



## Generalised Smooth Tests for the Generalised Pareto Distribution

B. De Boeck, *Department of Mathematical Modelling,  
Statistics and Bioinformatics, Ghent University, Ghent,  
Belgium. Email: deboeck@gmail.com*

O. Thas<sup>1</sup>, *Department of Mathematical Modelling,  
Statistics and Bioinformatics, Ghent University, Ghent,  
Belgium. Email: Olivier.Thas@UGent.be*

J. C. W. Rayner, *Centre for Statistical and Survey Methodology,  
School of Mathematics and Applied Statistics,  
University of Wollongong, NSW, Australia &  
School of Mathematical and Physical Sciences,  
University of Newcastle, NSW, Australia.  
Email: John.Rayner@newcastle.edu.au*

D. J. Best, *School of Mathematical and Physical Sciences,  
University of Newcastle, NSW, Australia.  
Email: John.Best@newcastle.edu.au*

*Received: July 21, 2011    Revised: September 23, 2011*

---

### Abstract

The generalised Pareto distribution (GPD) is often used to model extreme values. New smooth tests of goodness of fit are proposed for this distribution. Typical problems with the GPD are that not all moments exist and not all classical estimation procedures work well over the whole parameter space. The generalised smooth test has good powers within a subset of the parameter space for which other tests may not be defined or appropriate, and, conversely, the Anderson-Darling test performs well when the test proposed here does not.

*AMS Subject Classification: 62E20; 62G10.*

*Key-words: Goodness of fit; Orthonormal polynomials; Generalised smooth tests; Method of moments.*

---

<sup>1</sup>Corresponding Author

## 1. Introduction

Grimshaw (1993), Castillo and Hadi (1997) and Choulakian and Stephens (2001) give compelling examples of the application of the generalised Pareto distribution (GPD) to model exceedances over both known and fixed threshold. Early interest in the GPD focused on the method of estimation of the parameters. For an overview of the various estimation methods we refer to Zhang (2007).

Castillo and Hadi (1997) proposed an assessment of the GPD model involving a subjective judgement of the linearity of a scatter plot. Choulakian and Stephens (2001) applied the Cramér-von Mises and Anderson-Darling tests to the probability integral transformed observations, using the maximum likelihood estimators. Radouane and Crétois (2002) used Neyman smooth tests after the probability integral transformation. They essentially used the smooth tests first described by Thomas and Pierce (1979) and Kopecky and Pierce (1979). We, on the other hand, propose a generalised smooth test of the type of Rayner, Thas and Best (2009, chap. 9) based on method of moments estimation. Generalised smooth tests are diagnostic in the sense that at the rejection of the null hypothesis the individual components of the test statistic suggest in which moments the true distribution differs from the hypothesised GPD. The method of moments is used for the estimation of the nuisance parameters and the parametric bootstrap is recommended for  $p$ -value calculation. Our tests are implemented in the `smoothtest` R package which may be downloaded from the website, <http://biomath.ugent.be/~othas/smooth2>.

Testing for goodness of fit for the GPD suffers from some particular problems. First, not all moments exist and the number of moments depends on the parameters of the GPD. This may cause existence problems of our smooth test and any other test that relies on the method of moments estimators. Based on a simulation study we identify a subset of the parameter space for which our tests exist and performs well. Goodness of fit tests based on maximum likelihood estimators of the nuisance parameters suffer from other problems. For example, Smith (1984) showed that the rate of convergence of the MLE may be slower than  $n^{-1/2}$  and that the MLE may even not exist in yet another subset of the parameter space. Thus every test has actually its own *preferable range of application*, depending on the GPD parameters. More details follow later in the paper.

The structure of this paper is as follows. In the next section the GPD will be defined, and method of moments parameter estimation is discussed. In Section 3 the generalised smooth test of goodness of fit is described. Section 4 gives an empirical size and power study for comparing the new tests with those of Choulakian and Stephens (2001) and Radouane and Crétois (2002). Finally, some example data sets are analysed in Section 5 and a discussion follows in Section 6.

## 2. The generalised Pareto distribution and parameter estimation

### 2.1. The generalised Pareto distribution

The generalised Pareto distribution (GPD) was considered first by Pickands (1975). A random variable  $X$  has the GPD if it has cumulative distribution function

$$F_X(x; \sigma, \kappa) = \begin{cases} 1 - (1 - \kappa x / \sigma)^{1/\kappa} & \text{if } \kappa \neq 0 \\ 1 - \exp(-x/\sigma) & \text{if } \kappa = 0, \end{cases} \quad (2.1)$$

where the range of  $x$  is  $0 \leq x < \infty$  if  $\kappa \leq 0$  and  $0 \leq x < \sigma/\kappa$  if  $\kappa > 0$ , and in which  $\sigma > 0$  and  $-\infty < \kappa < \infty$ . Here  $\kappa$  is a *shape* parameter and  $\sigma$  a *scale* parameter. The ordinary Pareto distribution corresponds to  $\kappa < 0$  in the GPD family. The exponential distribution ( $\kappa = 0$ ) and the uniform distribution ( $\kappa = 1$ ) are special cases of the GPD.

The Pareto distribution is often used to model extreme values. In particular, every independent measurement of a certain characteristic, which exceeds a chosen threshold  $x_0$ , is assumed to be a realization of  $X + x_0$ . The exceedances over the threshold  $x_0$  are thus realizations of  $X$ . The GPD is just a more general way for modelling these exceedances over  $x_0$ . Note that in our setting the threshold  $x_0$  is always assumed known. We want to test whether exceedances over the known threshold  $x_0$  come from a GPD. Choulakian and Stephens (2001) describe how the threshold can be estimated using the GPD in more general extreme value problems.

If  $\kappa > 0$  the GPD has finite support which means there is a finite maximum value  $\sigma/\kappa$  for  $X$ . Moreover, when  $\kappa \geq 1/2$  the derivative of the probability density function makes a discontinuous jump at the end point  $\sigma/\kappa$ . This, among other factors like infinite variance for  $\kappa \leq -1/2$ , motivated Hosking and Wallis (1987) to restrict the shape parameter to  $-1/2 < \kappa < 1/2$ . Castillo and Hadi (1997), however, argued for not making these restrictions and gave examples where the estimates of  $\kappa$  exceed  $1/2$ . Thus, while most practical applications correspond to  $-1/2 < \kappa < 1/2$ , there is reason to allow for shape parameter values outside this interval.

**2.2. Parameter estimation**

The family of GPD densities corresponding to (2.1), and with the same range, is

$$f(x; \sigma, \kappa) = \begin{cases} (1/\sigma)(1 - \kappa x/\sigma)^{1/\kappa-1} & \text{if } \kappa \neq 0 \\ (1/\sigma) \exp(-x/\sigma) & \text{if } \kappa = 0, \end{cases}$$

and so it is routine to show that for the GPD and for every real  $r$

$$E[(1 - \kappa X/\sigma)^r] = 1/(1 + r\kappa),$$

but only if  $1 + r\kappa > 0$ , and otherwise the expectation does not exist. From this expression we find that the mean  $E[X] = \sigma/(1 + \kappa)$  exists if and only if  $\kappa > -1$ , and that similar results hold for the positive integer moments of the GPD. In particular, for  $r = 0, 1, 2, \dots$  we have

$$E[(X - E[X])^r] = \left(-\frac{\sigma}{\kappa}\right)^r \sum_{i=0}^r \binom{r}{i} \left(\frac{-1}{1 + \kappa}\right)^{r-i} \frac{1}{1 + i\kappa},$$

if  $\kappa > -1/r$ , and no existence otherwise. Therefore the mean of  $X$  does not exist if  $\kappa \leq -1$ , and the variance does not exist if  $\kappa \leq -1/2$ , and so on. This implies that method of moments estimators (MME) of the parameters of the GPD exist iff  $\kappa > -1/2$ . The MME of the parameters  $\kappa$  and  $\sigma$  are given by

$$\tilde{\kappa} = (\bar{X}^2/S^2 - 1)/2, \quad \tilde{\sigma} = \bar{X}(1 + \tilde{\kappa}),$$

Downloaded by [University of Newcastle (Australia)], [J. C. W. Rayner] at 17:51 17 February 2012

where  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  and  $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  are the mean and variance of the sample observations  $X_1, \dots, X_n$ . These estimators always exist, but are only consistent for a GPD with  $\kappa > -1/2$ . This immediately restricts the preferable range of application of MME based tests, irrespective of other disadvantages of MME.

MME is not the most efficient method of estimation for the GPD, and according to Hosking and Wallis (1987) MME is not reliable when  $\kappa < -0.2$ . MME also suffers from yet another problem, as Castillo and Hadi (1997) noted. They said that when  $\tilde{\kappa} > 0$  it is incidentally possible that sample observations fall outside the range implied by the MME. Indeed, when  $\tilde{\kappa} > 0$  the condition  $\tilde{\sigma}/\tilde{\kappa} > \max\{X_i\}_{1 \leq i \leq n}$  is not fulfilled occasionally. Nevertheless, one can still find good arguments for considering MME based tests.

The maximum likelihood estimation (MLE) based smooth tests discussed in Rayner and Best (1989) have useful approximate moment interpretations that are, however, only valid for testing distributions for which MLE and MME coincide. In particular, the smooth test statistic can then be decomposed into interpretable components in terms of deviations between sample and hypothesised distribution moments (Rayner, Best and Mathews, 1995). Henze and Klar (1996) and Klar (2000) showed that, at least theoretically, the components should be appropriately rescaled for this diagnostic property to hold. Moreover, they argued that only the use of MME results in components that are interpretable in terms of moments. Rayner, Thas and Best (2009) discussed generalised smooth tests that use MME. The components of the test statistic always have useful approximate moment interpretations, but use less efficient parameter estimation when testing for distributions for which MLE and MME do not coincide. Because MLE and MME do not coincide for the GPD, we thus have to choose between efficiency (MLE) and interpretability (MME). Small sample comparisons sometimes show powers for tests based on MME are as good as, or better than, powers for tests based on MLE (see, for example, Thas and Rayner (2005)), so the loss in efficiency does not necessarily lead to worse finite power properties. Moreover, Smith (1984) showed that the MLE  $\hat{\kappa}$  does not have its usual rate of convergence when  $1/2 \leq \kappa < 1$  and that it does not exist with probability asymptotically one when  $\kappa \geq 1$ , so the preferable range of application is also limited and there are thus also existence problems for the MLE based competitor tests. Because of these reasons we will only consider MME based generalised smooth tests with their interpretable components subsequently.

### 3. A generalised smooth test for the GPD

Suppose we wish to test if a random sample  $X_1, \dots, X_n$  comes from a distribution with probability density function  $f(x; \boldsymbol{\beta})$ , where  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of parameters. First define the family of *smooth order  $k$  alternatives* by

$$g_k(x; \boldsymbol{\theta}, \boldsymbol{\beta}) = C(\boldsymbol{\theta}, \boldsymbol{\beta}) \exp \left\{ \sum_{j=1}^k \theta_j h_j(x; \boldsymbol{\beta}) \right\} f(x; \boldsymbol{\beta}). \quad (3.1)$$

Here  $\theta_1, \dots, \theta_k$  are real-valued parameters,  $\boldsymbol{\theta}' = (\theta_1, \dots, \theta_k)$ , and  $C(\boldsymbol{\theta}, \boldsymbol{\beta})$  is a normalising constant (assumed to exist) that ensures that  $g_k(\cdot; \boldsymbol{\theta}, \boldsymbol{\beta})$  integrates to one. The polynomials  $h_j$  are of degree  $j$  and we take  $\{h_j(\cdot; \boldsymbol{\beta})\}_{j=0, \dots, k}$  to form a set of orthonormal polynomials

w.r.t.  $f(\cdot; \boldsymbol{\beta})$ , i.e. satisfying

$$\int_{-\infty}^{+\infty} h_i(x; \boldsymbol{\beta}) h_j(x; \boldsymbol{\beta}) f(x; \boldsymbol{\beta}) dx = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j. \end{cases}$$

These orthonormal polynomials  $h_j$  can be Gram-Schmidt constructed starting from  $h_0(x; \boldsymbol{\beta}) = 1$ , and have the advantage of leading to a test statistic which decomposes naturally if MLE and MME coincide. If we take the family of alternatives (3.1) broad enough by choosing the order  $k$  sufficiently large, testing for  $f(\cdot; \boldsymbol{\beta})$  becomes approximately equivalent to testing  $H : \boldsymbol{\theta} = \mathbf{0}$  against  $K : \boldsymbol{\theta} \neq \mathbf{0}$ . It is also interesting to note that the density in (3.1) is closely related to nonparametric orthogonal series density estimators; see Rayner, Thas and Best (2009, chap.10) for more details. To obtain interpretable test statistics, we use MME based *generalised smooth tests*; see Rayner, Thas and Best (2009, chap.9). Generalised smooth tests are basically generalised score tests (Hall and Mathiason, 1990; Boos, 1992), which are valid for the large class of *asymptotically linear estimators* of the parameters, to which, among others, the MLE and the M-estimators, and thus also the MME belong. We will use the generalised score test for testing  $H : \boldsymbol{\theta} = \mathbf{0}$  against  $K : \boldsymbol{\theta} \neq \mathbf{0}$  in (3.1). Thas and Rayner (2005) also considered MME based generalised smooth tests. A brief description of these tests follows.

The score statistic related to  $\theta_j$  in (3.1) is given by

$$V_j(\boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n h_j(X_i; \boldsymbol{\beta}),$$

for  $j = 1, \dots, k$ . This is the derivative of the log-likelihood function  $\log\{\prod_{i=1}^n g_k(X_i; \boldsymbol{\theta}, \boldsymbol{\beta})\}$  w.r.t.  $\theta_j$ , evaluated in  $\boldsymbol{\theta} = \mathbf{0}$ , and divided by  $\sqrt{n}$ . We denote the MME of  $\boldsymbol{\beta}$  under the null hypothesis as  $\tilde{\boldsymbol{\beta}}$ . Using  $\tilde{\boldsymbol{\beta}}$  for estimating  $\boldsymbol{\beta}$  ensures that the first  $p$   $V_j$ -statistics are equal to zero, because the MME estimation equations can be formulated as  $V_1(\tilde{\boldsymbol{\beta}}) = V_2(\tilde{\boldsymbol{\beta}}) = \dots = V_p(\tilde{\boldsymbol{\beta}}) = 0$ . We therefore remove the redundant first  $p$  terms from the smooth alternatives (3.1). Denoting  $\mathbf{V}'(\boldsymbol{\beta}) = (V_{p+1}(\boldsymbol{\beta}), \dots, V_k(\boldsymbol{\beta}))$  and  $\tilde{\mathbf{V}} = \mathbf{V}(\tilde{\boldsymbol{\beta}})$ , the generalised score test statistic for testing  $H : \boldsymbol{\theta} = \mathbf{0}$  against  $K : \boldsymbol{\theta} \neq \mathbf{0}$  is of the form  $\tilde{S}_k = \tilde{\mathbf{V}}' \tilde{\boldsymbol{\Sigma}}_0^{-1} \tilde{\mathbf{V}}$ , where  $\tilde{\boldsymbol{\Sigma}}_0$  is the MME plug-in estimator of the asymptotic variance covariance matrix of  $\tilde{\mathbf{V}}$  under the null hypothesis. As in Rayner, Thas and Best (2009) and Thas *et al.* (2009), we utilize the Cholesky decomposition of  $\tilde{\boldsymbol{\Sigma}}_0^{-1}$  for obtaining asymptotically independent components. The *Cholesky decomposition* states that if  $\mathbf{A}$  is a symmetric positive definite matrix, there exists an upper triangular matrix  $\mathbf{M}$  such that  $\mathbf{A} = \mathbf{M}\mathbf{M}'$ . Decomposing  $\tilde{\boldsymbol{\Sigma}}_0^{-1}$  into  $\mathbf{M}\mathbf{M}'$  and setting  $\tilde{\mathbf{V}}^* = (\tilde{V}_{p+1}^*, \dots, \tilde{V}_k^*)' = \mathbf{M}'\tilde{\mathbf{V}}$ , the generalised smooth test statistic becomes now

$$\tilde{S}_k = \tilde{\mathbf{V}}' \tilde{\boldsymbol{\Sigma}}_0^{-1} \tilde{\mathbf{V}} = (\mathbf{M}'\tilde{\mathbf{V}})'(\mathbf{M}'\tilde{\mathbf{V}}) = \tilde{\mathbf{V}}^{*t} \tilde{\mathbf{V}}^* = \sum_{r=p+1}^k \tilde{V}_r^{*2}, \tag{3.2}$$

The new components  $\tilde{V}_r^*$  (for  $r = p+1, \dots, k$ ) are, under the null hypothesis, asymptotically independent and standard normal, and  $\tilde{S}_k$  has asymptotically a  $\chi_{k-p}^2$  null distribution.

In fact, as in Dahiya and Gurland (1972), low order moments (from order 1 up to  $p$ ) are used to estimate the parameters (MME), and higher order moments (from order  $p+1$

up to  $k$ ) are used to construct the informative test statistic. Indeed, for  $r = p + 1, \dots, k$ ,  $\tilde{V}_r^*$  is a linear combination of the first  $r - p$  elements of  $\tilde{\mathbf{V}}$ , and thus involves polynomials up to order  $r$ . The expected value of  $\tilde{V}_r^*$  thus involves moments of  $X$  up to the  $r$ -th order and its variance involves moments of  $X$  up to the  $2r$ -th. In that way, the realization of  $\tilde{V}_r^*$  depends on the sample moments up to the  $r$ -th order and, neglecting that its standard deviation is determined under the null hypothesis and thus may be misspecified, the usual moment interpretation of the component tests holds: a significant  $\tilde{V}_r^*$  suggests differences between the sample and hypothesised moments up to  $r$ -th order. For more details we refer to Rayner, Best and Mathews (1995) and Rayner, Thas and Best (2009).

The above general framework can be applied directly to the GPD. For the GPD we have  $\boldsymbol{\beta}^t = (\sigma, \kappa)$  with  $p = 2$ . The orthonormal polynomials  $h_j$  are listed in Appendix A up to fourth order. For the higher order polynomials we rely on the recurrence relations of Rayner, Thas and De Boeck (2008). The matrix  $\boldsymbol{\Sigma}_0$  generally has no convenient simple form and depends on  $\kappa$ , but an explicit formula for  $k = 4$  is provided in Appendix B. We already know from Section 2 that MME-based generalised smooth tests have a limited range of application. Moreover, the dependence of the existence of the moments on the  $\kappa$  parameter further restricts the range of existence. Indeed, the construction of the  $r$ -th orthonormal polynomial requires all moments up to the  $2r$ -th order to exist. So it is necessary that  $\kappa > -1/(2r)$  for the  $r$ -th order orthonormal polynomial to exist. It follows that the order  $k$  generalised smooth test statistic  $\tilde{S}_k$  exists only if the estimated moments up to the  $2k$ -th order exist, or, equivalently, if  $\tilde{\kappa} > -1/(2k)$ . More precisely, the components  $\tilde{V}_r^*$  (for  $r = p + 1, \dots, k$ ) exist if and only if  $\tilde{\kappa} > -1/(2r)$ .

Radouane and Crétois (2002) developed a slightly different smooth test. They used different smooth alternatives of order  $k$ ,

$$g_k(x; \boldsymbol{\theta}, \boldsymbol{\beta}) = C(\boldsymbol{\theta}, \boldsymbol{\beta}) \exp \left\{ \sum_{j=1}^k \theta_j [1 - F_X(x; \boldsymbol{\beta})]^j \right\} f(x; \boldsymbol{\beta}),$$

which involve powers of  $1 - F_X(x; \boldsymbol{\beta})$  rather than orthonormal polynomials of  $x$  on the original measurement scale. Tests of this form, but with  $F_X(x; \boldsymbol{\beta})$  instead of  $1 - F_X(x; \boldsymbol{\beta})$ , were originally studied by Kopecky and Pierce (1979) and Thomas and Pierce (1979). The components of these smooth tests are not interpretable in terms of moments of the GPD.

#### 4. Size and power study

In this section we present the results of a simulation study. We considered the generalised smooth test, making use of the asymptotic null distribution as well as the bootstrap null distribution, the Neyman smooth tests (RC) of Radouane and Crétois (2002) and the Anderson-Darling (AD) and Cramér-von Mises (CvM) tests of Choulakian and Stephens (2001). For these tests the empirical levels are assessed and the empirical powers are compared. First we give some details about the implementation of the tests, and discuss the consequences for the simulation study.

The generalised smooth test statistic  $\tilde{S}_k$  exists only if  $\tilde{\kappa} > -1/(2k)$ , and is thus undefined for samples that lead to a smaller estimated shape parameter  $\tilde{\kappa}$ . We say that samples for which  $\tilde{\kappa} \leq -1/(2k)$  are "infeasible" for the generalised smooth test, because the corresponding test statistic  $\tilde{S}_k$  does not exist. Even if  $\tilde{S}_k$  exists for a certain sample, the moment

existence condition may also cause difficulties in the parametric bootstrap. In particular, the parametric bootstrap procedure involves repeatedly sampling from the fitted GPD, and for every bootstrap sample the parameters have to be estimated again. Although the shape estimate  $\tilde{\kappa}$  of the original sample may lie within the safe parameter region, for the estimated shape estimate of a bootstrap sample this may not necessarily be the case. Nevertheless, as the simulation results will indicate, this can often be approximately solved by just ignoring these "infeasible" bootstrap samples as long as they occur infrequently. This will be true as long as  $\tilde{\kappa}$  stays far enough away from the lower limit  $-1/(2k)$ . We have observed that this almost always happens when  $\tilde{\kappa}$  is positive. Summarizing, the necessary existence of estimated moments up to the  $2k$ -th order restricts the range of existence of the MME based generalised smooth tests to  $\tilde{\kappa} > -1/(2k)$ . And roughly speaking, as a result, the bootstrap based version of these tests is only reliable when  $\tilde{\kappa} > 0$ .

We also consider the MLE based AD, CvM and RC tests. For the maximum likelihood estimators numerical optimisation of the profile log-likelihood is required; see Choulakian and Stephens (2001, sec. 1). The MLE algorithm, which was acquired from Choulakian (personal communication, September 23, 2008), amounts to direct profile log-likelihood maximization using scaled observations obtained by dividing by the sample mean. Smith (1984) showed that the MLE  $\hat{\kappa}$  does not have its usual rate of convergence when  $1/2 \leq \kappa < 1$  and that it does not exist with probability asymptotically one when  $\kappa \geq 1$ , so the preferable range of application is also limited and there are thus also existence problems for the MLE based competitor tests.

Finally, the MME based RC test statistic does not exist when the condition  $\tilde{\kappa}/\tilde{\sigma} < 1/\max\{X_i\}_{1 \leq i \leq n}$  is not fulfilled. This occurs occasionally and unpredictably when  $\tilde{\kappa} > 0$ .

These competitor tests also suffer from existence problems, and we deal with them in the same way as with the generalised smooth tests. An infeasible sample for a certain test is again obtained when the test statistic for this sample does not exist. For all the competitor tests we recommend and use the parametric bootstrap because unrepresented simulation results demonstrated that when these tests are based on their asymptotic null distributions, the empirical sizes are further from the nominal level than for the parametric bootstrap based tests. Again, infeasible bootstrap samples will be ignored. Although Choulakian and Stephens (2001) used asymptotic critical values, the use of the parametric bootstrap for the AD, CvM and RC tests allows a fair comparison with the generalised smooth test. Moreover, because the empirical sizes for the parametric bootstrap based tests are generally larger than for the tests based on the asymptotic null distributions, larger powers are expected.

The number of infeasible Monte Carlo samples for a certain test statistic is always presented between parentheses and it refers to the group of tests immediate to the left of it. They are simply ignored in the calculation of its size and power.

All estimated test sizes and powers are based on 1000 Monte Carlo simulation runs, and all tests are performed at the nominal significance level of  $\alpha = 0.05$ . Sample sizes ranging from 30 to 1000 are considered and the common choice of  $k = 4$  is made. Parametric bootstraps are performed with 1000 bootstrap runs. We also calculated the empirical expected values of  $\tilde{\kappa}$  and  $\hat{\kappa}$  based on the 1000 Monte Carlo samples. These are approximations of  $E[\tilde{\kappa}]$  and  $E[\hat{\kappa}]$  under the alternative indicated. We note that we have altered the MME based RC test statistic  $\tilde{\Psi}_J^2$  from Radouane and Crétois (2002) since we suspect there is an error in its expression (in our opinion, in the second term the matrix should also be inverted and the entire test statistic should be divided by  $n$ ).

First, we have simulated under the null hypothesis in order to assess the bias of the tests. For all tests we have considered bootstrap versions, but we have also included the generalised smooth test based on the asymptotic  $\chi_{k-2}^2$  null distribution. To avoid too many existence problems of the generalised smooth test, this empirical size study is limited to nonnegative values of the shape parameter  $\kappa$ . The results are presented in Table 4.1. The sizes of our bootstrap tests seem satisfactory, except when the number of infeasible Monte Carlo samples is no longer negligible (less than 2% of the number of Monte Carlo runs), as is the case for  $\kappa = 0$  when  $n = 60$  or  $n = 100$ . Conversely, our asymptotic tests have empirical sizes which are too often too far above the nominal level  $\alpha$ , except when  $\kappa = 0$  in which case the empirical sizes of the asymptotic approximation are much smaller than the nominal level. These results indicate that the convergence to the asymptotic null distribution is too slow for the asymptotic approximation to be advocated. We therefore advise to use the bootstrap in most cases. The sizes of the bootstrap based AD, CvM and RC tests seem to be satisfactory.

Second, for the power study we have simulated under the following alternatives:

- The gamma distribution  $\Gamma(\alpha, s)$  with shape parameter  $\alpha > 0$ , scale parameter  $s > 0$  and density  $f_{\Gamma}(x; \alpha, s) = (x/s)^{\alpha-1} \exp(-x/s)/(s\Gamma(\alpha))$  if  $x \geq 0$ .
- The Weibull distribution  $W(\alpha, s)$  with shape parameter  $\alpha > 0$ , scale parameter  $s > 0$  and density  $f_W(x; \alpha, s) = \alpha (x/s)^{\alpha-1} \exp(-(x/s)^\alpha)/s$  if  $x \geq 0$ .

These alternatives are considered because they are well-known competitors of the GPD in modelling extreme value data and it is therefore particularly important to discriminate between these models. They were also considered by Choulakian and Stephens (2001) and Radouane and Crétois (2002). To make a meaningful comparison possible between our test and the MLE based goodness of fit tests, we mainly focussed on alternatives where  $\tilde{\kappa}$  and  $\hat{\kappa}$  are positive but smaller than one. In these cases our test and the MLE based competitors rarely have existence problems. Nevertheless, for some alternatives some tests (often the MME based RC test) frequently suffer from existence problems. We acknowledge that such outcomes with a non-negligible number of infeasible Monte Carlo samples have no statistical meaning, but we have included them anyway to demonstrate that these tests are useless to test against such alternatives.

The results are presented in Tables 4.2 and 4.3, and they show that our test performs better than the other goodness of fit tests when there are no severe existence problems or when  $\tilde{\kappa}$  is sufficiently large. The powers of our tests based on the asymptotic  $\chi^2$  distribution are superior to the corresponding bootstrap powers, but this could be expected and is unfair because the test size is too large. Therefore we further limit the discussion to the bootstrap version. Alternatives  $\Gamma(2, 2/3)$  and  $W(1.5, 2^{-1/1.5})$  as in Radouane and Crétois (2002), and  $\Gamma(2, 1)$  and  $W(1.25, 1)$  as in Choulakian and Stephens (2001) correspond to positive  $E[\tilde{\kappa}]$  and give powers of the MLE based tests which are generally considerably lower than those of the generalised smooth test  $\tilde{S}_4$ . The corresponding powers of the MME based RC test are also inferior to the powers of  $\tilde{S}_4$ , but unreliable as a result of the high number of infeasible Monte Carlo samples. In fact, the MME based RC test is unusable in practice for this shape parameter region. We also have to mention that the powers of the RC tests in Tables 4 and 5 of Radouane and Crétois (2002) are mostly considerably and inexplicably higher than the corresponding empirical powers we obtained for the alternatives  $\Gamma(2, 2/3)$  and



**Table 4.1.** Simulated levels for the generalised smooth tests based on  $\tilde{S}_4$ , and the bootstrap based Anderson-Darling, Cramér-von Mises and Neyman smooth tests. The number of unused infeasible Monte Carlo samples out of 1000 is reported in parentheses referring to the group of tests immediate to the left of it.

| $n$                              | estimated         |                     | asymptotic      |                 |               | bootstrap       |                 |               | AD<br>$A^2$ | CvM<br>$W^2$ | RC<br>$\hat{\Psi}_4^2$ | RC<br>$\tilde{\Psi}_4^2$ |     |       |       |
|----------------------------------|-------------------|---------------------|-----------------|-----------------|---------------|-----------------|-----------------|---------------|-------------|--------------|------------------------|--------------------------|-----|-------|-------|
|                                  | $E[\hat{\kappa}]$ | $E[\tilde{\kappa}]$ | $\tilde{V}_3^*$ | $\tilde{V}_4^*$ | $\tilde{S}_4$ | $\tilde{V}_3^*$ | $\tilde{V}_4^*$ | $\tilde{S}_4$ |             |              |                        |                          |     |       |       |
| GPD with $\beta^t = (3, 0.4)$    |                   |                     |                 |                 |               |                 |                 |               |             |              |                        |                          |     |       |       |
| 30                               | 0.46              | 0.53                | 0.055           | 0.083           | 0.087         | 0.055           | 0.051           | 0.044         | (0)         | 0.044        | 0.039                  | 0.056                    | (0) | 0.057 | (172) |
| 60                               | 0.43              | 0.47                | 0.046           | 0.102           | 0.100         | 0.065           | 0.059           | 0.057         | (0)         | 0.051        | 0.046                  | 0.054                    | (0) | 0.055 | (186) |
| 100                              | 0.42              | 0.44                | 0.035           | 0.097           | 0.078         | 0.048           | 0.070           | 0.060         | (0)         | 0.030        | 0.034                  | 0.045                    | (0) | 0.043 | (189) |
| 1000                             | 0.40              | 0.41                | 0.053           | 0.062           | 0.059         | 0.051           | 0.057           | 0.055         | (0)         | 0.054        | 0.058                  | 0.063                    | (0) | 0.055 | (171) |
| GPD with $\beta^t = (3, 0.15)$   |                   |                     |                 |                 |               |                 |                 |               |             |              |                        |                          |     |       |       |
| 30                               | 0.21              | 0.25                | 0.037           | 0.061           | 0.061         | 0.040           | 0.045           | 0.051         | (15)        | 0.042        | 0.048                  | 0.048                    | (0) | 0.045 | (72)  |
| 60                               | 0.18              | 0.20                | 0.040           | 0.043           | 0.050         | 0.047           | 0.036           | 0.039         | (1)         | 0.041        | 0.038                  | 0.055                    | (0) | 0.051 | (30)  |
| 100                              | 0.17              | 0.18                | 0.028           | 0.030           | 0.035         | 0.050           | 0.036           | 0.032         | (0)         | 0.043        | 0.046                  | 0.048                    | (0) | 0.040 | (17)  |
| 1000                             | 0.15              | 0.15                | 0.042           | 0.069           | 0.059         | 0.049           | 0.061           | 0.054         | (0)         | 0.049        | 0.042                  | 0.049                    | (0) | 0.063 | (1)   |
| GPD with $\beta^t = (1/1.42, 0)$ |                   |                     |                 |                 |               |                 |                 |               |             |              |                        |                          |     |       |       |
| 60                               | 0.05              | 0.05                | 0.024           | 0.017           | 0.025         | 0.030           | 0.030           | 0.027         | (69)        | 0.030        | 0.034                  | 0.055                    | (0) | 0.055 | (8)   |
| 100                              | 0.02              | 0.02                | 0.016           | 0.020           | 0.022         | 0.034           | 0.028           | 0.025         | (55)        | 0.051        | 0.048                  | 0.054                    | (0) | 0.045 | (2)   |
| 1000                             | 0.00              | 0.00                | 0.020           | 0.013           | 0.022         | 0.041           | 0.039           | 0.035         | (0)         | 0.049        | 0.046                  | 0.051                    | (0) | 0.049 | (0)   |

**Table 4.2.** Simulated powers under gamma alternatives for the generalised smooth test based on  $\tilde{S}_4$ , and the bootstrap based Anderson-Darling, Cramér-von Mises and Neyman smooth tests. The number of unused infeasible Monte Carlo samples out of 1000 is reported in parentheses referring to the group of tests immediate to the left of it.

| $n$              | estimated         |                     | asymptotic      |                 |               | bootstrap       |                 |               | AD<br>$A^2$ | CvM<br>$W^2$ | RC<br>$\hat{\Psi}_4^2$ | RC<br>$\tilde{\Psi}_4^2$ |     |       |       |
|------------------|-------------------|---------------------|-----------------|-----------------|---------------|-----------------|-----------------|---------------|-------------|--------------|------------------------|--------------------------|-----|-------|-------|
|                  | $E[\hat{\kappa}]$ | $E[\tilde{\kappa}]$ | $\tilde{V}_3^*$ | $\tilde{V}_4^*$ | $\tilde{S}_4$ | $\tilde{V}_3^*$ | $\tilde{V}_4^*$ | $\tilde{S}_4$ |             |              |                        |                          |     |       |       |
| $\Gamma(2, 2/3)$ |                   |                     |                 |                 |               |                 |                 |               |             |              |                        |                          |     |       |       |
| 60               | 0.57              | 0.36                | 0.801           | 0.575           | 0.796         | 0.740           | 0.511           | 0.694         | (0)         | 0.626        | 0.565                  | 0.552                    | (0) | 0.562 | (749) |
| 100              | 0.54              | 0.32                | 0.953           | 0.729           | 0.953         | 0.926           | 0.670           | 0.898         | (0)         | 0.906        | 0.855                  | 0.849                    | (0) | 0.820 | (861) |
| $\Gamma(2, 1)$   |                   |                     |                 |                 |               |                 |                 |               |             |              |                        |                          |     |       |       |
| 30               | 0.64              | 0.46                | 0.508           | 0.403           | 0.525         | 0.390           | 0.324           | 0.382         | (2)         | 0.183        | 0.194                  | 0.217                    | (0) | 0.252 | (595) |
| 50               | 0.58              | 0.38                | 0.721           | 0.546           | 0.716         | 0.628           | 0.473           | 0.588         | (0)         | 0.467        | 0.438                  | 0.408                    | (0) | 0.419 | (697) |
| 100              | 0.55              | 0.32                | 0.957           | 0.723           | 0.937         | 0.926           | 0.647           | 0.902         | (0)         | 0.895        | 0.840                  | 0.850                    | (0) | 0.786 | (869) |

$W(1.5, 2^{-1/1.5})$ . However, the powers of the AD test in Table 6 of Choulakian and Stephens (2001) are very similar to the corresponding bootstrap based powers we obtained for the alternatives  $\Gamma(2, 1)$  and  $W(1.25, 1)$ . We note that the test  $\tilde{V}_3^*$  performs a little better than  $\tilde{S}_4$ , and both are superior to  $\tilde{V}_4^*$ . Also, the powers of  $A^2$  are superior to those of  $W^2$  and  $\hat{\Psi}_4^2$ , which have similar performance.

Finally, we have also included the results obtained from the  $W(0.75, 1)$  alternative as in Choulakian and Stephens (2001) to illustrate the downside of our approach when  $\tilde{\kappa}$  is often negative. In contrast to all competitors, our tests are useless and their powers are unreliable because of the large number of infeasible Monte Carlo samples. It is clear that the powers of  $A^2$  and  $\tilde{\Psi}_4^2$  are similar and both superior to the powers of  $W^2$  and  $\hat{\Psi}_4^2$ . However, when choosing between  $A^2$  and  $\tilde{\Psi}_4^2$  we prefer the AD test for testing against alternatives that imply negative shape parameter estimates, because the MME based RC test assumes  $\kappa > -1/2$ .

**Table 4.3.** Simulated powers under Weibull alternatives for the generalised smooth test based on  $\tilde{S}_4$ , and the bootstrap based Anderson-Darling, Cramér-von Mises and Neyman smooth tests. The number of unused infeasible Monte Carlo samples out of 1000 is reported in parentheses referring to the group of tests immediate to the left of it.

| $n$                  | estimated         |                   | asymptotic      |                 |               | bootstrap       |                 |               | AD<br>$A^2$ | CvM<br>$W^2$ | RC<br>$\hat{\Psi}_4^2$ | RC<br>$\tilde{\Psi}_4^2$ |     |       |       |
|----------------------|-------------------|-------------------|-----------------|-----------------|---------------|-----------------|-----------------|---------------|-------------|--------------|------------------------|--------------------------|-----|-------|-------|
|                      | $E[\hat{\kappa}]$ | $E[\hat{\kappa}]$ | $\tilde{V}_3^*$ | $\tilde{V}_4^*$ | $\tilde{S}_4$ | $\tilde{V}_3^*$ | $\tilde{V}_4^*$ | $\tilde{S}_4$ |             |              |                        |                          |     |       |       |
| $W(1.5, 2^{-1/1.5})$ |                   |                   |                 |                 |               |                 |                 |               |             |              |                        |                          |     |       |       |
| 60                   | 0.65              | 0.44              | 0.651           | 0.521           | 0.659         | 0.557           | 0.413           | 0.564         | (0)         | 0.370        | 0.356                  | 0.322                    | (0) | 0.279 | (792) |
| 100                  | 0.62              | 0.39              | 0.872           | 0.704           | 0.867         | 0.839           | 0.630           | 0.823         | (0)         | 0.733        | 0.692                  | 0.639                    | (0) | 0.403 | (923) |
| $W(1.25, 1)$         |                   |                   |                 |                 |               |                 |                 |               |             |              |                        |                          |     |       |       |
| 30                   | 0.38              | 0.36              | 0.124           | 0.127           | 0.147         | 0.084           | 0.117           | 0.099         | (2)         | 0.037        | 0.046                  | 0.060                    | (0) | 0.081 | (198) |
| 50                   | 0.33              | 0.28              | 0.225           | 0.198           | 0.260         | 0.162           | 0.183           | 0.193         | (0)         | 0.112        | 0.119                  | 0.096                    | (0) | 0.131 | (236) |
| 100                  | 0.30              | 0.24              | 0.380           | 0.299           | 0.404         | 0.321           | 0.266           | 0.346         | (0)         | 0.268        | 0.253                  | 0.223                    | (0) | 0.227 | (299) |
| $W(0.75, 1)$         |                   |                   |                 |                 |               |                 |                 |               |             |              |                        |                          |     |       |       |
| 30                   | -0.14             | -0.31             | 0.000           | 0.017           | 0.010         | 0.034           | 0.017           | 0.015         | (593)       | 0.180        | 0.102                  | 0.140                    | (0) | 0.191 | (0)   |
| 50                   | -0.17             | -0.35             | 0.000           | 0.017           | 0.017         | 0.031           | 0.035           | 0.035         | (712)       | 0.264        | 0.165                  | 0.213                    | (0) | 0.274 | (0)   |
| 100                  | -0.20             | -0.38             | 0.000           | 0.000           | 0.000         | 0.014           | 0.042           | 0.021         | (856)       | 0.515        | 0.358                  | 0.379                    | (0) | 0.500 | (0)   |

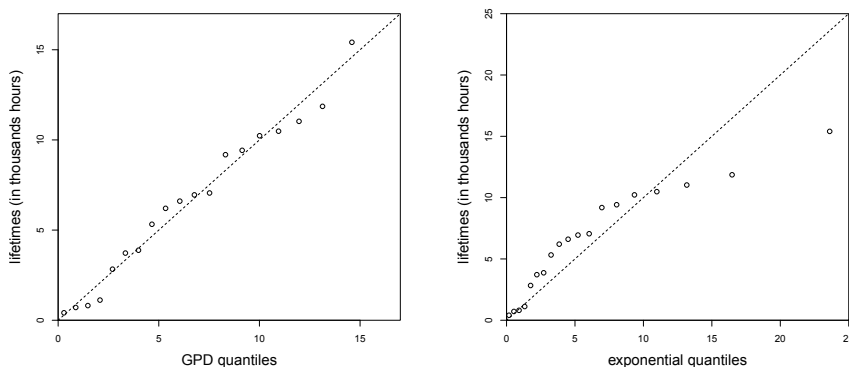
## 5. Examples

**Example 5.1 (Operational life time data).** Rayner, Thas and Best (2009, chap. 6) have analysed the operational life time data from Angus (1982) in their Example 6.3.1. They concluded that these data do not follow an exponential distribution. Here we reanalyse the data, but now we test the null hypothesis that the lifetimes come from a GPD, with threshold equal to the smallest observation, which is 2398 hours, and with this smallest observation removed from the data.

The parameter estimates are  $\hat{\kappa} = 0.69$  and  $\hat{\sigma} = 10972$ . We now find  $\tilde{S}_4 = 1.61$  with  $p$ -value 0.42. The Cholesky decomposition has been applied so that component  $\tilde{V}_3^*$  is only related to the third order moment, and  $\tilde{V}_4^*$  is a linear combination of  $\tilde{V}_3$  and  $\tilde{V}_4$ . This gives  $\tilde{V}_3^* = -0.33$  and  $\tilde{V}_4^* = 1.22$  with  $p$ -values equal to 0.692 and 0.458, respectively. This analysis suggests that the lifetimes may be well described by a GPD. Figure 5.1 shows a Q-Q plot. It is informative to compare this graph with a Q-Q plot of the same data for the exponential distribution (Figure 5.1).

The same conclusion is obtained with other tests. For example, the Anderson-Darling test of Choulakian and Stephens (2001) gives  $A^2 = 0.30$  with a  $p$ -value of at least 0.50, using their Table 2. This test uses maximum likelihood for the estimation of the parameters:  $\hat{\kappa} = 0.70$  and  $\hat{\sigma} = 11149$ . However, since  $0.5 \leq \hat{\kappa} < 1$ , the problem is referred to as *nonregular* according to Smith (1984), because the rate of convergence of the estimator is altered in this part of the parameter space. Order four Neyman smooth tests of Radouane and Crétois (2002) have also been applied to the data, both with the use of MLE and MME. With MLE we find  $\hat{\Psi}_4^2 = 1.52$  with  $p$ -value 0.824, and with MME we find  $\tilde{\Psi}_4^2 = 1.77$  with  $p$ -value 0.777. These tests again support the conclusion that the operational life times may be well described by a generalised Pareto distribution.

**Example 5.2 (Annual maximum December temperatures at Fair Isle).** Dupuis and Tsao (1998) fit a generalised extreme value distribution to 22 annual maximum December temperatures at Fair Isle Weather Station, UK for 1974 to 1995. The data, in degrees Celsius,



**Figure 5.1.** Generalised Pareto distribution Q-Q plot (left) and exponential Q-Q plot (right) of the Angus data.

are

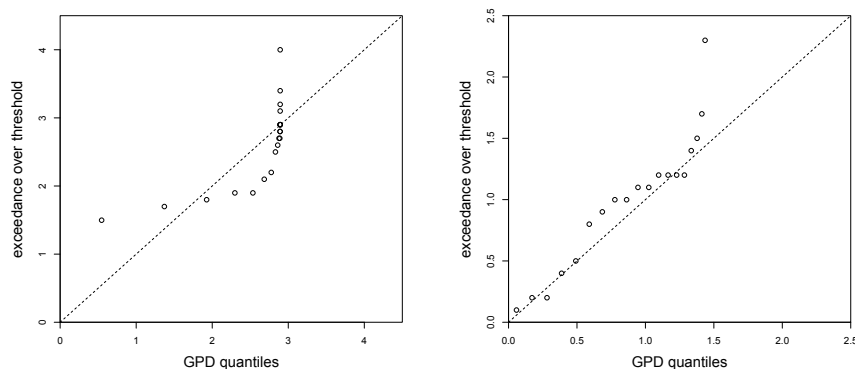
|      |      |      |      |     |      |      |      |      |      |       |
|------|------|------|------|-----|------|------|------|------|------|-------|
| 10.3 | 10.2 | 9.0  | 10.9 | 7.5 | 10.7 | 10.0 | 11.5 | 9.2  | 10.2 | 10.4  |
| 10.6 | 9.6  | 10.4 | 9.7  | 9.4 | 9.4  | 9.3  | 10.4 | 10.3 | 10.4 | 10.1. |

We proceed as Choulakian and Stephens (2001) by testing for goodness of fit for the GPD on the complete data set, setting the threshold at the smallest observed data point while removing this observation from the data, and later progressively increasing the threshold by removing the smallest observations one by one until the test gives an insignificant result. At most three consecutive observations are deleted. We use a generalised smooth test of order four, and since we use MME for the estimation of the two parameters, the first two components are zero.

Table 5.1 reports the parameter estimates,  $\tilde{S}_4$ ,  $\tilde{V}_3^*$  and  $\tilde{V}_4^*$  and their  $p$ -values for the Fair Isle data with thresholds 7.5, 9.0, and 9.2. For all three thresholds the generalised smooth tests based on  $\tilde{S}_4$  reject the null hypothesis, but for the two larger thresholds of 9.0 and 9.2 the rejection is less pronounced. Figure 5.2 shows two Q-Q plots, for the thresholds 7.5 and 9.2. The fit of the GPD is clearly better with the larger threshold, but still not good enough for the null hypothesis to be accepted by the smooth test.

The analyses seem to suggest that the skewnesses of the data sets with the larger thresholds are close to what is expected for a GPD, but that the fourth order moment, and perhaps also some higher order moments do not agree. Thus, the tail of the data does not behave as described by the GPD. Since the GPD is often used in risk analysis, in which the calculation of upper quantiles of the distribution is very important, we should not recommend using the GPD for further risk calculation with these data.

None of the other tests for goodness of fit that we discussed before could be applied to this data set. The MLE based tests of Choulakian and Stephens (2001) and Radouane and Crétois (2002) could not be used, because the MLE of  $\kappa$  does not exist for these data, which is suggested by  $\tilde{\kappa} > 1$ . The MME based Neyman smooth test of Radouane and Crétois (2002) could not be applied either. The reason is that the test statistic involves a logarithmic function which requires a positive argument. This is only guaranteed when  $\tilde{\kappa}/\tilde{\sigma} < 1/m$ , with  $m$  the largest sample observation after subtraction of the threshold, and this is not the



**Figure 5.2.** Generalised Pareto Distribution Q-Q plots of the Fair Isle data with thresholds 7.5 (left) and 9.2 (right).

case for this example.

We conclude this section with a remark on the potential inconsistency between the range of the sample observations and the parameter estimates. In both examples presented here the ratio  $\hat{\sigma}/\hat{\kappa}$ , which is for positive shape parameter the estimated maximum value of the support of the GPD, is smaller than the largest sample observation. The sample observations are thus not consistent with the estimated GPD. However, this does not prove that the sample did not arise from a GPD. The reason is that  $\hat{\sigma}/\hat{\kappa}$  is only an estimate of the true upper bound  $\sigma/\kappa$ . It is thus still meaningful to apply a goodness of fit test to formally test the distributional assumption.

**Table 5.1.** Results for the Fair Isle data.  $p$ -values are reported in parentheses.

|                 | threshold  |               |               |
|-----------------|------------|---------------|---------------|
|                 | 7.5        | 9.0           | 9.2           |
| $\hat{\kappa}$  | 8.65       | 1.54          | 1.22          |
| $\hat{\sigma}$  | 25.04      | 2.93          | 2.22          |
| $\tilde{S}_4$   | 110.06 (0) | 28.11 (0.013) | 27.60 (0.011) |
| $\tilde{V}_3^*$ | 3.44 (0)   | 1.73 (0.174)  | 1.21 (0.326)  |
| $\tilde{V}_4^*$ | 9.91 (0)   | 5.01 (0.012)  | 5.11 (0.020)  |

## 6. Discussion

In this paper we have constructed new smooth tests of goodness of fit for the GPD. Because the use of asymptotic critical values sometimes gives deviant test sizes for the generalised smooth test, we recommend the parametric bootstrap for critical value and  $p$ -value calculations. An advantage of these goodness of fit tests compared to other tests of fit is that the generalised smooth test statistic is decomposable into interpretable components. The most serious drawback is the existence problem. The test statistic  $\tilde{S}_k$  only exists if  $\hat{\kappa} > -1/(2k)$ , which also gives problems in the parametric bootstrap. However, while this

is undeniably a disadvantage, it was worth the effort to construct the generalised smooth test in addition to the known tests of goodness of fit. After all, it brings the interpretability, and, moreover, competitor tests are often challenged too by existence problems. We suggest first estimating  $\tilde{\kappa}$ , and if  $\tilde{\kappa} > 0$  we recommend our bootstrap based generalised smooth test with its good power and interpretable components. Otherwise, this test is not available and we advise the MLE based Anderson-Darling test. This means that for the GPD the choice of test is driven by the value of the estimated shape parameter  $\tilde{\kappa}$ . Our tests are implemented in the smoothest R package which may be downloaded from the website, <http://biomath.ugent.be/~othas/smooth2>.

## Appendices

### A. Orthonormal polynomials

The first 5 orthonormal polynomials w.r.t. the GPD are  $h_0(\cdot) = 1$  and  $h_j(\cdot; \boldsymbol{\beta})$  ( $j = 1, 2, 3, 4$ ) given by:

$$\begin{aligned} h_1(x; \sigma, \kappa) &= \frac{\sqrt{1+2\kappa}}{\sigma} [(1+\kappa)x + \sigma] \\ h_2(x; \sigma, \kappa) &= \frac{\sqrt{1+4\kappa}}{2\sigma^2} [(1+2\kappa)(1+3\kappa)x^2 - 4(1+2\kappa)\sigma x + 2\sigma^2] \\ h_3(x; \sigma, \kappa) &= \frac{\sqrt{1+6\kappa}}{6\sigma^3} [(1+3\kappa)(1+4\kappa)(1+5\kappa)x^3 - 9(1+3\kappa)(1+4\kappa)\sigma x^2 \\ &\quad + 18(1+3\kappa)\sigma^2 x - 6\sigma^3] \\ h_4(x; \sigma, \kappa) &= \frac{\sqrt{1+8\kappa}}{24\sigma^4} [(1+4\kappa)(1+5\kappa)(1+6\kappa)(1+7\kappa)x^4 \\ &\quad - 16(1+4\kappa)(1+5\kappa)(1+6\kappa)\sigma x^3 \\ &\quad + 72(1+4\kappa)(1+5\kappa)\sigma^2 x^2 - 96(1+4\kappa)\sigma^3 x + 24\sigma^4] \end{aligned}$$

As is clear by their expression, the  $j$ -th order polynomial exists if and only if  $\kappa > -1/(2j)$ .

### B. Variance covariance matrix

$\tilde{\boldsymbol{\Sigma}}_0$  is obtained by plugging the MME parameter estimate  $\tilde{\kappa}$  into the asymptotic variance covariance matrix of  $\tilde{\mathbf{V}}$  under  $H$ , say  $\boldsymbol{\Sigma}_0$ . The elements of the matrix  $\boldsymbol{\Sigma}_0$  for  $k = 4$  are

$$\begin{aligned} (\boldsymbol{\Sigma}_0)_{11} &= \frac{(1+4\kappa)(1+5\kappa)(72\kappa^4+78\kappa^3+35\kappa^2+6\kappa+1)}{(1+3\kappa)^3(1+2\kappa)^3} \\ (\boldsymbol{\Sigma}_0)_{12} &= \frac{24(1+5\kappa)(10\kappa^2+9\kappa+3)\kappa^3\sqrt{1+6\kappa}\sqrt{1+8\kappa}}{(1+3\kappa)^3(1+2\kappa)^3(1+4\kappa)} \\ (\boldsymbol{\Sigma}_0)_{21} &= (\boldsymbol{\Sigma}_0)_{12} \\ (\boldsymbol{\Sigma}_0)_{22} &= \frac{(1+5\kappa)(1+6\kappa)(8064\kappa^7+8976\kappa^6+4816\kappa^5+1456\kappa^4+460\kappa^3+115\kappa^2+16\kappa+1)}{(1+2\kappa)^3(1+3\kappa)^3(1+4\kappa)^3}. \end{aligned}$$

### Acknowledgements

Research supported by IUAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy). The authors also wish to thank Ian Robinson and Vartan Choulakian for their helpful contributions, and the reviewers for their insightful comments.

### References

- Angus, J.E., 1982. Goodness-of-fit tests for exponentiality based on loss-of-memory type functional equation. *Journal of Statistical Planning and Inference*, 6(3), 241–251.
- Boos, D.D., 1992. On generalized score tests. *The American Statistician*, 46(4), 327–333.
- Castillo, E., Hadi, A.S., 1997. Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association*, 92(440), 1609–1620.
- Choulakian, V., Stephens, M.A., 2001. Goodness-of-fit for the generalised Pareto distribution. *Technometrics*, 43(4), 478–484.
- Dahiya, R.C., Gurland, J., 1972. Goodness of fit tests for the gamma and exponential distributions. *Technometrics*, 14(3), 791–801.
- Dupuis, D.J., Tsao, M., 1998. A hybrid estimator for generalized Pareto and extreme-value distributions. *Communications in Statistics - Theory and Methods*, 27(4), 925–941.
- Grimshaw, S.D., 1993. Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics*, 35(2), 185–191.
- Hall, W.J., Mathiason, D.J., 1990. On large-sample estimation and testing in parametric models. *International Statistical Review*, 58, 77–97.
- Henze, N., Klar, B., 1996. Properly rescaled components of smooth tests of fit are diagnostic. *Australian Journal of Statistics*, 38, 61–74.
- Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29(3), 339–349.
- Klar, B., 2000. Diagnostic smooth tests of fit. *Metrika*, 52, 237–252.
- Kopecky, K.J., Pierce, D.A., 1979. Efficiency of smooth goodness-of-fit tests. *Journal of the American Statistical Association*, 74(366), 393–397.
- Pickands, J., 1975. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3, 119–131.
- Radouane, O., Crétois, E., 2002. Neyman smooth tests for the generalized Pareto distribution. *Communications in Statistics - Theory and Methods*, 31(7), 1067–1078.
- Rayner, J.C.W., Best, D.J., 1989. *Smooth Tests of Goodness of Fit*. Oxford University Press, New York.
- Rayner, J.C.W., Best, D.J., Mathews, K.L., 1995. Interpreting the skewness coefficient. *Communications in Statistics - Theory and Methods*, 24, 593–600.
- Rayner, J.C.W., Thas, O., Best, D.J., 2009. *Smooth Tests of Goodness of Fit, Using R, 2nd Edition*. Wiley.
- Rayner, J.C.W., Thas, O., De Boeck, B., 2008. A generalised Emerson recurrence relation. *Australian and New Zealand Journal of Statistics*, 50(3), 235–240.
- Smith, R., 1984. Threshold methods for sample extremes. In *Statistical Extremes and Applications*, Tiago de Oliveira J., (Editor), pages 621–635, Reidel, D., Dordrecht.
- Thas, O., Rayner, J.C.W., 2005. Smooth tests for the zero-inflated Poisson distribution. *Biometrics*, 61(3), 808–815.
- Thas, O., Rayner, J.C.W., Best, D.J., De Boeck, B., 2009. Informative statistical analyses using smooth goodness of fit tests. *Journal of Statistical Theory and Practice*, 3(3), 705–733.
- Thomas, D.R., Pierce, D.A., 1979. Neyman's smooth goodness-of-fit test when the hypothesis is composite. *Journal of the American Statistical Association*, 74(366), 441–445.
- Zhang, J., 2007. Likelihood moment estimation for the generalized Pareto distribution. *Australian and New Zealand Journal of Statistics*, 49(1), 69–77.