Pure and Applied Functional Analysis
Volume 6, Number 6, 2021, 1261–1271

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

RÉNYI ENTROPY AND CALIBRATION OF DISTRIBUTION TAILS

BOGDAN GRECHUK, MICHAEL ZABARANKIN, STAN URYASEV,
AND ALEXEI ZRAZHEVSKY

Abstract. The popular principle of Shannon entropy maximization subject to moment constraints yields distributions with light tails. Therefore, it is not suitable for estimation of probability distributions with heavy tails, which arise in various applications, including financial engineering, reliability theory, and climatology. However, with Rényi entropy in place of Shannon entropy, the principle yields distributions with heavy tails. This work obtains a general form of such distributions and proposes a novel method for estimating parameters of *generalized Pareto distribution* (GPD), which follows from Rényi entropy maximization subject to a constraint on expected value (this particular case was solved by Bercher (2008)). It also shows how a conditional tail GPD can be estimated based on quantile and CVaR (expected shortfall) regressions and, as an illustration, estimates a conditional tail GPD for Fidelity Magellan Fund return as a functions of stock indices. Quantile and CVaR regressions are implemented with the Portfolio Safeguard (PSG) optimization package, which has precoded errors for quantile and CVaR regressions.

1. Introduction

Method of moments is arguably one of the most popular techniques for estimating parameters of probability distributions. For example, parameters of normal distribution are estimated as the mean and variance of sample data. There are two factors that contribute to the popularity of this approach: (i) simplicity/clarity and (ii) stability/robustness. Simplicity/clarity makes it popular in the engineering community since it requires no elaborate knowledge of statistics and probability theory. Stability/robustness means that 10–15 data sample points are often sufficient for reliable estimation of the distribution parameters in question. However, it should be noted that the mean and standard deviation are integral characteristics of a whole distribution—they typically depend mostly on the data around the most probable central part of the distribution and represent distribution tails poorly. Yet precisely distribution tails is the main focus in various engineering applications, particularly in structural safety and reliability. Typically, they are described by quantiles and probabilities of exceedance rather than by moments of the whole distribution.

2020 *Mathematics Subject Classification.* 94A17, 46N10.
Key words and phrases. Maximum entropy principle, Renyi entropy, heavy tails, moments, generalized Pareto distribution (GPD), quantile regression, conditional value-at-risk (CVaR), CVaR regression, expected shortfall regression.

Another well-known approach for estimating the probability distribution of a random variable with insufficient information is *entropy maximization* [8], which prescribes to choose distribution $P(x)$ maximizing the Shannon entropy

$$H(P) = -\int P(x)\ln(P(x))dx$$

subject to any given information / constraints on $P(x)$. This approach yields distributions with exponentially decreasing tails. In particular, if the constraints depend only on distribution of the random variable, then any solution to Shannon entropy maximization problem must have log-concave density [7].

However, in several applications, e.g., financial engineering, reliability theory and climatology [1, 2, 13], the probability density functions of corresponding random variables have heavy tails, for example, as in power-law distributions that maximize Rényi (or Tsallis) entropy [16]. In particular, Rényi entropy maximization subject to a constraint on the expected value yields a *generalized Pareto distribution* (GPD) [2], and the same problem with an additional constraint on a deviation measure, e.g., standard deviation and mean-absolute deviation, was solved in [3, 6, 9].

The contribution and organization of this work are as follows. Section 2 solves Rényi entropy maximization subject to constraints on generalized moments: a solution has heavy tails and, as a particular case, includes GPD. Section 3 proposes *harmonic method* for estimating GPD parameters and shows how conditional tail GPD can be estimated based on quantile and CVaR regressions. As an illustration, Section 4 tests the harmonic method on artificial data and estimates conditional tail GPD for the return of the Fidelity Magellan fund that consists of different stock indices. Section 5 concludes the work. Appendix 6 presents the proof of the theorem establishing solution for Rényi entropy maximization subject to moment constraints.

2. Estimation of probability distributions with heavy tails

Let $\mathbb{R} = (-\infty, \infty)$ be the real line, $\mathbb{R}^+ = [0, \infty)$ the set of nonnegative real numbers, and $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ the extended real line, and let $[x]_+ = \max\{x, 0\}$ for any $x \in \mathbb{R}$.

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, where Ω denotes the designated space of future states ω, Σ is a σ-algebra of sets in Ω, and $\mathbb{P}$ is a probability measure on (Ω, Σ). A random variable (r.v.) is any measurable function from Ω to $\mathbb{R}$. An r.v. X is continuously distributed if there exists a Lebesgue integrable function $f_X : \mathbb{R} \to \mathbb{R}^+$ such that $\mathbb{P}[a < X < b] = \int_a^b f_X(t)dt$ for all $a, b \in \overline{\mathbb{R}}$ with $a < b$. The function f_X is called the probability density function (PDF) of X. We assume that the probability space $(\Omega, \Sigma, \mathbb{P})$ is *atomless*, i.e., there exists a continuously distributed r.v. $X : \Omega \to \mathbb{R}$.

A continuously distributed r.v. X has support (a, b), $-\infty \leqslant a < b \leqslant \infty$, if $\mathbb{P}[a < X < b] = \int_a^b f_X(t)dt = 1$. Let $L^1_+(a, b)$ be the set of all functions $f : (a, b) \to \mathbb{R}^+$ such that $\int_a^b f(t)dt < +\infty$. Given $f \in L^1_+(a, b)$, there exists a continuously distributed r.v. X with support (a, b) and PDF $f_X = f$ if and only if $\int_a^b f(t)dt = 1$.

For any continuously distributed r.v. X with support (a,b), its Rényi differential entropy of order $\varkappa$ is defined by [14]:

$$H_\varkappa(f_X) = \frac{1}{1-\varkappa}\ln\int_a^b (f_X(t))^\varkappa dt, \qquad \varkappa > 0, \quad \varkappa \neq 1.$$

Let $l(a,b)$ be the set of locally integrable functions $\phi : (a,b) \to \mathbb{R}$, i.e., such that $\int_K |\phi(x)|dx < +\infty$ for any compact subset K of (a,b). Entropy maximization with the Renyi entropy subject to moment constraints is formulated[1] as [17, §5.3]:

$$\max_{f\in L^1_+(a,b)} H_\varkappa(f) \quad \text{subject to} \quad \int_a^b \phi_k(t)f(t)\,dt = \mu_k, \quad k = 0,1,\dots,m, \tag{2.1}$$

where $-\infty \leqslant a < b \leqslant +\infty$, $\phi_0(t) = 1$, $\mu_0 = 1$, $\phi_k(t) \in l(a,b)$, $k = 1,\dots,m$, $\mu_k \in \mathbb{R}$, $k = 1,\dots,m$.

Proposition 2.1. *If a solution to* (2.1) *has a finite Rényi entropy, then it is unique.*

Proof. Let f_1 and f_2 be two different solutions to (2.1) such that $-\infty < H_\varkappa(f_1) = H_\varkappa(f_2) < +\infty$, or, equivalently,

$$0 < \int_a^b (f_1(t))^\varkappa dt = \int_a^b (f_2(t))^\varkappa dt < +\infty.$$

Since $g(y) = y^\varkappa$ is strictly concave on $\mathbb{R}^+$ for $0 < \varkappa < 1$ and is strictly convex for $\varkappa > 1$, this implies that

$$\int_a^b (f(t))^\varkappa dt \leqslant \frac{1}{2}\int_a^b (f_1(t))^\varkappa dt + \frac{1}{2}\int_a^b (f_2(t))^\varkappa dt = \int_a^b (f_1(t))^\varkappa dt, \quad 1 \leqslant \varkappa,$$

where $f = (f_1 + f_2)/2$. Consequently, $H_\varkappa(f) > H_\varkappa(f_1)$. Since f satisfies all the constrains in (2.1), this contradicts the optimality of f_1. □

Theorem 2.2. *Let $\varkappa > 0$ and $\varkappa \neq 1$. If there exist real numbers $\lambda_0^*,\dots,\lambda_m^*$ such that*

$$f_0(t) = \left[\sum_{k=0}^m \lambda_k^*\,\phi_k(t)\right]_+^{\frac{1}{\varkappa-1}}, \quad a < t < b, \tag{2.2}$$

is finite on (a,b) and satisfies the constraints in (2.1)*, then f_0 solves* (2.1)*.*

Proof. See Appendix 6. □

In [6, Appendix B], (2.2) is obtained by the Lagrange multipliers technique.

If $\sum_{k=0}^m \lambda_k^*\,\phi_k(t) > 0$, $a < t < b$, (2.2) simplifies.

Corollary 2.3. *Let $\varkappa > 0$ and $\varkappa \neq 1$. If there exist real numbers $\lambda_0^*,\dots,\lambda_m^*$ such that*

$$\sum_{k=0}^m \lambda_k^*\,\phi_k(t) > 0, \quad a < t < b, \tag{2.3}$$

[1]If the integral $\int_a^b \phi_k(t)f(t)\,dt$ does not exist for some k, then f does not satisfy the corresponding constraint.

and function

$$f_0(t) = \left(\sum_{k=0}^{m} \lambda_k^* \phi_k(t)\right)^{\frac{1}{\varkappa-1}}, \quad a < t < b, \tag{2.4}$$

satisfies the constraints in (2.1)*, then* f_0 *solves* (2.1).

With (2.4), the constraints in (2.1) yield a system for $\lambda_0^*, \dots, \lambda_m^*$:

$$\int_a^b \left(\sum_{k=0}^{m} \lambda_k^* \phi_k(t)\right)^{\frac{1}{\varkappa-1}} \phi_i(t)dt = \mu_i, \quad i = 0, 1, \dots, m. \tag{2.5}$$

Example 2.4 (generalized Pareto distribution (GPD))**.** Let $a \in \mathbb{R}$ be arbitrary, $b = \infty$, $m = 1$, $\phi_1(t) = t$, $\mu_1 = \mu > a$, and $\frac{1}{2} < \varkappa < 1$. Then the solution of (2.1) is determined by (see [2])

$$f(t) = \frac{\varkappa}{2\varkappa - 1} \frac{1}{\mu - a} \left(1 + \frac{1-\varkappa}{2\varkappa - 1} \frac{t-a}{\mu - a}\right)^{\frac{1}{\varkappa-1}}, \quad a < t. \tag{2.6}$$

Detail. Substitution $t = s + a$ reduces the problem to (2.1) with $a' = 0$, $b = \infty$, $m = 1$, $\phi_1(s) = s$, and $\mu_1 = \mu - a$. In this case, condition (2.3) simplifies to

$$\lambda_0 + \lambda_1 s > 0, \quad 0 < s,$$

which holds provided that $\lambda_0 \geqslant 0$ and $\lambda_1 > 0$. System (2.5) takes the form

$$\frac{1-\varkappa}{\varkappa\lambda_1} \lambda_0^{\frac{\varkappa}{\varkappa-1}} = 1, \qquad \frac{(1-\varkappa)^2}{\lambda_1^2 \varkappa(2\varkappa-1)} \lambda_0^{\frac{2\varkappa-1}{\varkappa-1}} = \mu - a,$$

which has a closed-form solution:

$$\lambda_0 = \left(\frac{\varkappa}{(2\varkappa-1)(\mu-a)}\right)^{\varkappa-1}, \qquad \lambda_1 = \frac{1-\varkappa}{\varkappa} \left(\frac{\varkappa}{(2\varkappa-1)(\mu-a)}\right)^{\varkappa},$$

and (2.4) yields (2.6). □

3. Parameter estimation for generalized Pareto distribution

GPD PDF (2.6) has three parameters: a, μ , and $\varkappa$. While input parameters a and μ can be readily estimated from observations, $\varkappa$ can be estimated by likelihood maximization. Suppose there is a sample $x_1, \dots, x_n$ of n independent and identically distributed observations drawn from the PDF (2.6), in which a and μ are given. The likelihood function with (2.6) is determined by

$$\begin{aligned} \ell(\varkappa; x_1, \dots, x_n) &= \prod_{i=1}^{n} f(x_i|\varkappa) \\ &= \left(\frac{1}{\mu-a}\right)^n \left(\frac{\varkappa}{2\varkappa-1}\right)^n \prod_{i=1}^{n} \left(1 + \frac{1-\varkappa}{2\varkappa-1} \frac{x_i - a}{\mu - a}\right)^{\frac{1}{\varkappa-1}}, \end{aligned}$$

and $\varkappa$ is estimated by the solution of likelihood maximization

$$\widehat{\varkappa} \in \arg \max_{\frac{1}{2}<\varkappa<1} \ell(\varkappa; x_1, \dots, x_n).$$

3.1. **Harmonic method.** Another approach to estimating $\varkappa$ is to find $\varkappa$ from the condition

$$\int_a^{\infty} \ln(t-a) f(t) dt = C, \tag{3.1}$$

where C is estimated from observations as the expected value of $\ln(t-a)$. In this case, let $\eta = \mu - a$, $s = t - a$, and

$$g(s) = f(s+a) = \frac{\varkappa}{2\varkappa - 1}\frac{1}{\eta}\left(1 + \frac{s}{z\,\eta}\right)^{\frac{1}{\varkappa - 1}}, \qquad z = \frac{2\varkappa - 1}{1 - \varkappa},$$

where $f(t)$ is given by (2.6). If $\eta > 0$ and $\frac{1}{2} < \varkappa < 1$, then $z \in (0, +\infty)$ and

$$F(\varkappa, \eta) \equiv \int_0^{\infty} g(s) \ln s \, ds = -H_z + \ln z + \ln \eta, \quad H_z = \int_0^1 \frac{1 - t^z}{1 - t} dt,$$

where H_z is the harmonic number [4]. The condition (3.1) can be written as

$$p(z) = \widetilde{C}, \tag{3.2}$$

where

$$p(z) = H_z - \ln z, \qquad \widetilde{C} = \ln \eta - C.$$

Proposition 3.1. *Equation* (3.2) *has a unique solution with respect to z if and only if $\widetilde{C} > \gamma$, where $\gamma = \lim_{z\to\infty}(H_z - \ln z) = 0.57721566$ is the Euler constant [11].*

Proof. Since

$$\frac{d}{dz} H_z = \int_0^1 \frac{-t^z \ln t}{1 - t} dt < \int_0^1 t^{z-1} dt = \frac{1}{z},$$

$p'(z) < 0$ on $(0, \infty)$, and $p(z)$ is a strictly decreasing function on $(0, \infty)$: it decreases from $\lim_{z\to 0} p(z) = +\infty$ to $\lim_{z\to\infty} p(z) = \gamma$. □

Thus, $\varkappa$ can be found from (3.2) as follows: for a sample $x_1, \ldots, x_n$, calculate $\widetilde{C} = \ln\left(\frac{1}{n}\sum_{i=1}^n x_i\right) - \frac{1}{n}\sum_{i=1}^n \ln x_i$, and if $\widetilde{C} > \gamma = 0.57721566$, $\widehat{z}$ is a solution of $H_z - \ln z = \widetilde{C}$ and $\varkappa$ is estimated by $\widehat{\varkappa} = \frac{1+\widehat{z}}{2+\widehat{z}}$. This will be called *harmonic method.* If $\widetilde{C} < \gamma$, then estimate does not exist. In this case, $\widetilde{C}$ can be increased by dropping some of $x_1, \ldots, x_n$.[2]

3.2. **Parameter estimation of conditional tail distribution.** The GPD is often used to model a distribution tail. A conditional tail distribution can be modeled by the GPD estimated for positive residuals of either quantile regression

$$Y = \boldsymbol{c}^\top \boldsymbol{X} + e, \quad Q_\alpha(e) = 0, \tag{3.3}$$

or CVaR regression

$$Y = \boldsymbol{c}^\top \boldsymbol{X} + e, \qquad \mathrm{CVaR}_\alpha(e) = 0, \tag{3.4}$$

where $\boldsymbol{X}$ and Y are independent and dependent random vector and random variable, respectively. Let matrix $\widehat{\mathbf{X}}$ and vector $\widehat{\mathbf{Y}}$ be realizations of $\boldsymbol{X}$ and Y, respectively:

[2]The procedure is as follows. Let $\widetilde{C}_{i_1,\ldots,i_m}$ correspond to the set $x_1, \ldots, x_n$ without $x_{i_1}, \ldots, x_{i_m}$, and let $(i_1^*, \ldots, i_m^*) \in \arg\max_{i_1,\ldots,i_m} \widetilde{C}_{i_1,\ldots,i_m}$. Set $m = 1$. If $\widetilde{C}_{i_1^*,\ldots,i_m^*} > \gamma$ then stop, otherwise increase m by 1 and so on.

$(\widehat{\mathbf{Y}}, \widehat{\mathbf{X}})$ is called an extended design matrix. In the α-CVaR regression, the right tail of the distribution of residuals is assumed to start from residuals' α-quantile, so that α-quantile must be subtracted from the residuals of α-CVaR regression to calculate positive right tail (the GPD is estimated for positive samples). The procedure for estimating a conditional tail GPD based on either (3.3) or (3.4) is as follows.

(1) Find $\mathbf{c}$ by minimizing the Koenker and Basset error [10] of e in (3.3) and by minimizing the Rockafellar error [5, 15, mixed-quantile quadrangle] of e in (3.4). Let $\widehat{\mathbf{c}}$ be the corresponding solution and let $\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{c}}^\top \mathbf{X}$ be residuals. For (3.4), define adjusted residuals to be $\widetilde{\mathbf{e}} = \boldsymbol{e} - Q_\alpha(\boldsymbol{e})$.
(2) Estimate μ in the GPD as the average value of positive residuals of $\boldsymbol{e}$ for (3.3) and as the average value of positive *adjusted* residuals of $\widetilde{\mathbf{e}}$ for (3.4). Let $\widehat{\mu}$ denote the corresponding estimate.
(3) Use MLE or harmonic method to estimate $\varkappa$ in the GPD for positive residuals of $\boldsymbol{e}$ for (3.3) and for positive *adjusted* residuals of $\widetilde{\mathbf{e}}$ for (3.4). Let $\widehat{\varkappa}$ denote the corresponding estimate.
(4) For some observation $\boldsymbol{X}_0$ of $\mathbf{X}$, calculate conditional α-quantile, $Q_\alpha(\boldsymbol{Y}|\boldsymbol{X}_0) = \widehat{\mathbf{c}}^\top \boldsymbol{X}_0$, for (3.3) and conditional α-CVaR, $\mathrm{CVaR}_\alpha(\boldsymbol{Y}|\boldsymbol{X}_0) = \widehat{\mathbf{c}}^\top \boldsymbol{X}_0$, for (3.4). Conditional tail GPD with $(\widehat{\mu}, \widehat{\varkappa})$ starts from the conditional estimate of the quantile, $\widehat{a} = \widehat{\mathbf{c}}^\top \boldsymbol{X}_0$, for (3.3) and from $\widehat{a} = \mathrm{CVaR}_\alpha(\boldsymbol{Y}|\boldsymbol{X}_0) - Q_\alpha(\boldsymbol{e})$ for (3.4).

4. Case study: Numerical experiments

The α-quantile regression (3.3) and α-CVaR regression (3.4) are implemented with AORDA Portfolio Safeguard (PSG) package (http://www.aorda.com/).[3]

4.1. Parameter estimation with artificial data.

This section illustrates the approach for estimating parameter $\varkappa$ of GPD suggested in §3. We generated 100 samples with size of $n = 1250$ and 100 samples with $n = 12500$ from GPD with $a = 0$, $\mu_0 = 0.7778$, $\varkappa_0 = 0.9091$. In this case, μ is estimated as the average value of each sample, whereas $\varkappa$ is estimated as the solution of (3.2) (harmonic method). For comparison, $\varkappa$ is also estimated with the standard maximum likelihood (ML) method. Table 1 shows minimum, mean and maximum values, as well as the deviation $\Delta =$ (maximum $-$ minimum) of the relative errors $(\varkappa_0 - \widehat{\varkappa})/\varkappa_0$ and $(\mu_0 - \widehat{\mu})/\mu_0$ between the true and estimated values of $\varkappa$ and μ over 100 samples.

Relative errors in Table 1 show that the harmonic method estimates are better on average than the ML estimates, although they have a larger deviation. The harmonic method uses only two characteristics of the sample: the logarithm of the average sample and the average of the logarithms of the sample, and, in contrast to the ML method, it does not require solving an optimization problem. Both harmonic and ML methods yield similar results with low relative errors.

4.2. Conditional tail parameter estimation with artificial data: quantile regression.

Samples of $\boldsymbol{Y}$ and $\boldsymbol{X}$ in the α-quantile regression (3.3) are generated as follows:

[3] For the case study data, codes, and solutions, see `http://uryasev.ams.stonybrook.edu/index.php/research/testproblems/advanced-statistics/case-study-renyi-entropy-maximization/`

TABLE 1. Minimum, mean and maximum values and the deviation of the relative error between the true and estimated values of $\varkappa$ for 100 samples with $n = 1250$ and for 100 samples with $n = 12500$ drawn from a GPD with $a = 0$, $\mu = 0.7778$, $\varkappa = 0.9091$.

method	sample size	parameter	min (%)	mean (%)	max (%)	Δ (%)
	1250	μ	-10.82	-0.28	5.49	16.32
ML	1250	$\varkappa$	-8.9	-0.36	8.81	17.71
harmonic	1250	$\varkappa$	-8.61	-0.27	11.51	20.13
	12500	μ	-2.42	-0.08	2.01	4.44
ML	12500	$\varkappa$	-1.79	-0.03	1.98	3.78
harmonic	12500	$\varkappa$	-2.91	-0.02	2.86	5.77

(1) Generate n random samples $\boldsymbol{X} = \{X_i, i = 1, \ldots, n\}$ from the uniform normal distribution $U(0, 1)$.
(2) Generate random samples $\boldsymbol{e} = \{e_i, i = 1, \ldots, n\}$ from the standard normal distribution $N(0, 1)$.
(3) For $\alpha = 0.75$, calculate quantile q_α for $\boldsymbol{e}$ and the set $G_1 = \{e_i : e_i < q_\alpha, i = 1, \ldots, n\}$. Let n_1 be the number of elements in G_1.
(4) With PDF (2.6) and parameters $a = 0$, $\mu_0 = 0.7778$, $\varkappa_0 = 0.9091$ generate samples $\nu_1, \ldots, \nu_{n-n_1}$.
(5) Combine generated data: $\Omega = G_1 \bigcup \{q_\alpha, \nu_1, \ldots, \nu_{n-n_1}\}$.
(6) Randomly mix elements in the set Ω.
(7) Use $Y_i = 5 + 4X_i + \omega_i$, $\omega_i \in \Omega$, $i = 1, \ldots, n$ to generate $\boldsymbol{Y} = \{Y_i, i = 1, \ldots, n\}$.

GPD parameters are then estimated based on generated 100 samples with the sample size of $n = 5000$ and 100 samples with $n = 50000$. The four-step procedure from §3.2 then uses a 0.75-quantile regression, which yields 1250 positive residuals for $n = 5000$ and 12500 for $n = 50000$. Table 2 shows the minimum, mean and maximum values as well as the deviation Δ = (maximum $-$ minimum) of the relative errors $(\varkappa_0 - \widehat{\varkappa})/\varkappa_0$ and $(\mu_0 - \widehat{\mu})/\mu_0$ between the true and estimated values of $\varkappa$ and μ over 100 samples for the corresponding sample size. Results in Table 2 are similar to those in Table 1.

4.3. **Tail parameter estimation for Fidelity Magellan fund: quantile and CVaR regressions.** As yet another illustration, parameters of conditional tail GPD are estimated based on 0.75-quantile and 0.75-CVaR regressions for 1264 daily returns ($N = 1264$) of the Fidelity Magellan fund regressed against four factors: the Russell Value Index (RUJ), RUSSELL 1000 VALUE INDEX (RLV), Russell 2000 Growth Index (RUO) and Russell 1000 Growth Index (RLG). The four-step procedure from §3.2 uses the harmonic method for positive residuals ($n = 316$) and yields $\widehat{\mu} = 0.0041$ and $\widehat{\varkappa} = 0.791$ for the 0.75-quantile regression and $\widehat{\mu} = 0.0040$ and $\widehat{\varkappa} = 0.859$ for the 0.75-CVaR regression, where in the latter, the residuals are shifted by $Q_\alpha(\boldsymbol{e}) = 0.0041$. Conditional 0.75-quantile and conditional 0.75-CVaR, calculated for some observations of the four factors, are then $\widehat{a} = 0.0217$ and $\mathrm{CVaR}_{0.75}(\boldsymbol{Y}|\boldsymbol{X}_0) = 0.026$, respectively. In the 0.75-CVaR regression, the estimated

Table 2. Minimum, mean and maximum values and the standard deviation of the relative error between the true and estimated values of μ and $\varkappa$ over 100 samples for corresponding sample size, where samples are generated through the 0.75-quantile regression (3.3) with the four-step procedure from §3.2.

method	sample size	parameter	min, %	mean (%)	max (%)	Δ (%)
	5000	μ	−11.05	−0.41	5.64	16.7
ML	5000	$\varkappa$	−9.19	−0.51	8.39	17.58
harmonic	5000	$\varkappa$	−8.86	−1.22	8.54	17.41
	50000	μ	−2.38	−0.11	1.59	4.33
ML	50000	$\varkappa$	−1.86	−0.06	1.95	3.81
harmonic	50000	$\varkappa$	−2.89	−0.18	2.6	5.5

0.75-quantile of residuals is 0.0041, i.e., $Q_{0.75}(\boldsymbol{e}) = 0.0041$, and the conditional tail starts from $\widehat{a} = 0.026 - 0.0041 = 0.0219$. Figure 1 shows bar-charts of residuals and probability densities for conditional tail GPD estimated based on the 0.75-quantile and 0.75-CVaR regressions.

5. Summary

This work obtains a general solution of the Rényi entropy maximization problem with moment constraints and proposes novel methods for estimating parameters of a generalized Pareto distribution (GPD), which is a particular case of that solution. For example, GPD's parameter $\varkappa$ can be estimated by the *harmonic method*, which requires solving a transcendental equation—conditions guaranteeing existence and uniqueness of a solution to that equation are provided. This work also shows how conditional tail GPD can be estimated based on quantile and CVaR regressions. As an illustration, it estimates conditional tail GPD for the Fidelity Magellan fund return with Portfolio Safeguard (PSG) package, which has precoded error functions for quantile and CVaR regressions.

6. Appendix: Proof of theorem 2.2

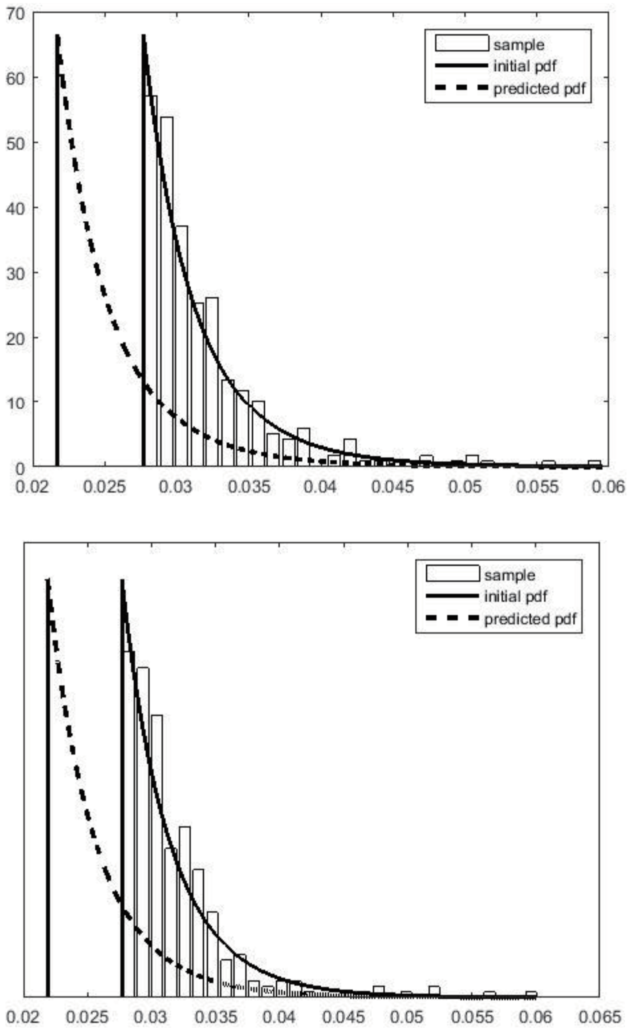
Let f be an arbitrary PDF with support (a, b) satisfying the constraints in (2.1). Then

$$\int_a^b \phi_k(t) f(t)\, dt = \mu_k = \int_a^b \phi_k(t) f_0(t) dt, \quad k = 0, 1, \dots, m.$$

By multiplying k-th equation by λ_k^* and adding the equations, we obtain

$$\int_a^b g(t) f(t)\, dt = \int_a^b g(t) f_0(t) dt,$$

where $g(t) = \sum_{k=0}^m \lambda_k^* \phi_k(t)$. Let J^+ and J^- be subsets of (a, b) on which $g(t) > 0$, $g(t) \leqslant 0$, respectively. Note that if $0 < \varkappa < 1$, then f_0 is finite-valued on (a, b) only if J_- is an empty set. If $\varkappa > 1$, then J_- may be non-empty, and $f_0(t) = 0$, $t \in J_-$.



(a)

(b)

FIGURE 1. Estimation of conditional tail GPD with (a) the quantile regression (3.3) and (b) CVaR regression (3.4): histograms of (a) positive residuals of 0.75-quantile regression and (b) positive adjusted residuals of 0.75-CVaR regression both starting from 0.75-quantile of historical return of the Magellan fund (dependent variable), the GPD fitted to (a) positive residuals and (b) positive adjusted residuals (continuous curve), and the conditional tail GPD starting from the quantile estimate for a new observation of independent factors (dotted curve).

We obtain

$$\int_a^b g(t)f_0(t)dt = \int_{J^+} g(t)f_0(t)dt + \int_{J^-} g(t)f_0(t)dt = \int_a^b [g(t)]_+ f_0(t)\,dt + 0.$$

and consequently,

$$\int_a^b [g(t)]_+ f(t)\,dt = \int_{J^+} g(t)f(t)\,dt \geqslant \int_a^b g(t)f(t)\,dt = \int_a^b [g(t)]_+ f_0(t)\,dt.$$

This inequality can be rewritten as

$$\int_{\mathbb{R}} f_0^{\varkappa-1}(t)f(t)dt \geqslant \int_{\mathbb{R}} f_0^{\varkappa-1}(t)f_0(t)\,dt,$$

where we have used the definition of f_0. If $\varkappa > 1$, this is equivalent to

$$\left(\int_{\mathbb{R}} f_0^{\varkappa-1}(t)f(t)dt\right)^{1/(1-\varkappa)} \leqslant \left(\int_{\mathbb{R}} f_0^{\varkappa}(t)dt\right)^{1/(1-\varkappa)}. \tag{6.1}$$

If $0 < \varkappa < 1$, then J_- is an empty set, and consequently, (6.1) holds as equality.

Lutwak et al. [12, Lemma 1] showed that inequality

$$\frac{\left(\int_{\mathbb{R}} g^{\varkappa-1}(t)f(t)dt\right)^{1/(1-\varkappa)}\left(\int_{\mathbb{R}} g^{\varkappa}(t)dt\right)^{1/\varkappa}}{\left(\int_{\mathbb{R}} f^{\varkappa}(t)dt\right)^{1/(\varkappa(1-\varkappa))}} \geqslant 1 \tag{6.2}$$

holds for all densities f and g for which the integrals exist and finite. If f and g have support (a,b) then all integrals $\int_{\mathbb{R}}$ in (6.2) can be replaced by $\int_a^b$. By applying (6.2) with $g = f_0$, we obtain

$$\begin{aligned}\left(\int_a^b f^{\varkappa}(t)dt\right)^{1/(\varkappa(1-\varkappa))} &\leqslant \left(\int_a^b f_0^{\varkappa-1}(t)f(t)dt\right)^{1/(1-\varkappa)}\left(\int_a^b f_0^{\varkappa}(t)dt\right)^{1/\varkappa}\\ &\leqslant \left(\int_a^b f_0^{\varkappa}(t)dt\right)^{1/(\varkappa(1-\varkappa))},\end{aligned}$$

where the second inequality follows from (6.1). This implies that $H_{\varkappa}(f_0) \geqslant H_{\varkappa}(f)$.

Acknowledgement

Research was partially supported by AFOSR grant FA9550-18-1-0391 on Massively Parallel Approaches for Buffered Probability Optimization and Application.

References

[1] M. Asadi and I. Bayramoglu, *The mean residual life function of a k-out-of-n structure at the system level*, IEEE Trans. Reliab. **55** (2006), 314–318.

[2] J. F. Bercher, *Tsallis distribution as a standard maximum entropy solution with 'tail' constraint*, Phys. Lett. A **372** (2008), 5657–5659.

[3] J. Costa, A. Hero and C. Vignat, *On solutions to multivariate maximum-entropy problems*, Lecture Notes in Computer Science (A. Rangarajan, M. Figueiredo, J. Zerubia, eds.), Springer-Verlag, Berlin, **2683** (2003), 211–228.

[4] O. Espinosa and V.H. Moll, *A Generalized polygamma function*, Integral Transform Spec. Funct. **15** (2004), 101–115.

[5] A. Golodnikov, V. Kuzmenko and S. Uryasev, *CVaR Regression Based on the Relation between CVaR and Mixed-Quantile Quadrangles*, J. of Risk and Fin. Manag. **12** (2019), 1–22.

[6] B. Grechuk, A. Molyboha and M. Zabarankin, *Maximum entropy principle with general deviation measures*, Math. of Oper. Res. **34** (2009), 445–467.

[7] B. Grechuk and M. Zabarankin, *Regression Analysis: Likelihood, Error and Entropy*, Math. Program. **174** (2019), 145–166.

[8] E. T. Jaynes, *Information theory and statistical mechanics*, Phys. Rev. **106** (1957), 620–630.

[9] O. Johnson and C. Vignat, *Some results concerning maximum Rényi entropy distributions*, Ann. I. H. Poincaré – PR **43** (2007), 339–351.
[10] R. Koenker and G. Bassett, *Regression Quantiles*, Econometrica **46** (1978), 33–50.
[11] J. C. Lagarias, *Euler's constant: Euler's work and modern developments*, Bull. Am. Math. Soc. **50** (2013), 527–628.
[12] E. Lutwak, D. Yang and G. Zhang, *Cramer-Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information*, IEEE Trans. Inf. Theory **51** (2005), 473–478.
[13] A. J. V. Montfort and J. V. Witter, *Generalized Pareto distribution applied to rainfall depths*, Hydrol. Sci. J. **31** (1986), 151–162.
[14] A. Rényi, *On measures of information and entropy*, Proc. 4th Berkeley Symp. on Math., Stat. and Prob., Univ. of California Press, Berkeley, CA, 1961, 547–561.
[15] R. T. Rockafellar and S. Uryasev, *The Fundamental Risk Quadrangle in Risk Management, Optimization and Statistical Estimation*, Surv. Oper. Res. Manage. Sci. **18** (2013), 33–53.
[16] C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, J. Stat. Phys. **52** (1988), 479–487.
[17] M. Zabarankin and S. Uryasev, *Statistical Decision Problems: Selected Concepts and Portfolio Safeguard Case Studies*, Series: Springer Optimization and Its Applications, **85**, Springer, 2014.

Manuscript received November 30 2019
revised May 24 2020

B. Grechuk
University of Leicester, UK
E-mail address: bg83@leicester.ac.uk

M. Zabarankin
Stevens Institute of Technology, USA
E-mail address: mzabaran@stevens.edu

S. Uryasev
Stony Brook University, USA
E-mail address: stanislav.uryasev@stonybrook.edu

A. Zrazhevsky
American Optimal Decisions, Inc., USA
E-mail address: alex.zrazhevsky@aorda.com