Supplementary material to “A new hybrid estimation method for the generalized Pareto distribution” by C. Wang and G. Chen

Convergence Problem associated with the Maximum Likelihood and the Maximum Goodness-of-fit Methods

It is well known that severe convergence problem may be encountered when computing the maximum likelihood (ML) estimators and the maximum goodness-of-fit (MGF) estimators for the generalized Pareto distribution. However, most readers will assume this is for $k > 0.5$ where the regularity conditions are not satisfied. We will show some simulation evidence that occurrence of the convergence problem may depend on the effective sample size and even when k is in the regular range.

First, we use the algorithm provided by Professor Grimshaw for computing the ML estimates (?) and carry out a simulation study with 10,000 replications. We use the criterion to determine a convergence problem if the algorithm returned “MLE does not converge”. The default initial value used in the algorithm is

$$\theta_L = 2(X_{(1)} - \bar{X}) / (X_{(1)}^2) \quad \text{or} \quad \theta_U = 1/X_{(n)} - \varepsilon \quad \text{with } \varepsilon = 10^{-6}/\bar{X}.$$

The convergence problem encountered is summarized in the following Table 1.

Table 1: Number of times a convergence problem occurred when computing the ML estimates based on 10,000 replications, and $\sigma = 1$ is fixed.

n	$k = -3$	$k = -2$	$k = -1$	$k = -0.2$	$k = 0.2$	$k = 0.4$	$k = 0.9$	$k = 1.2$
10	96	315	1310	4501	7680	8949	9889	9973
20	5	0	26	449	1998	4159	9464	9924
50	8	0	0	1	13	107	7619	9914
100	14	1	0	0	0	0	5386	9953
200	29	0	0	0	0	0	2719	9986
500	50	0	0	0	0	0	395	10000

The results in Table 1 are in agreement with the conclusions of ? and ?. In many real cases where $k > 0$ and $n < 25$, the ML method of GPD may have convergence problem, and the situation of no maximum can occur when k approaches and beyond 0.5. However, from the above table we can also find that in some very heavy-tailed cases, that is, $k < -2$, some extremely large observations may significantly alter the convergence of the algorithm even though the sample size is large. To overcome this computational problem in such cases, a carefully selected initial point $\theta^{(0)}$ is recommended.

Next, we examine the computational issue associated with the MGF estimators. In particular, we discuss the MGF method based on the classic Cramér-von Mises (CM) statistic, the

Anderson-Darling (AD) statistic and also the right-sided and left-sided Anderson-Darling (ADR, ADL) statistics introduced in ?. In terms of the GPD with cdf $F(x; \sigma, k)$, the definition of these four EDF statistics are given in Luceño (2006). The search for the MGF estimators requires two-dimensional numerical optimization which may cause convergence problems in some cases. We use the criterion to determine a convergence problem if the output value of `convergence` is other than “`successful`”. The following Table 2 summarizes the convergence problem for computing the MGF estimates using the R function `gpdmgf` in the package `POT`. The default starting point used in the algorithm is $(\sigma^{(0)}, k^{(0)}) = (\bar{X}, 0)$. The simulation study is based on 10,000 replications and the tolerance level is set to be $10^{-6}/\bar{X}$.

Table 2: Number of times a convergence problem occurred when computing the MGF estimates using the R function `gpdmgf` in the package `POT` with the default starting values based on 10,000 replications, and $\sigma = 1$ is fixed.

Method	n	$k = -3$	$k = -2$	$k = -1$	$k = -0.2$	$k = 0.2$	$k = 0.4$	$k = 0.9$	$k = 1.2$
CM	20	6087	4269	495	0	0	0	0	10
	50	7759	5692	354	0	0	0	0	0
	100	8484	6741	239	0	0	0	0	0
	200	8676	7593	156	0	0	0	0	0
AD	20	3581	1461	40	0	0	0	3	212
	50	7517	4029	718	0	0	0	0	53
	100	9539	7973	3019	3	0	0	0	16
	200	9990	9782	6237	8	0	0	0	5
ADR	20	3978	2455	158	0	0	0	1	322
	50	7820	4730	722	0	0	0	3	128
	100	9727	8099	3019	3	0	0	0	50
	200	9993	9783	6237	8	0	0	0	11
ADL	20	7259	4198	295	0	0	0	0	0
	50	8636	4889	143	0	0	0	0	0
	100	9077	5327	76	0	0	0	0	0
	200	9116	5705	42	0	0	0	0	0

According to the results in Table 2, the MGF method generally works well in the cases when k approaches and beyond 0.5, where the ML method arises problems. But in contrast with the ML method, the MGF method can not successfully deal with the very heavy-tailed cases, e.g., $k < -2$, and it is relatively slow as expected. In such case, a well specified starting point $(\sigma^{(0)}, k^{(0)})$ could be critical.

Finite-Sample Performances of Different Estimation Methods

The following Tables 3 and 4 summarize the performances of several estimation methods for the GPD parameters that have been frequently referred in the literature, together with the proposed new hybrid estimation method based on the ordinary Anderson-Darling statistic and the right-sided Anderson-Darling statistic as discussed in Luceño (2006).

Table 3: Summary of the finite-sample performances for different estimation methods based on 10,000 times simulation with sample size $n = 50$, and $\sigma = 1$ fixed.

Methods	$\hat{\sigma}$											
	MOM	PWM	MLE	LME	EBM	EBM*	CM	AD	ADR	ADL	HYB _{AD}	HYB _{ADR}
$k = -5$												
Bias(σ)	Inf	Inf	0.1787	0.4062	0.8552	0.2163	0.2061	0.1398	0.2309	0.1442	0.1258	0.1801
RMSE(σ)	(Inf)	(Inf)	(0.6747)	(1.1650)	(1.4204)	(0.7047)	(0.7735)	(0.6551)	(0.8933)	(0.6502)	(0.6489)	(0.8504)
$k = -3$												
Bias(σ)	Inf	Inf	0.1166	0.2022	0.2309	0.1267	0.1197	0.0815	0.1207	0.0905	0.0688	0.0880
RMSE(σ)	(Inf)	(Inf)	(0.4945)	(0.6418)	(0.5694)	(0.4978)	(0.5190)	(0.4730)	(0.5518)	(0.4901)	(0.4657)	(0.5309)
$k = -1$												
Bias(σ)	7.1646	0.2704	0.0667	0.0770	0.0581	0.0408	0.0570	0.0317	0.0446	0.0452	0.0217	0.0271
RMSE(σ)	(196.93)	(0.4570)	(0.3252)	(0.3282)	(0.3095)	(0.3078)	(0.3364)	(0.3115)	(0.3178)	(0.3505)	(0.3057)	(0.3114)
$k = -0.5$												
Bias(σ)	0.3919	0.0503	0.0607	0.0557	0.0165	0.0184	0.0450	0.0195	0.0294	0.0358	0.0109	0.0154
RMSE(σ)	(0.6334)	(0.2562)	(0.2835)	(0.2720)	(0.2544)	(0.2587)	(0.2985)	(0.2716)	(0.2691)	(0.3189)	(0.2672)	(0.2657)
$k = 0.2$												
Bias(σ)	0.0290	0.0125	0.0717	0.0317	-0.0236	-0.0181	0.0319	0.0029	0.0116	0.0250	-0.0036	0.0018
RMSE(σ)	(0.2097)	(0.2269)	(0.2338)	(0.2101)	(0.1945)	(0.1889)	(0.2523)	(0.2181)	(0.2092)	(0.2779)	(0.2162)	(0.2088)
$k = 1.5$												
Bias(σ)	0.0362	0.0203	NA	0.0077	-0.0756	-0.1969	0.0236	0.0005	0.0066	0.0257	-0.0123	-0.0060
RMSE(σ)	(0.2776)	(0.2515)	(NA)	(0.1645)	(0.1457)	(0.2176)	(0.2033)	(0.1739)	(0.1630)	(0.2379)	(0.1697)	(0.1611)
$k = -5$												
Bias(k)	4.5146	4.0097	0.0607	0.1277	0.4896	0.0922	0.0409	-0.0402	-0.0126	-0.0038	0.0090	0.0066
RMSE(k)	(4.5146)	(4.0097)	(0.8589)	(0.9634)	(0.8331)	(0.8570)	(0.9629)	(0.9006)	(0.9064)	(1.0474)	(0.8640)	(0.9063)
$k = -3$												
Bias(k)	2.5202	2.0266	0.0490	0.0895	0.1493	0.0591	0.0326	-0.0268	-0.0056	-0.0017	0.0047	0.0054
RMSE(k)	(2.5203)	(2.0269)	(0.5754)	(0.6222)	(0.5533)	(0.5677)	(0.6565)	(0.6056)	(0.5995)	(0.7392)	(0.5746)	(0.5968)
$k = -1$												
Bias(k)	0.5950	0.2818	0.0440	0.0538	0.0387	0.0205	0.0264	-0.0171	-0.0004	0.0009	-0.0009	0.0042
RMSE(k)	(0.5996)	(0.3234)	(0.2981)	(0.2953)	(0.2778)	(0.2878)	(0.3820)	(0.3260)	(0.3044)	(0.4679)	(0.2977)	(0.2968)
$k = -0.5$												
Bias(k)	0.2243	0.0742	0.0482	0.0453	0.0076	0.0087	0.0255	-0.0174	-0.0010	0.0027	-0.0035	0.0030
RMSE(k)	(0.2542)	(0.1934)	(0.2348)	(0.2237)	(0.2131)	(0.2196)	(0.3282)	(0.2633)	(0.2378)	(0.4157)	(0.2373)	(0.2298)
$k = 0.2$												
Bias(k)	0.0293	0.0090	0.0714	0.0323	-0.0245	-0.0174	0.0266	-0.0215	-0.0036	0.0073	-0.0098	-0.0002
RMSE(k)	(0.1567)	(0.1872)	(0.1750)	(0.1574)	(0.1493)	(0.1409)	(0.2794)	(0.1985)	(0.1702)	(0.3655)	(0.1810)	(0.1650)
$k = 1.5$												
Bias(k)	0.0615	0.0238	NA	0.0121	-0.1247	-0.3565	0.0432	-0.0053	0.0083	0.0525	-0.0219	-0.0093
RMSE(k)	(0.5137)	(0.4517)	(NA)	(0.2516)	(0.2263)	(0.3818)	(0.3426)	(0.2751)	(0.2513)	(0.4455)	(0.2654)	(0.2474)

Table 4: Summary of the finite-sample performances for different estimation methods based on 10,000 times simulation with sample size $n = 200$, and $\sigma = 1$ fixed.

Methods	$\hat{\sigma}$											
	MOM	PWM	MLE	LME	EBM	EBM*	CM	AD	ADR	ADL	HYB _{AD}	HYB _{ADR}
$k = -5$												
Bias(σ)	Inf	Inf	0.0386	0.0865	0.3013	0.0490	0.0437	0.0304	0.0487	0.0312	0.0272	0.0376
RMSE(σ)	(Inf)	(Inf)	(0.2607)	(0.3848)	(0.4635)	(0.2646)	(0.2808)	(0.2619)	(0.3128)	(0.2605)	(0.2616)	(0.3079)
$k = -3$												
Bias(σ)	Inf	Inf	0.0261	0.0458	0.0650	0.0289	0.0267	0.0183	0.0272	0.0200	0.0152	0.0195
RMSE(σ)	(Inf)	(Inf)	(0.2093)	(0.2557)	(0.2296)	(0.2099)	(0.2169)	(0.2076)	(0.2287)	(0.2138)	(0.2068)	(0.2262)
$k = -1$												
Bias(σ)	7.3732	0.1948	0.0148	0.0177	0.0139	0.0094	0.0130	0.0071	0.0103	0.0100	0.0045	0.0059
RMSE(σ)	(111.48)	(0.2616)	(0.1463)	(0.1486)	(0.1448)	(0.1447)	(0.1553)	(0.1482)	(0.1487)	(0.1641)	(0.1465)	(0.1479)
$k = -0.5$												
Bias(σ)	0.2736	0.0178	0.0129	0.0126	0.0026	0.0038	0.0101	0.0042	0.0066	0.0076	0.0019	0.0029
RMSE(σ)	(0.3591)	(0.1260)	(0.1268)	(0.1258)	(0.1236)	(0.1243)	(0.1402)	(0.1317)	(0.1287)	(0.1509)	(0.1299)	(0.1282)
$k = 0.2$												
Bias(σ)	0.0063	0.0022	0.0170	0.0060	-0.0092	-0.0062	0.0066	-0.0020	-0.0002	0.0047	-0.0034	-0.0024
RMSE(σ)	(0.0995)	(0.1090)	(0.0979)	(0.0982)	(0.0932)	(0.0930)	(0.1194)	(0.1053)	(0.0997)	(0.1319)	(0.1048)	(0.1002)
$k = 1.5$												
Bias(σ)	0.0074	0.0039	NA	-0.0033	-0.0673	-0.1037	0.0039	-0.0019	-0.0007	0.0027	-0.0085	-0.0055
RMSE(σ)	(0.1249)	(0.1181)	(NA)	(0.0770)	(0.0902)	(0.1171)	(0.0920)	(0.0812)	(0.0767)	(0.1042)	(0.0786)	(0.0759)
$\hat{k}$												
$k = -5$												
Bias(k)	4.5036	4.0024	0.0176	0.0343	0.2350	0.0276	0.0122	-0.0073	-0.0001	0.0011	0.0050	0.0045
RMSE(k)	(4.5036)	(4.0024)	(0.4265)	(0.4821)	(0.3930)	(0.4257)	(0.4769)	(0.4441)	(0.4474)	(0.5182)	(0.4284)	(0.4482)
$k = -3$												
Bias(k)	2.5050	2.0067	0.0137	0.0237	0.0497	0.0165	0.0094	-0.0049	0.0006	0.0008	0.0029	0.0032
RMSE(k)	(2.5050)	(2.0068)	(0.2845)	(0.3108)	(0.3004)	(0.2836)	(0.3243)	(0.2983)	(0.2962)	(0.3638)	(0.2844)	(0.2950)
$k = -1$												
Bias(k)	0.5373	0.2002	0.0111	0.0137	0.0104	0.0058	0.0069	-0.0034	0.0007	0.0004	0.0004	0.0016
RMSE(k)	(0.5382)	(0.2234)	(0.1438)	(0.1453)	(0.1412)	(0.1427)	(0.1866)	(0.1598)	(0.1507)	(0.2263)	(0.1458)	(0.1460)
$k = -0.5$												
Bias(k)	0.1476	0.0282	0.0114	0.0112	0.0013	0.0024	0.0063	-0.0037	0.0001	0.0003	-0.0007	0.0007
RMSE(k)	(0.1643)	(0.1093)	(0.1094)	(0.1080)	(0.1066)	(0.1076)	(0.1586)	(0.1286)	(0.1177)	(0.1980)	(0.1148)	(0.1119)
$k = 0.2$												
Bias(k)	0.0068	0.0019	0.0182	0.0065	-0.0083	-0.0051	0.0055	-0.0085	-0.0041	0.0005	-0.0048	-0.0027
RMSE(k)	(0.0730)	(0.0897)	(0.0685)	(0.0703)	(0.0651)	(0.0644)	(0.1301)	(0.0902)	(0.0785)	(0.1671)	(0.0814)	(0.0745)
$k = 1.5$												
Bias(k)	0.0125	0.0041	NA	-0.0055	-0.1104	-0.1752	0.0072	-0.0043	-0.0018	0.0046	-0.0139	-0.0087
RMSE(k)	(0.2318)	(0.2132)	(NA)	(0.1167)	(0.1425)	(0.1929)	(0.1482)	(0.1241)	(0.1162)	(0.1764)	(0.1197)	(0.1148)