Nonparametric extremal quantile regression

Stéphane Girard

INRIA Rhône-Alpes & LJK
http://mistis.inrialpes.fr/people/girard/

July, 2013

joint work with Abdelaati Daouia (Catholic University of Louvain) and Laurent Gardes (Université de Strasbourg)

Nonparametric quantile regression

- Conditional quantiles: Given a vector of regressors $X \in \mathbb{R}^p$ and a response variable $Y \in \mathbb{R}$, define

$$q(\alpha|x) = \bar{F}^{\leftarrow}(\alpha|x) = \inf\{y,\ \bar{F}(y|x) \le \alpha\}$$

for $\alpha \in (0,1)$, where $\bar{F}(y|x) = \mathbb{P}(Y > y|X = x)$.

- Nonparametric estimators: Let (X_i, Y_i), $i = 1, \ldots, n$ be iid copies of (X, Y).

$$\begin{aligned} \hat{q}_n(\alpha|x) &= \hat{\bar{F}}_n^{\leftarrow}(\alpha|x) = \inf\{y,\ \hat{\bar{F}}_n(y|x) \le \alpha\} \\ \hat{\bar{F}}_n(y|x) &= \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \mathbb{I}\{Y_i > y\} \Big/ \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \end{aligned}$$

where the kernel K is a bounded pdf on $\mathbb{R}^p$ with support included in the unit ball, and $h_n \to 0$ is the window-width.

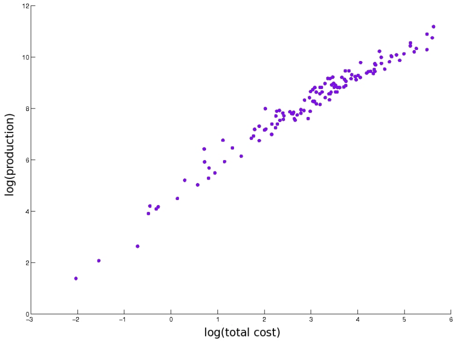
Extremal quantile regression

Objective: Extending the asymptotics further into the tails of the conditional distribution by considering $\alpha = \alpha_n \to 0$ as $n \to \infty$.

- Is it still reasonable to use $\hat{q}_n(\alpha_n|x)$ for estimating $q(\alpha_n|x)$ when $\alpha_n \to 0$?
- If not, how can we adapt $\hat{q}_n(\alpha_n|x)$ to this situation?

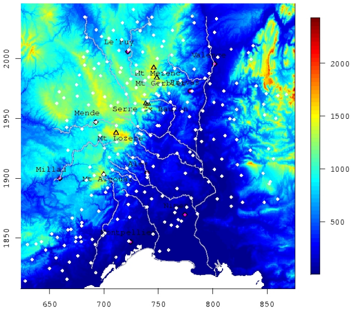
Extremal quantile regression

Motivating example 1: American electric utility companies ($p = 1$, $n = 123$).



Goal: Identification of the set of the most efficient firms.

Motivating example 2: Extreme rainfall as a function of the geographical location ($p = 3$), south of France.



$X = \{$longitude, lattitude, altitude$\}$, $Y =$rainfall (mm).
Goal: Estimation of the 100-year return level.

Related works:

- Parametric models for exceedances over high thresholds:
  Davison and Smith (1990), Smith (1989)
- Semi-parametric approaches:
  Hall and Tajvidi (2000), Beirlant and Goegebeur (2003)
- Extreme quantiles in the linear regression model:
  Chernozhukov (2005), Jurecková (2007), Wang et al. (2012)
- Local polynomial fitting of the GEV distribution:
  Davison and Ramesh (2000)
- Local polynomial fitting of a Generalized Pareto Distribution:
  Beirlant and Goegebeur (2004)
- Spline estimators fitted by maximum penalized likelihood:
  Chavez-Demoulin and Davison (2005)

Assumption: von-Mises condition

The function $\bar{F}(\cdot|x)$ is twice differentiable and

$$\lim_{y \uparrow y_F(x)} \frac{\bar{F}(y|x)\bar{F}''(y|x)}{(\bar{F}')^2(y|x)} = \gamma(x) + 1$$

where

- $y_F(x) = q(0|x) \in (-\infty, \infty]$ is the endpoint of Y given $X = x$,
- $\gamma(x)$ is the conditional extreme-value index,
- $\bar{F}'(\cdot|x)$ and $\bar{F}''(\cdot|x)$ are respectively the first and the second derivatives of $\bar{F}(\cdot|x)$.

This condition implies the existence of an auxiliary function $a(\cdot|x)$ such that, for all $t > 0$ as $\alpha \to 0$:

$$\frac{q(t\alpha|x) - q(\alpha|x)}{a(q(\alpha|x)|x)} \to K_{\gamma(x)}(1/t) := \int_1^{1/t} v^{\gamma(x)-1} dv.$$

Regularity assumptions

Notations

- g is the probability density function of X,
- $d(x, x')$ is the Euclidean distance between x and x' in $\mathbb{R}^p$,
- $B(x, h_n)$ is the ball centered at x with radius h_n.

The oscillation of the conditional survival function is controlled by

$$\Delta_\kappa(x, \alpha_n) := \sup_{(x',\beta)\in B(x,h_n)\times[\alpha_n\kappa,\alpha_n]} \left| \frac{\bar{F}(q(\beta|x)|x')}{\beta} - 1 \right|$$

where $\kappa \in (0, 1)$.

Assumption: $|g(x) - g(x')| \leq c_g d(x, x')$, where $c_g > 0$.

Asymptotic normality of $\hat{q}_n(\alpha_n|x)$

Theorem 1

Let $0 < \tau_J < \cdots < \tau_2 < \tau_1 \leq 1$ where $J > 0$ and let $x \in \mathbb{R}^p$ such that $g(x) > 0$.
If $\alpha_n \to 0$ and there exists $\kappa \in (0, \tau_J)$ such that

$$nh_n^p \alpha_n \to \infty, \;\; nh^p \alpha_n (h_n \vee \Delta_\kappa(x, \alpha_n))^2 \to 0,$$

then

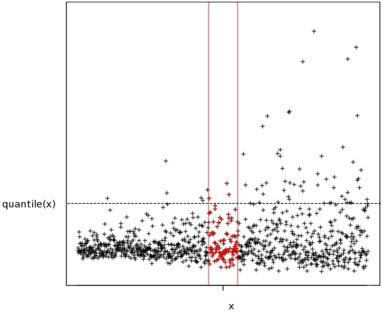
$$\left\{ f(q(\alpha_n|x)|x) \sqrt{nh_n^p \alpha_n^{-1}} \left(\hat{q}_n(\tau_j \alpha_n|x) - q(\tau_j \alpha_n|x) \right) \right\}_{j=1,\ldots,J}$$

is asymptotically Gaussian, centered, with covariance matrix $\|K\|_2^2/g(x)\Sigma(x)$ where $\Sigma_{j,j'}(x) = (\tau_j \tau_{j'})^{-\gamma(x)} \tau_{j \wedge j'}^{-1}$.

Analog of Berlinet et al.(2001), Theorem 6.4.

Conditions on the sequences α_n and h_n

$nh_n^p\alpha_n \to \infty$: Necessary and sufficient condition for the almost sure presence of at least one point in the region $B(x, h_n) \times [q(\alpha_n|x), +\infty)$ of $\mathbb{R}^p \times \mathbb{R}$.



$nh^p\alpha_n(h \vee \Delta_\kappa(x, \alpha_n))^2 \to 0$: The biais induced by the smoothing in negligible compared to the variance.

Asymptotic normality of $\hat{q}_n(\alpha_n|x)$

Corollary

1) Under the assumptions of Theorem 1,

$$\left\{ \sqrt{nh^p\alpha_n} \frac{q(\alpha_n|x)}{a(q(\alpha_n|x)|x)} \left(\frac{\hat{q}_n(\tau_j\alpha_n|x)}{q(\tau_j\alpha_n|x)} - 1 \right) \right\}_{j=1,\dots,J}$$

is asymptotically Gaussian, centered, with covariance matrix $\|K\|_2^2/g(x)\tilde{\Sigma}(x)$ where $\tilde{\Sigma}_{j,j'}(x) = (\tau_j\tau_{j'})^{-(\gamma(x)\wedge 0)}\tau_{j\wedge j'}^{-1}$.

2) If, moreover, $\gamma(x) > 0$, then

$$\left\{ \sqrt{nh^p\alpha_n} \left(\frac{\hat{q}_n(\tau_j\alpha_n|x)}{q(\tau_j\alpha_n|x)} - 1 \right) \right\}_{j=1,\dots,J}$$

is asymptotically Gaussian, centered, with covariance matrix $\|K\|_2^2\gamma^2(x)/g(x)\tilde{\Sigma}(x)$ where $\tilde{\Sigma}_{j,j'}(x) = \tau_{j\wedge j'}^{-1}$.

Regression quantiles of higher order

- **Goal:** Estimate $q(\beta_n|x)$ with $\beta_n/\alpha_n \to 0$ and $nh^p\alpha_n \to \infty$.
- **Idea:** The von-Mises condition implies that

$$b(t,\alpha|x) := \frac{q(t\alpha|x) - q(\alpha|x)}{a(q(\alpha|x)|x)} - K_{\gamma(x)}(1/t) \to 0$$

for all $t > 0$ as $\alpha \to 0$.

This allows to build a new estimator of $q(\beta_n|x)$:

$$\tilde{q}_n(\beta_n|x) = \hat{q}_n(\alpha_n|x) + K_{\hat{\gamma}_n(x)}(\alpha_n/\beta_n)\, \hat{a}_n(x).$$

where $\hat{\gamma}_n(x)$ and $\hat{a}_n(x)$ are two estimators of $\gamma(x)$ and $a(q(\alpha_n|x)|x)$ respectively.

Asymptotic normality of $\tilde{q}_n(\beta_n|x)$

Theorem 2

Let $\alpha_n \to 0$, $\beta_n/\alpha_n \to 0$ and suppose there exists $\Lambda_n \to 0$ such that $\Lambda_n \log(\alpha_n/\beta_n) \to 0$, $\Lambda_n^{-1} b(\beta_n/\alpha_n, \alpha_n|x)/K'_{\gamma(x)}(\alpha_n/\beta_n) \to 0$ and

$$\Lambda_n^{-1}\left(\hat{\gamma}_n(x) - \gamma(x), \frac{\hat{a}_n(x)}{a(q(\alpha_n|x)|x)} - 1, \frac{\hat{q}_n(\alpha_n|x) - q(\alpha_n|x)}{a(q(\alpha_n|x)|x)}\right)^t$$

converges in distribution to $\zeta(x)$ a $\mathbb{R}^3$ random vector. Then,

$$\Lambda_n^{-1}\left(\frac{\tilde{q}_n(\beta_n|x) - q(\beta_n|x)}{a(q(\alpha_n|x)|x)K'_{\gamma(x)}(\alpha_n/\beta_n)}\right) \xrightarrow{d} c(x)^t \zeta(x)$$

where $c(x)^t = \left(1, -(\gamma(x) \wedge 0), (\gamma(x) \wedge 0)^2\right)$.

Analog of de Haan and Ferreira (2006), Theorem 4.3.1.

Illustration: Refined Pickands estimators

Estimators of $\gamma(x)$ and $a(q(\alpha_n|x)|x)$ adapted from Drees (1995):

$$\hat{\gamma}_n^{\mathrm{RP}}(x) = \frac{1}{\log r} \sum_{j=1}^{J-2} \pi_j \log \left(\frac{\hat{q}_n(\tau_j \alpha_n|x) - \hat{q}_n(\tau_{j+1}\alpha_n|x)}{\hat{q}_n(\tau_{j+1}\alpha_n|x) - \hat{q}_n(\tau_{j+2}\alpha_n|x)} \right)$$

$$\hat{a}_n^{\mathrm{RP}}(x) = \frac{1}{K_{\hat{\gamma}_n^{\mathrm{RP}}(x)}(r)} \sum_{j=1}^{J-2} \pi_j \, r^{j\hat{\gamma}_n^{\mathrm{RP}}(x)} (\hat{q}_n(\tau_j \alpha_n|x) - \hat{q}_n(\tau_{j+1}\alpha_n|x))$$

where

- $\tau_j = r^{j-1}$, $j = 1, \dots, J$, with $J \geq 3$ and $r \in (0,1)$.
- (π_j) is a sequence of weights summing to one.

Theorem 2 holds with $\Lambda_n^{-1} = \sqrt{nh^p \alpha_n}$.

Comparison of 3 estimators

- Kernel estimator without extrapolation $\hat{q}_n(\beta_n|x)$,
- Kernel estimator with extrapolation $\tilde{q}_n^{\mathrm{RP}}(\beta_n|x)$ based on Refined Pickands estimators ($J = 3$ and $r = 1/3$).

$$\hat{\gamma}_n^{\mathrm{RP}}(x) = \frac{1}{\log r} \log \left(\frac{\hat{q}_n(\alpha_n|x) - \hat{q}_n(r\alpha_n|x)}{\hat{q}_n(r\alpha_n|x) - \hat{q}_n(r^2\alpha_n|x)} \right)$$

$$\hat{a}_n^{\mathrm{RP}}(x) = \frac{r^{\hat{\gamma}_n^{\mathrm{RP}}(x)}}{K_{\hat{\gamma}_n^{\mathrm{RP}}(x)}(r)} (\hat{q}_n(\alpha_n|x) - \hat{q}_n(r\alpha_n|x))$$

- Local polynomial fitting of a Generalized Pareto Distribution (GPD): Beirlant and Goegebeur (2004)

Local polynomial fitting of a GPD

Let $x \in \mathbb{R}^p$, let h be bandwidth and k_{xh} a number of exceedances.

- Compute $Y^x_{1,n_{xh}} \leq \cdots \leq Y^x_{n_{xh},n_{xh}}$ the order statistics corresponding to the n_{xh} values Y_i for which $X_i \in B(x,h)$.
- Compute the exceedances $Z^x_i = Y^x_{n_{xh}-i+1,n_{xh}} - Y^x_{n_{xh}-k_{xh},n_{xh}}$, $i = 1, \ldots, k_{xh}$.
- Fit a GPD density $\tilde{g}(., \sigma, \gamma)$ by maximizing

$$\sum_{i=1}^{k_{xh}} \log \tilde{g}\left(Z^x_i; \sum_{j=0}^{p_1} \beta_{1j}(X_i - x)^j, \sum_{j=0}^{p_2} \beta_{2j}(X_i - x)^j\right) K\left(\frac{X_i - x}{h}\right)$$

w.r.t $(\beta_{10}, \cdots, \beta_{1p_1}, \beta_{20}, \cdots, \beta_{2p_2})$ to get the estimates

$$\hat{q}^{\text{GPD}}_n(\beta_n|x) = Y^x_{n_{xh}-k_{xh},n_{xh}} + \frac{\hat{\beta}_{10}}{\hat{\beta}_{20}}\left[\left(\frac{n_{xh}\,\beta_n}{k_{xh}}\right)^{-\hat{\beta}_{20}} - 1\right]$$

Tuning parameters

- Number of upper order statistics k_{xh} in $\hat{q}_n^{\text{GPD}}(\beta_n|x)$ and $\alpha_n := k_{xh}/n_{xh}$ in $\tilde{q}_n^{\text{RP}}(\beta_n|x)$.
- Smoothing parameter h for all the considered estimators $\hat{q}_n(\beta_n|x)$, $\tilde{q}_n^{\text{RP}}(\beta_n|x)$ and $\hat{q}_n^{\text{GPD}}(\beta_n|x)$.

For all the considered estimators:

- The tuning parameters h and k_{xh} were selected by minimizing the MSE estimated on 400 replications.
- The kernel function was chosen to be

$$K(t) = \frac{35}{32}(1 - t^2)^3 \mathbb{I}\{-1 \leq t \leq 1\}.$$

Simulated model

Simulations from the model of Ruppert et al. (2003):

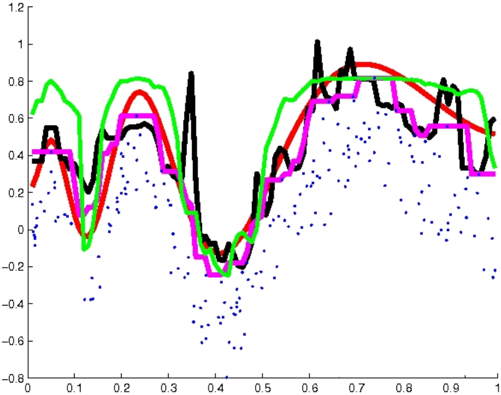
$$Y_i = \mu(X_i) + \sigma(X_i)\, U_i, \quad i = 1, \ldots, n.$$

where

- $n = 200$
- The X_i are i.i.d standard uniform ($p = 1$).
- The U_i given $X_i = x$ are independent **standard Gaussian**, **Beta**$(\nu(x), \nu(x))$ or **Student** $t_{[\nu(x)]+1}$ with

$$\begin{aligned}
\mu(x) &= \sqrt{x(1-x)}\,\sin\left(\frac{2\pi(1+2^{-7/5})}{x+2^{-7/5}}\right)\\
\sigma(x) &= (1+x)/10\\
\nu(x) &= \left\{\left(\frac{1}{10}+\sin(\pi x)\right)\left(\frac{11}{10}-\frac{1}{2}\exp\{-64(x-1/2)^2\}\right)\right\}^{-1}
\end{aligned}$$

Typical realization in case $Y|X$ is Gaussian and $\beta_n = 0.005$



Red: true quantile $q(\beta_n|x)$, Black: $\tilde{q}_n^{\mathrm{RP}}(\beta_n|x)$, Green: $\hat{q}_n^{\mathrm{GPD}}(\beta_n|x)$, Magenta: $\hat{q}_n(\beta_n|x)$.

Results

$\beta_n = 0.05$

	MSE			Bias		
	$\hat{q}_n^{\text{GPD}}(\beta_n\|x)$	$\tilde{q}_n^{\text{RP}}(\beta_n\|x)$	$\hat{q}_n(\beta_n\|x)$	$\hat{q}_n^{\text{GPD}}(\beta_n\|x)$	$\tilde{q}_n^{\text{RP}}(\beta_n\|x)$	$\hat{q}_n(\beta_n\|x)$
Gaussian	0.018	0.011	0.011	0.097	0.000	0.006
Student	0.135	0.031	0.077	0.153	-0.013	0.087
Beta	0.036	0.009	0.002	0.158	0.050	0.013

$\beta_n = 0.01$

	MSE			Bias		
	$\hat{q}_n^{\text{GPD}}(\beta_n\|x)$	$\tilde{q}_n^{\text{RP}}(\beta_n\|x)$	$\hat{q}_n(\beta_n\|x)$	$\hat{q}_n^{\text{GPD}}(\beta_n\|x)$	$\tilde{q}_n^{\text{RP}}(\beta_n\|x)$	$\hat{q}_n(\beta_n\|x)$
Gaussian	0.028	0.026	0.016	0.095	-0.078	-0.036
Student	0.692	0.111	0.682	0.089	-0.089	-0.096
Beta	0.066	0.014	0.003	0.207	0.052	0.021

$\beta_n = 0.005$

	MSE			Bias		
	$\hat{q}_n^{\text{GPD}}(\beta_n\|x)$	$\tilde{q}_n^{\text{RP}}(\beta_n\|x)$	$\hat{q}_n(\beta_n\|x)$	$\hat{q}_n^{\text{GPD}}(\beta_n\|x)$	$\tilde{q}_n^{\text{RP}}(\beta_n\|x)$	$\hat{q}_n(\beta_n\|x)$
Gaussian	0.031	0.035	0.020	0.086	-0.098	-0.052
Student	1.023	0.292	0.978	-0.045	-0.162	-0.260
Beta	0.079	0.016	0.004	0.224	0.054	0.024

Tuning parameters selection in practice

A heuristical approach for $\tilde{q}_n^{\text{RP}}(\beta_n|x)$.

Step 1. For each $h \in \mathcal{H}$ and $k = 1, \dots, n^\star_{xh} - 1$:

- Compute the estimator $\tilde{q}_n^{\text{RP}}(\beta_n|x; h, k)$,
- Compute the standard deviation of the estimates over a small window of successive values of k.

Select the value of k_{xh} as the minimizer of the standard deviation.

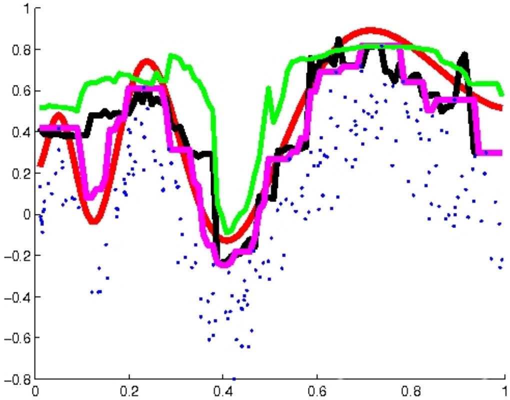
Step 2. For each $h \in \mathcal{H}$:

- Compute the estimator $\tilde{q}_n^{\text{RP}}(\beta_n|x; h, k_{xh})$,
- Compute the standard deviation of the estimates over a small window of successive values of h.

Select the value of h_x as the minimizer of the standard deviation.

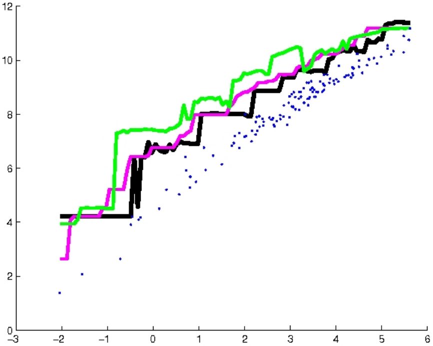
The same approach is adopted for $\hat{q}_n^{\text{GPD}}(\beta_n|x)$.

Typical realization in case $Y|X$ is Gaussian and $\beta_n = 0.005$



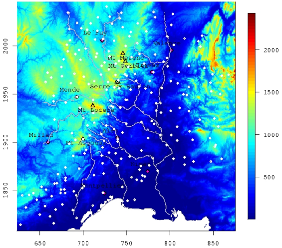
Red: true quantile $q(\beta_n|x)$, Black: $\tilde{q}_n^{\mathrm{RP}}(\beta_n|x)$, Green: $\hat{q}_n^{\mathrm{GPD}}(\beta_n|x)$, Magenta: $\hat{q}_n(\beta_n|x)$.

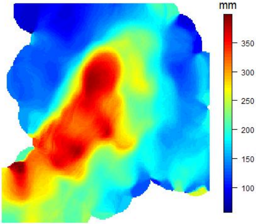
Data : $Y = \log(Q)$ with Q being the firm production output, $X = \log(C)$, with C being the total cost involved in the production. $n = 123$ firms.
Objective: Identification of the most efficient firms ($\beta_n = 1/n$).



Black: $\tilde{q}_n^{\mathrm{RP}}(\beta_n|x)$, Green: $\hat{q}_n^{\mathrm{GPD}}(\beta_n|x)$, Magenta: $\hat{q}_n(\beta_n|x)$.

Data. Y: Daily rainfalls (mm) measured at 523 raingauge stations from 1958 to 2000, X: Three dimensional geographical location (longitude, latitude and altitude).
Objective. Estimation of the 100-year return level.





Conclusion

- \+ Extremal quantile regression with fully nonparametric (kernel) methods,
- \+ Theoretical properties similar to the one-dimensional extreme-value theory,
- – Selection of tuning parameters is a difficult issue,
- – Curse of dimensionality when p is large.

References

- L. Gardes, S. Girard and A. Lekina. Functional nonparametric estimation of conditional extreme quantiles, *Journal of Multivariate Analysis*, 101, 419–433, 2010.
- L. Gardes and S. Girard. Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels, *Extremes*, 13, 177–204, 2010.
- A. Daouia, L. Gardes, S. Girard and A. Lekina. Kernel estimators of extreme level curves, *Test*, 20, 311–333, 2011.
- L. Gardes and S. Girard. Functional kernel estimators of large conditional quantiles, *Electronic Journal of Statistics*, 6, 1715–1744, 2012.
- A. Daouia, L. Gardes, and S. Girard. On kernel smoothing for extremal quantile regression, *Bernoulli*, 19, 2557–2589, 2013.
- J. El Methni, L. Gardes and S. Girard. Nonparametric estimation of extreme risks from conditional heavy-tailed distributions, *Scandinavian Journal of Statistics*, to appear, 2014.