Renyi Entropy Maximization and Regression of Distribution Tail

Based on “Renyi Entropy and Calibration of Distribution Tails” by Grechuk et al.

Background

The objective of this lecture is to demonstrate a new methodology for conditional estimating of a distribution tail. It is based on approaches for Rényi entropy maximization subject to constrains on generalized moments. 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. However, it should be noted that the mean and standard deviation are integral characteristics of a whole distribution — they pick up performance of the data around the most probable central part of the distribution and typically represent distribution tails poorly.

Another well-known approach for estimating the probability distribution of a random variable with insufficient information is entropy maximization [7], which prescribes to choose distribution $P(x)$ maximizing the Shannon entropy

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

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 [6].

However, in several applications, e.g., financial engineering, reliability theory and climatology [2] [1] [11], the probability density functions of corresponding random variables have heavy tails, for example, as in power-law distributions that maximize Rényi (or Tsallis) entropy [13]. A tail distribution is the probability density function (pdf) of the excesses $X_a$ of a random variate $X$ over a threshold $a$. Then it is shifted by $a$ toward the origin and is normalized to 1 [2]. The problem of Rényi entropy maximization subject to a constraint on a deviation measure, e.g., standard deviation and mean-absolute deviation, was solved in [3] [8] [5]. This Lecture addresses Rényi entropy maximization subject to constrains on generalized moments and proposes approaches for estimating parameters in the corresponding maximum entropy distributions.

This lecture consists of two parts. Part 1 describes Method of Moments with Rényi Entropy for estimation of distributions. Part 2 describes how to build a conditional distribution of the distribution tail, under condition that values of some explanatory factors are observed.

Part 1: Method of Moments with Renyi Entropy for Estimation of Distributions

Let $\mathbb{R} = (-\infty, \infty)$ be the real line, $\mathbb{R}^+ = [0, \infty)$ the set of non-negative 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 $\Omega$ be a probability space, i.e., a measure space such that the measure of the whole space is equal to one, and let $\mathbb{P}$ be the corresponding probability measure. A random variable (r.v.) is any measurable function from $\Omega$ to $\mathbb{R}$. An r.v. $X$ is continuously distributed if there exists a Lebesgue integrable function $f_X : \mathbb{R} \rightarrow \mathbb{R}^+$ such that $\mathbb{P}[a < X < b] = \int_a^b f_X(t) dt$ for all $a, b \in \mathbb{R}$ such that $a < b$. The function $f_X$ is called the probability density function (PDF) of $X$. We will assume that $\Omega$ contains at least one continuously distributed r.v. (such probability spaces are called atomless).

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

For any continuously distributed r.v. \( X \) with support \((a, b)\), its Rényi differential entropy of order \( \kappa \) is defined by [12]

\[
H^{\kappa}(f_{X}) = \frac{1}{1-\kappa} \ln \int_{a}^{b} (f_{X}(t))^\kappa \, dt, \quad \kappa > 0, \quad \kappa \neq 1.
\]  

(1)

Let \( l(a,b) \) be the set of locally integrable functions \( \phi : (a, b) \rightarrow \mathbb{R} \), i.e., such that \( \int_{K} |\phi(x)| \, dx < +\infty \) for any compact subset \( K \) of \((a, b)\). The maximum entropy principle with the Rényi entropy and moment constraints is formulated as:

\[
\max_{f \in L_{+}^{1}(a,b)} H^{\kappa}(f) \quad \text{subject to} \quad \int_{a}^{b} \phi_{k}(t) f(t) \, dt = \mu_{k}, \quad k = 0, 1, \ldots, m,
\]

(2)

where \(-\infty \leq a < b \leq +\infty, \phi_{0}(t) = 1, \mu_{0} = 1, \phi_{k}(t) \in l(a,b), k = 1, \ldots, m, \mu_{k} \in \mathbb{R}, k = 1, \ldots, m.\)

**Proposition 1.** If a solution to (2) has finite Rényi entropy, then it is unique.

**Proof.** Let \( f_{1} \) and \( f_{2} \) be two different solutions to (2), such that \(-\infty < H^{\kappa}(f_{1}) = H^{\kappa}(f_{2}) < +\infty\), or, equivalently,

\[
0 < \int_{a}^{b} (f_{1}(t))^\kappa \, dt = \int_{a}^{b} (f_{2}(t))^\kappa \, dt < +\infty.
\]

Since \( g(y) = y^{\kappa} \) is strictly concave for \( 0 < \kappa < 1 \) and strictly convex for \( \kappa > 1 \), this implies that

\[
\int_{a}^{b} (f(t))^\kappa \, dt \leq \frac{1}{2} \int_{a}^{b} (f_{1}(t))^\kappa \, dt + \frac{1}{2} \int_{a}^{b} (f_{2}(t))^\kappa \, dt = \int_{a}^{b} (f_{1}(t))^\kappa \, dt, \quad 1 \leq \kappa,
\]

where \( f = (f_{1} + f_{2})/2. \) Hence, \( H^{\kappa}(f) > H^{\kappa}(f_{1}). \) Since \( f \) satisfies all the constrains in (2), this contradicts the optimality of \( f_{1}. \)

**Theorem 1.** Let \( \kappa > 0 \) and \( \kappa \neq 1 \). If there exist real numbers \( \lambda_{0}^{*}, \lambda_{1}^{*}, \ldots, \lambda_{m}^{*} \) such that

\[
f_{0}(t) = \left[ \sum_{k=0}^{m} \lambda_{k}^{*} \phi_{k}(t) \right]^{\frac{1}{\kappa-1}}, \quad a < t < b,
\]

(3)

is finite on \((a, b)\) and satisfies the constraints in (2), then \( f_{0} \) solves (2).

In [5] Appendix B, (3) is obtained by the Lagrange multipliers technique.

If \( \sum_{k=0}^{m} \lambda_{k}^{*} \phi_{k}(t) > 0, a < t < b, \) (3) simplifies.

**Corollary 1.** Let \( \kappa > 0 \) and \( \kappa \neq 1 \). If there exist real numbers \( \lambda_{0}^{*}, \lambda_{1}^{*}, \ldots, \lambda_{m}^{*} \) such that

\[
\sum_{k=0}^{m} \lambda_{k}^{*} \phi_{k}(t) > 0, \quad a < t < b,
\]

(4)

and function

\[
f_{0}(t) = \left( \sum_{k=0}^{m} \lambda_{k}^{*} \phi_{k}(t) \right)^{\frac{1}{\kappa-1}}, \quad a < t < b,
\]

(5)

satisfies the constraints in (2), then \( f_{0} \) solves (2).
With (5), the constraints in (2) yield a system for \( \lambda_0^*, \lambda_1^*, \ldots, \lambda_m^* \):

\[
\int_a^b \left( \sum_{k=0}^{m} \lambda_k^* \phi_k(t) \right)^{1/\kappa} \phi_i(t) dt = \mu_i, \quad i = 0, 1, \ldots, m. \tag{6}
\]

_Generalized Pareto Distribution (GPD)_

For arbitrary \( a, b = \infty, m = 1, \phi_1(t) = t, \mu_1 = \mu > a, \) and \( \frac{1}{2} < \kappa < 1 \), the solution of (2) is determined by

\[
f(t) = \frac{\kappa}{2\kappa - 1} \frac{1}{\mu - a} \left( 1 + \frac{1 - \kappa}{2\kappa - 1} \frac{t - a}{\mu - a} \right)^{1/\kappa}, \quad a < t. \tag{7}
\]

GPD is standardly specified by three parameters: location \( \eta \), scale \( \sigma \), and shape \( \xi \). The probability density function is

\[
f_{\xi, \eta, \sigma}(t) = \frac{1}{\sigma} \left( 1 + \frac{\xi(t - \eta)}{\sigma} \right)^{-(\frac{1}{\xi} - 1)}.\]

Formulas for translation of our parameters to standard GPD parameters: \( \eta = a, \xi = \frac{1 - \kappa}{\kappa}, \sigma = \frac{(2\kappa - 1)(\mu - a)}{\kappa} \).

**Detail.** Substitution \( t = s + a \) reduced the problem to (2) with \( a' = 0, b = \infty, m = 1, \phi_1(s) = s, \) and \( \mu_1 = \mu - a \). In this case, condition (4) simplifies to

\[
\lambda_0 + \lambda_1 s > 0, \quad 0 < s,
\]

and holds provided that \( \lambda_0 \geq 0 \) and \( \lambda_1 > 0 \). System (6) takes the form

\[
\frac{1 - \kappa}{\kappa \lambda_1} \lambda_0^{\frac{\kappa - 1}{\kappa - 1}} = 1, \quad \frac{(1 - \kappa)^2}{\lambda_1 \kappa (2\kappa - 1)} \lambda_0^{\frac{2\kappa - 1}{\kappa - 1}} = \mu - a,
\]

which has a closed-form solution:

\[
\lambda_0 = \left( \frac{\kappa}{(2\kappa - 1)(\mu - a)} \right)^{\kappa - 1}, \quad \lambda_1 = \frac{1 - \kappa}{\kappa \left( \frac{\kappa}{(2\kappa - 1)(\mu - a)} \right)^{\kappa}},
\]

and (5) yields (7).

The PDF (7) has three parameters:

1. \( a \) = location parameter = beginning of the tail
2. \( \mu \) = scale parameter = mean value of the tail
3. \( \kappa \) = shape parameter

Usually, it is easy to estimate, \( a \) and \( \mu \), and some effort is needed for estimation of \( \kappa \). One of the approaches to estimate \( \kappa \) is by likelihood maximization.

**Maximum Likelihood: Estimation of \( \kappa \)**
Suppose there is a sample \( x_1, \ldots, x_n \) of \( n \) independent and identically distributed observations drawn from the PDF (7), in which \( a \) and \( \mu \) are given. The likelihood function with (7) is determined by

\[
\ell(\kappa; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\kappa) = \left( \frac{1}{\mu - a} \right)^n \left( \frac{\kappa}{2\kappa - 1} \right)^n \prod_{i=1}^{n} \left( 1 + \frac{1 - \kappa x_i - a}{2\kappa - 1} \mu - a \right)^{-\frac{1}{\kappa}},
\]

and \( \kappa \) is estimated by the solution of likelihood maximization

\[
\hat{\kappa} \in \arg \max_{\frac{1}{2} < \kappa < 1} \ell(\kappa; x_1, \ldots, x_n).
\]

“Harmonic” Method: Estimation of \( \kappa \)

Find \( \kappa \) from the condition

\[
\int_{a}^{\infty} \ln(t - a) f(t) \, dt = \hat{C},
\]

where \( \hat{C} = \frac{1}{n} \sum_{i=1}^{n} \ln(x_i - a) \) is estimated as the expected value of \( \ln(t - a) \) from observations.

Denote \( \eta = \mu - a, s = t - a, \)

\[
g(s) = f(s + a) = \frac{\kappa}{2\kappa - 1} \eta \left( 1 + \frac{s}{\eta} \right)^\frac{1}{\kappa - 1}, \quad z = \frac{2\kappa - 1}{1 - \kappa},
\]

where \( f(t) \) is given by (7). If \( \eta > 0 \) and \( \frac{1}{2} < \kappa < 1 \), then \( z \in (0, +\infty) \) and

\[
F(\kappa, \eta) \equiv \int_{0}^{\infty} g(s) \ln s \, ds = -H_z + \ln z + \ln \eta, \quad H_z = \int_{0}^{1} \frac{1 - t^z}{1 - t} \, dt,
\]

where \( H_z \) is the harmonic number [4]. The condition (8) can be written as

\[
p(z) = \tilde{C},
\]

where \( p(z) = H_z - \ln z, and \tilde{C} = \ln \eta - C. \)

Numerical Experiments

To illustrate the two approaches for estimating \( \kappa \), 100 samples with sample size of 1250 and 100 samples with sample size of 12500 are sampled from a GPD distribution with \( a = 0, \eta = 0.7778 \text{ and } \kappa = 0.9091 \). In this case, \( \eta \) is estimated as the average value of each sample. Table 1 shows the minimum, mean and maximum values as well as the deviation = (maximum - minimum) of the relative errors \( (\kappa_0 - \hat{\kappa}) / \kappa_0 \text{ and } (\eta_0 - \hat{\eta}) / \eta_0 \) between the true and estimated values of \( \kappa \) and \( \eta \) over 100 samples with sample size of 1250 and over 100 samples with sample size of 12500.

Relative errors from Table 1 show, that the estimates obtained by the harmonic method are better on average than the ML estimates, although they have a larger deviation for different generated samples. Note, that the harmonic method uses only a few derivative characteristics of the sample (the logarithm of the average sample and the average of the logarithms of the sample). In general, the results obtained by both methods gave similar results with low relative errors.

Since GPD distribution is often used to model the tails of another distribution, the important problem is to estimate distribution parameters not for all samples, but only for tail part of them. In the Part 2 we consider problems of parameters estimation of GPD for distributions tails.
Table 1: Results of the First Numerical Experiments

<table>
<thead>
<tr>
<th>Method</th>
<th>Tail Size</th>
<th>Parameter</th>
<th>min (%)</th>
<th>mean (%)</th>
<th>max (%)</th>
<th>deviation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>1,250</td>
<td>η</td>
<td>-10.82</td>
<td>-0.28</td>
<td>5.49</td>
<td>16.32</td>
</tr>
<tr>
<td>Harmonik</td>
<td>1,250</td>
<td>κ</td>
<td>-8.9</td>
<td>-0.36</td>
<td>8.81</td>
<td>17.71</td>
</tr>
<tr>
<td>ML</td>
<td>12,500</td>
<td>η</td>
<td>-2.42</td>
<td>-0.08</td>
<td>2.01</td>
<td>4.44</td>
</tr>
<tr>
<td>Harmonik</td>
<td>12,500</td>
<td>κ</td>
<td>-1.79</td>
<td>-0.03</td>
<td>1.98</td>
<td>3.78</td>
</tr>
</tbody>
</table>

Part 2: Estimation of Conditional Distribution Tail with Quantile and CVaR regression

Classical least squares regression is a widely used statistical procedure for modeling the relationship between several independent (or explanatory) variables, \( \theta_1, \ldots, \theta_I \), and a dependent (or response) random variable, \( \theta_0 \). Independent random variables are frequently called factors. Let \( \theta_{10}, \ldots, \theta_{J0} \) be \( J \) observations of the dependent random value \( \theta_0 \), and \( \theta_{1i}, \ldots, \theta_{ji} \) be \( J \) corresponding observations of independent variables \( \theta_i, i = 1, \ldots, I \). Observations of the dependent and independent variables are usually placed in the following Extended Design Matrix.

Table 2: Extended Design Matrix

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variables (factors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_{10} )</td>
<td>( \theta_{11} ) ( \theta_{12} ) ( \ldots ) ( \theta_{1I} )</td>
</tr>
<tr>
<td>( \theta_{20} )</td>
<td>( \theta_{21} ) ( \theta_{22} ) ( \ldots ) ( \theta_{2I} )</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots ) ( \ldots ) ( \ldots ) ( \ldots )</td>
</tr>
<tr>
<td>( \theta_{J0} )</td>
<td>( \theta_{J1} ) ( \theta_{J2} ) ( \ldots ) ( \theta_{JI} )</td>
</tr>
</tbody>
</table>

Standardly, observations from the dependent variable, \( \theta_{10}, \ldots, \theta_{J0} \), are estimated by a linear combination of independent variables,

\[
\hat{\theta}_{j0} = x_0 + x_1 \hat{\theta}_{j1} + \cdots + x_I \hat{\theta}_{jI} + \epsilon_j, \quad j = 1, \ldots, J,
\]

where \( x_0 \) is an intercept, and \( x_1, \ldots, x_I \) are regression coefficients. Terms \( \epsilon_j \) are observations of random value accounting for the surplus variability or scatter in \( \theta_0 \) that cannot be explained by \( \theta_1, \ldots, \theta_I \).

Conditional Tail Estimation: Quantile Regression

Classical regression (with Mean-Square Error) estimates mean value of the dependent variable. The quantile regression estimates any quantile of the distribution of the dependent variable, rather than a single mean value. We use \( \alpha \)-quantile regression for building tail distribution, which is approximated by GPD distribution. Consider the problem of GPD parameters estimation.

Let \( \hat{x}_0 \) and \( \hat{x}_i, i = 1, \ldots, I \) be optimal estimates of coefficients for \( \alpha \)-quantile regression. Given these coefficients calculate residuals

\[
\hat{\epsilon}_j = \theta_{j0} - (\hat{x}_0 + \sum_{i=1}^{I} \hat{\theta}_{ji} \hat{x}_i), \quad j = 1, \ldots, J
\]
and choose only positive (tail) residuals. Using positive residuals, estimate GPD parameter $\eta$ as their average, and estimate GPD parameter $\varkappa$ using MLE or harmonic method. Let $(\tilde{\theta}_1, \ldots, \tilde{\theta}_I)$ be new observation of independent variables. Then conditional tail GPD distribution begins from the estimate of $\alpha$-quantile, $\hat{x}_0 + \sum_{i=1}^I \tilde{\theta}_i \hat{x}_i$.

To demonstrate this approach we generated random samples using the following regression

$$\theta_{0j} = a + b \theta_{1j} + \epsilon_j, \quad j = 1, \ldots, J$$

where

- $a, b$ are constants, $a = 5, b = 4$;
- $\theta_{1j}$ is sampled from uniform on interval $[0, 1]$;
- $\epsilon_j$ is sampled from $N(0, 1)$.

Let $q_a$ be $\alpha$-quantile of sampled residuals $\epsilon_j$ and $\hat{\epsilon}_1, \ldots, \hat{\epsilon}_{(1-\alpha)J}$ be GPD random $(1 - \alpha)J$ samples with $a = q_a, \mu = q_a + 0.7778, \varkappa = 0.9091$. Replace the largest $(1 - \alpha)J$ residuals $\epsilon_j$ by $\hat{\epsilon}_1, \ldots, \hat{\epsilon}_{(1-\alpha)J}$ and randomly mix the new set of residuals. Denote the new set of residuals by $\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_{(1-\alpha)J}$.

As a result we obtain new random sample:

$$\tilde{\theta}_{0j} = a + b \theta_{1j} + \tilde{\epsilon}_j, \quad j = 1, \ldots, J.$$ 

This model was used to generate 100 samples with sample size of 5,000 and 100 samples with sample size of 50,000. In this case, $\eta$ was estimated as the average value of each sample, and $\varkappa$ was estimated by MLE and harmonic methods. Table 3 shows the minimum, mean and maximum values as well as the deviation = (maximum - minimum) of the relative errors $(\varkappa_0 - \hat{\varkappa})/\varkappa_0$ and $(\eta_0 - \hat{\eta})/\eta_0$ between the true and estimated values of $\varkappa$ and $\eta$ over 100 samples with sample size of 5,000 and over 100 samples with sample size of 50,000.

<table>
<thead>
<tr>
<th>Method</th>
<th>Tail Size</th>
<th>Parameter</th>
<th>min (%)</th>
<th>mean (%)</th>
<th>max (%)</th>
<th>deviation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>1,250</td>
<td>$\eta$</td>
<td>-11.05</td>
<td>-0.41</td>
<td>5.64</td>
<td>16.7</td>
</tr>
<tr>
<td>Harmonik</td>
<td>1,250</td>
<td>$\varkappa$</td>
<td>-9.19</td>
<td>-0.51</td>
<td>8.39</td>
<td>17.58</td>
</tr>
<tr>
<td>ML</td>
<td>12,500</td>
<td>$\eta$</td>
<td>-2.38</td>
<td>-0.11</td>
<td>1.59</td>
<td>4.33</td>
</tr>
<tr>
<td>Harmonik</td>
<td>12,500</td>
<td>$\varkappa$</td>
<td>-1.86</td>
<td>-0.06</td>
<td>1.95</td>
<td>3.81</td>
</tr>
</tbody>
</table>

**Conditional Tail Estimation: CVaR Regression**

We use CVaR regression for building tail distribution, which is approximated by GPD distribution. Consider the problem of GPD parameters estimation.

Run CVaR regression. Let $\hat{x}_0$ and $\hat{x}_i, i = 1, \ldots, I$, be optimal estimates of coefficients for CVaR regression. After that, we calculate residuals

$$\hat{\epsilon}_j = \theta_{j0} - (\hat{x}_0 + \sum_{i=1}^I \theta_{ji} \hat{x}_i), \quad j = 1, \ldots, J$$
choose only positive (tail) residuals, and estimate GPD for the set of positive residuals. Then conditional tail GDP distribution begins from the estimate of $\alpha$-CVaR, $\hat{x}_0 + \sum_{i=1}^{l} \hat{\theta}_i \hat{x}_i$.

**Case Study**

We applied the described above procedures for parameter estimation of tail distribution of the residuals of $\alpha$-quantile regression and $CVaR_\alpha$ regression for real data. We considered the return-based style classification of a mutual fund. The procedure regresses fund return by several indices as explanatory variables. We considered the $\alpha = 0.75$ quantile and $CVaR_\alpha$ regression of gains (returns) and losses (returns with minus sign) of Fidelity Magellan Fund with Russell Value Index (RUI), RUSSELL 1000 VALUE INDEX (RLV), Russell 2000 Growth Index (RUO) and Russell 1000 Growth Index (RLG). Number of daily returns = 1264. We used two methods for estimating GPD $\kappa$ parameter: maximum likelihood and harmonic. Estimates of GPD parameters are shown in Table 5.

<table>
<thead>
<tr>
<th>Method</th>
<th>Tail</th>
<th>$\eta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantile regression</td>
<td>Right</td>
<td>0.0041</td>
</tr>
<tr>
<td>Quantile regression</td>
<td>Left</td>
<td>0.0045</td>
</tr>
<tr>
<td>CVaR regression</td>
<td>Right</td>
<td>0.0040</td>
</tr>
<tr>
<td>CVaR regression</td>
<td>Left</td>
<td>0.0043</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Tail</th>
<th>Estimator</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantile regression</td>
<td>Right</td>
<td>MLE</td>
<td>0.909</td>
</tr>
<tr>
<td>Quantile regression</td>
<td>Left</td>
<td>MLE</td>
<td>1.021</td>
</tr>
<tr>
<td>CVaR regression</td>
<td>Right</td>
<td>MLE</td>
<td>0.808</td>
</tr>
<tr>
<td>CVaR regression</td>
<td>Left</td>
<td>MLE</td>
<td>1.055</td>
</tr>
<tr>
<td>Quantile regression</td>
<td>Right</td>
<td>Harmonic</td>
<td>0.791</td>
</tr>
<tr>
<td>Quantile regression</td>
<td>Left</td>
<td>Harmonic</td>
<td>0.727</td>
</tr>
<tr>
<td>CVaR regression</td>
<td>Right</td>
<td>Harmonic</td>
<td>0.859</td>
</tr>
<tr>
<td>CVaR regression</td>
<td>Left</td>
<td>Harmonic</td>
<td>0.976</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Tail</th>
<th>Estimator</th>
<th>$\xi$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantile regression</td>
<td>Right</td>
<td>MLE</td>
<td>0.099</td>
<td>0.003</td>
</tr>
<tr>
<td>Quantile regression</td>
<td>Left</td>
<td>MLE</td>
<td>-0.02</td>
<td>0.004</td>
</tr>
<tr>
<td>CVaR regression</td>
<td>Right</td>
<td>MLE</td>
<td>0.236</td>
<td>0.003</td>
</tr>
<tr>
<td>CVaR regression</td>
<td>Left</td>
<td>MLE</td>
<td>-0.052</td>
<td>0.004</td>
</tr>
<tr>
<td>Quantile regression</td>
<td>Right</td>
<td>Harmonic</td>
<td>0.263</td>
<td>0.003</td>
</tr>
<tr>
<td>Quantile regression</td>
<td>Left</td>
<td>Harmonic</td>
<td>0.375</td>
<td>0.002</td>
</tr>
<tr>
<td>CVaR regression</td>
<td>Right</td>
<td>Harmonic</td>
<td>0.163</td>
<td>0.003</td>
</tr>
<tr>
<td>CVaR regression</td>
<td>Left</td>
<td>Harmonic</td>
<td>0.024</td>
<td>0.004</td>
</tr>
</tbody>
</table>
Conditional tails GPD are presented in the following graphs.

![Figure 1: Conditional Right GPD Tails: MLE Method](image1)

Left (right) figure is based on 0.75-VaR (0.75-CVaR) regression of returns of right tail of the Magellan fund; GPD parameter $\kappa$ for residuals is calibrated with MLE method; GPD red curve starts from 0.75-VaR (0.75-CVaR) of fund return; GPD blue curve starts from the estimate of 0.75-VaR (0.75-CVaR) for a new observation.

![Figure 2: Conditional Left GPD Tails: MLE Method](image2)

Left (right) figure is based on 0.75-VaR (0.75-CVaR) regression of returns of left tail of the Magellan fund; GPD parameter $\kappa$ for residuals is calibrated with MLE method; GPD red curve starts from 0.75-VaR (0.75-CVaR) of fund return; GPD blue curve starts from the estimate of 0.75-VaR (0.75-CVaR) for a new observation.

![Figure 3: Conditional Right GPD Tail: Harmonic Method](image3)

Figure 3: Conditional Right GPD Tail: Harmonic Method.
Left (right) figure is based on 0.75-VaR (0.75-CVaR) regression of returns of right tail of the Magellan fund; GPD parameter $\kappa$ for residuals is calibrated with harmonic method; GPD red curve starts from 0.75-VaR (0.75-CVaR) of fund return; GPD blue curve starts from the estimate of 0.75-VaR (0.75-CVaR) for a new observation.

Figure 4: Conditional Left GPD Tail: Harmonic Method

Left (right) figure is based on 0.75-VaR (0.75-CVaR) regression of returns of left tail of the Magellan fund; GPD parameter $\kappa$ for residuals is calibrated with harmonic method; GPD red curve starts from 0.75-VaR (0.75-CVaR) of fund return; GPD blue curve starts from the estimate of 0.75-VaR (0.75-CVaR) for a new observation.

References


