on data above $x_0 = 10$. Indeed, the MLE's evaluations are now substantially reduced. Fourth, overall the robust procedures tend to provide lower estimates of risk at the most extreme levels of significance than the MLE. This actually can easily be seen from the PR-plots (see Figure 2): near the 100th percentile, MLE's residuals are below the fitted line whilst the PM's and MTM's are above. Fifth, when we employ data-perturbation scenarios, robust estimators' risk evaluations are quite stable compared to those of the MLE.

TABLE 8: Point estimates of various value-at-risk measures computed by employing empirical, parametric (MLE), and robust parametric (PM and MTM) methods.

Scenario	Estimation Method	VaR(β) $\beta = 0.10$	$\beta = 0.05$	$\beta = 0.01$	$\beta = 0.001$	$\beta = 0.0001$
All Data ($x_0 = 1$)	MLE	5.73	9.0	25	101	408
	PM3	5.48	8.2	20	65	207
	MTM3	5.40	8.1	20	68	228
	MTM4	5.57	8.4	21	70	230
	EMPIRICAL	5.56	10.1	26	145	263
$x_0 = 10$	MLE	5.96	10.1	27	95	306
	PM3	5.68	10.1	25	69	166
	MTM3	5.16	10.1	26	67	147
	MTM4	5.46	10.1	27	78	199
$x_0 = 10$ and replace $x = 263$ with $x = 350$	MLE	6.04	10.1	27	98	331
	PM3	5.68	10.1	25	69	166
	MTM3	5.16	10.1	26	67	147
	MTM4	5.46	10.1	27	78	199
$x_0 = 10$ and add $x = 350$	MLE	6.20	10.1	29	117	469
	PM3	5.43	10.1	26	66	147
	MTM3	5.15	10.1	27	71	164
	MTM4	5.48	10.1	28	86	241

3.3 Contract Pricing

Consider now estimation of the pure premium for an insurance benefit equal to the amount by which a loss exceeds l (million DKK) with a maximum benefit of m. That is, if the fire damage is X with distribution function F, we seek

$$\Pi[F] = \int_{l}^{l+m} (x - l)\, \mathrm{d}F(x) + m\big(1 - F(l+m)\big). \tag{3.2}$$

Since $\Pi[F]$ is a functional of the underlying loss distribution F, we can estimate it by replacing F with its estimate. To accomplish that, we employ three approaches: empirical, parametric (MLE), and robust parametric (PM and MTM). In addition, we also provide (estimated) standard errors of the premium $\Pi[F]$ estimates. To find parametric estimates of the errors, we use the delta method applied to the transformation of parameter estimators given by equation (3.2) together with the MLE, PM, and MTM asymptotic distributions, which have been discussed earlier. For the empirical estimation of standard errors, we use the classical central limit theorem and have that

$$\Pi[\widehat{F}_n] \;\sim\; \mathcal{AN}\big(\Pi[F],\, n^{-1}V[F]\big),$$

where $V[F] = \int_l^{l+m}(x-l)^2\, \mathrm{d}F(x) + m^2(1-F(l+m)) - (\Pi[F])^2$ and $\widehat{F}_n$ denotes the empirical distribution function. Further, the reliability of premium estimates is studied by employing two data-perturbation scenarios. In the first scenario (labeled "Replace Top 10 with ~ 350"), we make the ten largest losses even larger by replacing them with $351, 352, \ldots, 360$. In the second scenario (labeled "Replace Top 10 with ~ 100"), we replace the ten largest losses with $101, 102, \ldots, 110$. Table 9 summarizes our numerical investigations for contract pricing.

Table 9: Empirical, parametric (MLE), and robust parametric (PM and MTM) point estimates of $\Pi[F]$, for selected insurance contracts and under two data-perturbation scenarios. Estimated standard errors of the premium estimates are presented in parentheses.

Scenario	Estimation Method	Insurance contract specified by (l,m)			
		$(l,m)=(2,3)$ Premium	$(l,m)=(5,10)$ Premium	$(l,m)=(20,20)$ Premium	$(l,m)=(50,50)$ Premium
Original Data ($x_0=1$)	MLE	0.69 (0.021)	0.51 (0.037)	0.16 (0.025)	0.09 (0.021)
	PM3	0.69 (0.024)	0.44 (0.051)	0.10 (0.032)	0.04 (0.021)
	MTM3	0.67 (0.023)	0.43 (0.056)	0.10 (0.035)	0.05 (0.024)
	MTM4	0.70 (0.023)	0.46 (0.043)	0.11 (0.027)	0.05 (0.019)
	EMPIRICAL	0.66 (0.023)	0.54 (0.043)	0.17 (0.034)	0.08 (0.041)
Replace Top 10 with ~ 350	MLE	0.70 (0.021)	0.55 (0.038)	0.19 (0.028)	0.12 (0.026)
	PM3	0.69 (0.024)	0.44 (0.051)	0.10 (0.032)	0.04 (0.021)
	MTM3	0.67 (0.023)	0.43 (0.056)	0.10 (0.035)	0.05 (0.024)
	MTM4	0.70 (0.023)	0.46 (0.043)	0.11 (0.027)	0.05 (0.019)
	EMPIRICAL	0.66 (0.023)	0.54 (0.043)	0.17 (0.034)	0.08 (0.041)
Replace Top 10 with ~ 100	MLE	0.69 (0.021)	0.52 (0.037)	0.17 (0.026)	0.09 (0.022)
	PM3	0.69 (0.024)	0.44 (0.051)	0.10 (0.032)	0.04 (0.021)
	MTM3	0.67 (0.023)	0.43 (0.056)	0.10 (0.035)	0.05 (0.024)
	MTM4	0.70 (0.023)	0.46 (0.043)	0.11 (0.027)	0.05 (0.019)
	EMPIRICAL	0.66 (0.023)	0.54 (0.043)	0.17 (0.034)	0.08 (0.041)

We observe that premium estimates for the contract $(l, m) = (2, 3)$, which covers events that are quite likely but have relatively low economic impact, are practically identical. And the data-perturbation scenarios have no effect on any of the estimators, including the MLE. For more extreme coverages, i.e., for contracts $(l, m) = (5, 10)$ and higher, which cover low-probability-but-high-consequence events, the estimates diverge. The MLE and empirical estimates are usually close to each other whilst the robust estimates form their own cluster which yields fairly different prices from those of the MLE and empirical methods. Further, for very extreme layers of losses such as $(l, m) = (20, 20)$, $(50, 50)$, the parametric estimators produce smaller standard errors than the empirical approach. Finally, since both data contaminations occur outside the defined layers, one would assume that the premium should not change. This property is exhibited by the empirical and robust approaches, but the MLE is clearly affected by contamination.

4 Discussion

In this article, we have introduced a new method for robust fitting of the GPD. It is based on 'trimmed moments' and therefore called the method of trimmed moments (MTM). Its large- and small-sample properties have been explored and compared to some well-established standard approaches (MLE and MM) as well as to some less-known but conceptually sound methods such as 'percentile matching' (PM). We have found that the new MTM procedure is computationally attractive, it possesses competitive efficiency properties and provides sufficient protection against various data contamination sources. A connection between the MTMs and PMs has also been established. Thus it is not surprising at all that these two procedures have some desirable properties in common. In particular, they offer a variety of robustness-efficiency trade-offs and, unlike other existing proposals in the literature, both are applicable and valid for the entire parameter space of the GPD (i.e., for $-\infty < \gamma < \infty$).

In addition, the favorable theoretical and computational properties of the new method translate into accurate risk evaluations as well as fair pricing. Indeed, as it is confirmed by our numerical illustrations, the value-at-risk estimates based on the robust procedures show more stability under various scenarios of data-perturbation than those based on the MLE. Also, robustly estimated contract prices for extreme layers of losses are less volatile than the empirical ones, and more outlier resistant than the MLE-based prices.

5 Acknowledgment

The first author gratefully acknowledges the support provided by a grant from the Actuarial Profession (United Kingdom).

A Appendix: Auxiliary Results

A.1 Differentiation of $\boldsymbol{\mu_i}$

The derivative of a trimmed moment μ_j, given by (2.11), with respect to the shape parameter γ is necessary for computations of the variance-covariance matrix $\mathbf{\Sigma_3}$, given by (2.12). The general case is found by straightforward differentiation. The special cases $\gamma = -1$ and $\gamma = 0$ represent the limit of the general case when $\gamma \to -1$ and $\gamma \to 0$, respectively. Thus, we have:

$$\mu_j' = -\sigma \times \begin{cases} 1 - \dfrac{\log(1-a_j) - \log(b_j)}{1-a_j-b_j} - \dfrac{\log^2(1-a_j) - \log^2(b_j)}{2(1-a_j-b_j)}, & \text{if } \gamma = -1, \\ 1 - \dfrac{(1-a_j)\log(1-a_j) - b_j\log(b_j)}{1-a_j-b_j} + \dfrac{(1-a_j)\log^2(1-a_j) - b_j\log^2(b_j)}{2(1-a_j-b_j)}, & \text{if } \gamma = 0, \\ \dfrac{1}{\gamma(\gamma+1)}\left[(2\gamma+1)(\mu_j/\sigma) - 1 + \dfrac{(1-a_j)^{\gamma+1}\log(1-a_j) - b_j^{\gamma+1}\log(b_j)}{1-a_j-b_j}\right], & \text{otherwise.} \end{cases}$$

A.2 Approximation of $\boldsymbol{\sigma^2_{ij}}$

The entries σ^2_{ij} appear in the variance-covariance matrix $\mathbf{\Sigma}_{**}$ which is a part of $\mathbf{\Sigma_3}$, given by (2.12). Instead of pursuing exact formulas for these entries, we use the bivariate trapezoidal rule.

Let us define the region $R = [a, b] \times [c, d]$ which is a subset of $[0, 1) \times [0, 1)$. Next, we divide each interval, $[a, b]$ and $[c, d]$, into k subintervals and define

$$\begin{aligned} h_x &= (b-a)/k, \qquad x_i = a + ih_x, \quad i = 0, \ldots, k, \\ h_y &= (d-c)/k, \qquad y_j = c + jh_y, \quad j = 0, \ldots, k, \end{aligned}$$

where the number k is fixed.

The double integral over R to be approximated is of the form

$$I = \int_a^b \int_c^d g(x, y) f'(x) f'(y) \, \mathrm{d}y \, \mathrm{d}x.$$

Using the composite trapezoidal rule in both spatial directions yields

$$I \approx I_k = \sum_{i=0}^{k}\sum_{j=0}^{k} g(x_i, y_j)\, f'(x_i) f'(y_j)\, w_i w_j,$$

where

$$w_i = h_x \times \begin{cases} 1/2, & \text{if } i = 0 \text{ or } i = k, \\ 1, & \text{otherwise,} \end{cases} \qquad \text{and} \qquad w_j = h_y \times \begin{cases} 1/2, & \text{if } j = 0 \text{ or } j = k, \\ 1, & \text{otherwise.} \end{cases}$$

If the integrand $g(\cdot,\cdot)\, f'(\cdot) f'(\cdot)$ is sufficiently smooth, then the order of convergence is $\mathcal{O}(h^2)$.

The derivatives $f'(x)$ and $f'(y)$ are approximated by

$$\begin{aligned} f'(x_i) &\approx \frac{1}{2h_x}\big(f(x_{i+1}) - f(x_{i-1})\big), \qquad i = 1, \ldots, k, \\ f'(y_j) &\approx \frac{1}{2h_y}\big(f(y_{j+1}) - f(y_{j-1})\big), \qquad j = 1, \ldots, k, \end{aligned}$$

where x_i, y_j, h_x and h_y are given above. Here, if the function f is assumed to be sufficiently smooth, then the order of convergence is $\mathcal{O}(h^2)$. This means that the *error-of-convergence* ratios,

$$\text{EOC} = \frac{I - I_{k/2}}{I - I_k},$$

approach 4 if the true solution is known. If not, then the following ratios

$$\text{EOC} \approx \frac{I_k - I_{k/2}}{I_{k/2} - I_{k/4}}$$

approach 4.

Finally, note that all the smoothness conditions made above are satisfied by the GPD.

References

[1] Brazauskas, V. (2009). Robust and efficient fitting of loss models: diagnostic tools and insights. *North American Actuarial Journal* (in press).

[2] Brazauskas, V. and Serfling, R. (2000). Robust and efficient estimation of the tail index of a single-parameter Pareto distribution (with discussion). *North American Actuarial Journal*, **4**(4), 12–27. Discussion: **5**(3), 123–126. Reply: **5**(3), 126–128.

[3] Brazauskas, V., Jones, B.L., and Zitikis, R. (2009). Robust fitting of claim severity distributions and the method of trimmed moments. *Journal of Statistical Planning and Inference*, **139**(6), 2028–2043.

[4] Castillo, E. and Hadi, A.S. (1997). Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association*, **92**, 1609–1620.

[5] Cebrian, A.C., Denuit, M., and Lambert, P. (2003). Generalized Pareto fit to the Society of Actuaries' large claims database. *North American Actuarial Journal*, **7**(3), 18–36.

[6] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society* (Series B), **52**, 393–442.

[7] Dell'Aquila, R. and Embrechts, P. (2006). Extremes and robustness: a contradiction? *Financial Markets and Portfolio Management*, **20**(1), 103–118.

[8] Dupuis, D.J. (1999). Exceedances over high thresholds: a guide to threshold selection. *Extremes*, **1**(3), 251–261.

[9] He, X. and Fung, W.K. (1999). Method of medians for lifetime data with Weibull models. *Statistics in Medicine*, **18**, 1993–2009.

[10] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, **29**(3), 339–349.

[11] Johnson, N.L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, 2nd edition; Volume 1. Wiley, New York.

[12] Juárez, S.F. and Schucany, W.R. (2004). Robust and efficient estimation for the generalized Pareto distribution. *Extremes*, **7**(3), 237–251.

[13] Klugman, S.A., Panjer, H.H., and Willmot, G.E. (2004). *Loss Models: From Data to Decisions*, 2nd edition. Wiley, New York.

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

[15] McNeil A.J., Frey R., and Embrechts P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

[16] Peng, L. and Welsh, A.H. (2001). Robust estimation of the generalized Pareto distribution. *Extremes*, **4**(1), 53–65.

[17] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

[18] Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, **72**, 67–90.