Inference for extremal regression with dependent heavy-tailed data

Gilles Stupfler (University of Angers)

Joint work with Abdelaati Daouia (Toulouse School of Economics) and Antoine Usseglio-Carleve (Avignon Université)

30th birthday of the Laboratoire Manceau de Mathématiques
Le Mans, France, 24th May 2024

Motivating example: Tornado loss data

The data (NOAA Storm Prediction Center) contains, for each tornado in the eastern US between 1 January 2010 and 31 December 2019:

- The associated monetary loss,
- Starting and ending latitude and longitude,
- Length and width of the area traveled over.

We consider the loss per surface unit $Y = \text{loss}/(\text{length}\times\text{width})$, with covariate $\boldsymbol{X} \in \mathbb{R}^2$ being the average latitude and longitude of a tornado.

This results in a sample $(\boldsymbol{X}_t, Y_t)$ of size $n \approx 6{,}000$, including the 2011 Joplin, Missouri, tornado which caused a total loss of 2.8 billion USD.

Motivating question

What are the financial consequences of a high-impact tornado as a function of the geographical area it hits?

Tornado loss data - Exploratory analysis

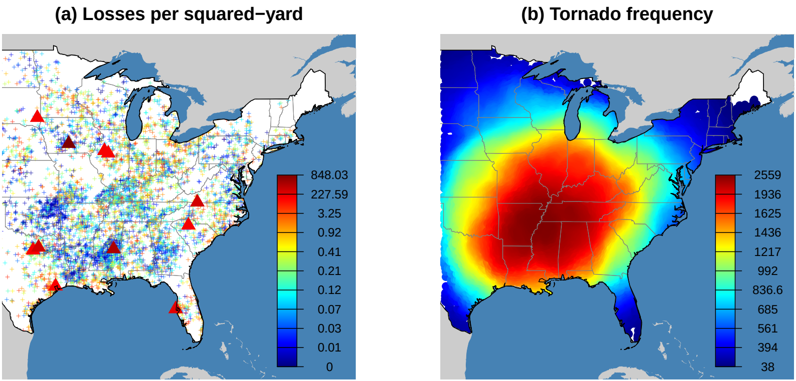
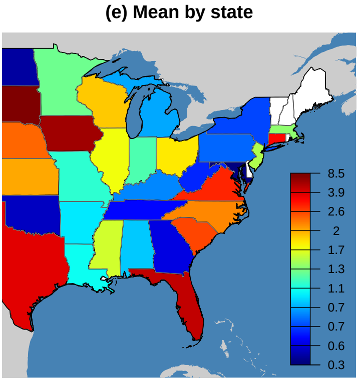


Figure: Tornado loss data. Left: Data across the eastern half of the US, right: Local number of observations.

Tornado loss data - State-by-state analysis



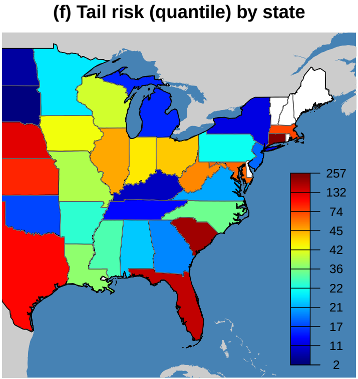


Figure: Tornado loss data. Left: Unconditional statewide estimation, using the sample average, right: Using the bias-reduced extreme quantile estimator of Gomes and Pestana (2007) at level 0.995.

Such region-by-region maps seem to be standard in insurance.

One major issue is that they are inherently non-smooth geographically: there are, among others, sharp jumps along the borders of

- The states in the Midwest,
- The states in New England,
- Florida-Georgia-South Carolina.

Inference, in the form of confidence intervals, is not obvious in this setting where data are dependent in time and nonstationary in space.

Context of the talk: conditional risk assessment

The (risk) variable Y is recorded alongside a random covariate $\boldsymbol{X} \in \mathbb{R}^p$:

- Y = individual stock price, $\boldsymbol{X}$ = stock market index $\in \mathbb{R}$,
- Y = loss variable following a climate event, $\boldsymbol{X}$ = latitude and longitude $\in \mathbb{R}^2$.

No assumption is made on the link between Y and $\boldsymbol{X}$, and we seek estimators of extreme quantiles and/or expectiles of Y given $\boldsymbol{X}$.

$\Rightarrow$ The estimation approach has to be nonparametric.

Difficulty: Quantiles and expectiles are not expectations. Otherwise, one would immediately think about nonparametric regression!

Two key (known) observations provide the basis for this work.

- The expectile of Y is in fact a quantile, of the c.d.f. E defined as

$$E(y) = 1 - \frac{\mathbb{E}((Y-y)\mathbb{1}\{Y>y\})}{2\,\mathbb{E}((Y-y)\mathbb{1}\{Y>y\}) + y - \mathbb{E}(Y)}.$$

  This function E is readily estimated.
- If q is the quantile function related to F then $q_\tau \le t \Leftrightarrow \tau \le F(t)$.

Both expectile and quantile estimation fall under the umbrella of quantile estimation (first point)...

... and therefore under the umbrella of c.d.f. estimation (second point). Indeed, if $\widehat{F}_n$ estimates F and $\widehat{q}_n$ is the left-continuous inverse of $\widehat{F}_n$,

$$\mathbb{P}(\widehat{q}_{n,\tau} \le t) = \mathbb{P}(\tau \le \widehat{F}_n(t)).$$

Position within the literature and contribution

Existing approaches to extreme conditional quantile and expectile estimation use kernel regression to estimate $F(\cdot|\boldsymbol{x})$ and $E(\cdot|\boldsymbol{x})$.

They assume i.i.d. data and do not discuss inference.

Main contribution of the paper

Study the estimation of, and Gaussian inference about, extreme conditional quantiles and expectiles for stationary $\alpha-$mixing data.

We do so in a conditional heavy (right) tail model.

Goal of the presentation

Give a flavor of the theory that can be obtained and briefly discuss a few data-generating processes that can be handled.

Estimation procedure - Intermediate levels

We concentrate on extreme quantile estimation.

Let $(\boldsymbol{X}_t, Y_t)$, $t \geq 1$, be strictly stationary, and assume that $\boldsymbol{X}$ has p.d.f. g. The conditional c.d.f. $F(y|\boldsymbol{x})$ of Y given $\boldsymbol{X} = \boldsymbol{x}$ is estimated by

$$\widehat{F}_n(y|\boldsymbol{x}) = \frac{1}{nh_n^p\,\widehat{g}_n(\boldsymbol{x})} \sum_{t=1}^{n} \mathbb{1}_{\{Y_t \leq y\}} K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_t}{h_n}\right)$$

$$\text{with } \widehat{g}_n(\boldsymbol{x}) = \frac{1}{nh_n^p} \sum_{t=1}^{n} K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_t}{h_n}\right).$$

Here K is a kernel p.d.f. on $\mathbb{R}^p$ and $h_n \to 0$. Set

$$\widehat{q}_n(\tau|\boldsymbol{x}) = \inf\left\{y \in \mathbb{R} \,|\, \widehat{F}_n(y|\boldsymbol{x}) \geq \tau\right\}.$$

This estimator was considered in *e.g.* Daouia *et al.* (2011, 2013). We first show that it is asymptotically normal at intermediate levels.

Model and assumptions

Condition $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ - Conditional heavy tails

There exist $\gamma(\boldsymbol{x}) > 0$, $\rho(\boldsymbol{x}) \leq 0$ and a positive or negative measurable function $A(\cdot|\boldsymbol{x})$ converging to 0 at infinity such that for any $y > 0$,

$$\lim_{s \to \infty} \frac{\overline{F}(sy|\boldsymbol{x})/\overline{F}(s|\boldsymbol{x}) - y^{-1/\gamma(\boldsymbol{x})}}{A(1/\overline{F}(s|\boldsymbol{x})|\boldsymbol{x})} = \frac{y^{-1/\gamma(\boldsymbol{x})}}{\gamma^2(\boldsymbol{x})} \int_1^y u^{\rho(\boldsymbol{x})/\gamma(\boldsymbol{x})-1} \, \mathrm{d}u$$

Condition $\mathcal{M}$ - Time dynamics

The data $(\boldsymbol{X}_t, Y_t)$, $t \geq 1$, form a stationary $\alpha-$mixing sequence of copies of $(\boldsymbol{X}, Y)$ satisfying assumption $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$.

Recall that $(\boldsymbol{X}_t, Y_t)$ is $\alpha-$mixing if, with $\mathcal{F}_a^b = \sigma(\{(\boldsymbol{X}_j, Y_j), a \leq j \leq b\})$,

$$\alpha(n) = \sup_{k \geq 1} \sup_{A \in \mathcal{F}_1^k} \sup_{B \in \mathcal{F}_{k+n}^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \to 0 \text{ as } n \to \infty.$$

Condition $\mathcal{A}(l_n, r_n)$ - Big-block/small-block

There exist sequences (l_n) and (r_n) such that $l_n \to \infty$, $r_n \to \infty$, $l_n/r_n \to 0$, $r_n/n \to 0$ and $n\,\alpha(l_n)/r_n \to 0$ as $n \to \infty$.

This ensures asymptotic independence of "big blocks" separated by "small blocks" of sample means.

We require regularity conditions on $\boldsymbol{X}$ and $Y|\boldsymbol{X}$. Fix a norm $\|\cdot\|$ on $\mathbb{R}^p$.

Condition $\mathcal{KS}$ - Kernel

K is bounded and symmetric (*i.e.* $K(\boldsymbol{u}) = K(-\boldsymbol{u})$) with a support contained in the unit closed $\|\cdot\|-$ball.

Condition $\mathcal{D}_g$ - Regularity

$g(\boldsymbol{x}) > 0$ and g is continuously differentiable in a neighborhood of $\boldsymbol{x}$ with a Lipschitz continuous gradient at $\boldsymbol{x}$.

Condition $\mathcal{D}_\omega$ - Variation in conditional extreme value behavior

For y large enough, the function $\overline{F}(y|\cdot)$ is differentiable at $\boldsymbol{x}$, the function $y \mapsto \nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x}) / \log(y)$ has a limit $\boldsymbol{\mu}(\boldsymbol{x}) \in \mathbb{R}^p$ as $y \to \infty$, and there exists $r > 0$ such that

$$\frac{1}{\|\boldsymbol{x}' - \boldsymbol{x}\|^2} \left| \frac{1}{\log(y)} \log \frac{\overline{F}(y|\boldsymbol{x}')}{\overline{F}(y|\boldsymbol{x})} - (\boldsymbol{x}' - \boldsymbol{x})^\top \frac{\nabla_{\boldsymbol{x}} \log \overline{F}(y|\boldsymbol{x})}{\log(y)} \right| < \infty$$

is uniformly bounded for $\boldsymbol{x}' \in B(\boldsymbol{x}, r)$ and y large.

Conditions $\mathcal{KS}$, $\mathcal{D}_g$ and $\mathcal{D}_\omega$ amount to (being able to use the) twice differentiability of the distribution of $\boldsymbol{X}$ and of the extreme value behavior of Y given $\boldsymbol{X}$.

We finally make two anti-clustering assumptions, one on $\boldsymbol{X}$ and the other on Y given $\boldsymbol{X}$.

Condition $\mathcal{B}_p$ - Anti-clustering in small balls

There exists an integer $t_0 \geq 1$ such that

$$1 \leq t < t_0 \Rightarrow \lim_{r \to 0} r^{-p} \, \mathbb{P}(\boldsymbol{X}_1 \in B(\boldsymbol{x}, r), \boldsymbol{X}_{t+1} \in B(\boldsymbol{x}, r)) = 0$$

$$\text{and } \limsup_{r \to 0} \sup_{t \geq t_0} r^{-2p} \, \mathbb{P}(\boldsymbol{X}_1 \in B(\boldsymbol{x}, r), \boldsymbol{X}_{t+1} \in B(\boldsymbol{x}, r)) < \infty.$$

Condition $\mathcal{B}_p$ means that values of $(\boldsymbol{X}_t)$ close in time might concentrate somewhat around $\boldsymbol{x}$ but those far apart in time cannot.

Under $\mathcal{D}_g$, $\mathcal{B}_p$ is automatically true if $(\boldsymbol{X}_t)$ is $\beta-$mixing, because then $(\boldsymbol{X}_1, \boldsymbol{X}_{t+1})$ has a p.d.f. g_t uniformly bounded in t around $(\boldsymbol{x}, \boldsymbol{x})$.

When $p \geq 2$, the causal and invertible AR(p) process does not satisfy this boundedness condition, but satisfies assumption $\mathcal{B}_p$ with $t_0 = p$.

Condition $\mathcal{B}_\Omega$ - Anti-clustering on conditional extremes

There exist $h, z > 0$ such that $\Omega_h(z|\boldsymbol{x}) < \infty$, where

$$\Omega_h(z|\boldsymbol{x}) = \sup_{t \geq 1} \sup_{\substack{\boldsymbol{x}', \boldsymbol{x}'' \in B(\boldsymbol{x}, h) \\ y, y' \geq z}} \frac{\mathbb{P}(Y_1 > y, Y_{t+1} > y' | \boldsymbol{X}_1 = \boldsymbol{x}', \boldsymbol{X}_{t+1} = \boldsymbol{x}'')}{\sqrt{\overline{F}(y|\boldsymbol{x}')\overline{F}(y'|\boldsymbol{x}'')}}.$$

Condition $\mathcal{B}_\Omega$ means that a joint conditional extreme value of (Y_1, Y_{t+1}) cannot be much more likely than a marginal conditional extreme of Y_1, uniformly across time and locally uniformly across the covariate space.

It is (interestingly) much weaker in spirit than analogue conditions in the unconditional setting.

Theorem (Intermediate quantiles - A.D., G.S. and A.U.-C. *AoS* 2023)

Assume that $\sum_{j=1}^{\infty} j^{\eta}\alpha(j) < \infty$ *for some* $\eta > 1$. *Let* $\tau_n \uparrow 1$ *and* $h_n \to 0$ *be such that* $nh_n^p(1-\tau_n) \to \infty$ *and* $r_n h_n^p \to 0$, *and suppose*

- $nh_n^p(1-\tau_n)\{h_n \log(1-\tau_n)\}^4 \to 0$,
- $\sqrt{nh_n^p(1-\tau_n)}A((1-\tau_n)^{-1}|\boldsymbol{x}) = \mathrm{O}(1)$,
- *There is* $\delta > 0$ *such that* $r_n(r_n/\sqrt{nh_n^p(1-\tau_n)})^{\delta} \to 0$.

Then

$$\sqrt{nh_n^p(1-\tau_n)}\left(\frac{\widehat{q}_n(\tau_n|\boldsymbol{x})}{q(\tau_n|\boldsymbol{x})} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})}\gamma^2(\boldsymbol{x})\right).$$

Certain conditions can be dropped if the data-generating process satisfies stronger mixing conditions (relevant for examples!)

A stronger, joint convergence result for several intermediate quantile estimators is available and shall prove useful for the estimation of $\gamma(\boldsymbol{x})$.

Remarks

The asymptotic distribution with dependent $(\boldsymbol{X}_t, Y_t)$ is the same as in the i.i.d. setting (see Daouia *et al.*, 2011). This is specific to nonparametric regression problems (see *e.g.* Linton and Xiao, 2013).

The optimal (over τ_n and h_n) rate of convergence is $n^{2\rho(\boldsymbol{x})/(2-(p+4)\rho(\boldsymbol{x}))}$, for intermediate conditional quantiles with $\tau_n = 1 - Cn^{-2/(2-(p+4)\rho(\boldsymbol{x}))}$.

When $\rho(\boldsymbol{x}) \to -\infty$, corresponding to the pure conditional Pareto model, the optimal convergence rate is $n^{-2/(p+4)}$ (for central quantiles).

This is the optimal convergence rate $n^{-2/(p+4)}$ for central conditional quantiles under twice differentiability. (Improvement w.r.t. literature!)

Estimation procedure - Extreme levels

At properly extreme levels, $\widehat{q}_n(\tau_n|\boldsymbol{x})$ is not consistent. But

$$q(\tau'|\boldsymbol{x}) \approx \left(\frac{1-\tau'}{1-\tau}\right)^{-\gamma(\boldsymbol{x})} q(\tau|\boldsymbol{x}) \text{ when } \tau, \tau' \uparrow 1.$$

Plugging in a consistent estimator $\widehat{\gamma}(\boldsymbol{x})$ of $\gamma(\boldsymbol{x})$ yields a conditional Weissman-type estimator of $q(\tau_n'|\boldsymbol{x})$ when τ_n' is extreme:

$$\widehat{q}^{W}_{n,\tau_n}(\tau_n'|\boldsymbol{x}) = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}(\boldsymbol{x})} \widehat{q}_n(\tau_n|\boldsymbol{x}), \text{ where } \frac{1-\tau_n'}{1-\tau_n} \to 0.$$

We use a (somewhat) Hill-type estimator from Daouia *et al.* (2011):

$$\widehat{\gamma}^{(J)}_{\tau_n}(\boldsymbol{x}) = \frac{1}{\log(J!)} \sum_{j=1}^{J} \log\left(\frac{\widehat{q}_n(1-(1-\tau_n)/j|\boldsymbol{x})}{\widehat{q}_n(\tau_n|\boldsymbol{x})}\right), \text{ for a fixed } J \geq 2.$$

Theorem (Extreme quantiles - A.D., G.S. and A.U.-C. *AoS* 2023)

Assume also that $\rho(\boldsymbol{x}) < 0$ *and* $\sqrt{nh_n^p(1-\tau_n)}A\left((1-\tau_n)^{-1}|\boldsymbol{x}\right) \to 0$*. Let* $\tau_n' \uparrow 1$ *be such that* $\sqrt{nh_n^p(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$*. Then*

$$\frac{\sqrt{nh_n^p(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{q}^{W}_{n,\tau_n}(\tau_n'|\boldsymbol{x})}{q(\tau_n'|\boldsymbol{x})} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\int_{\mathbb{R}^p} K^2}{g(\boldsymbol{x})}\gamma^2(\boldsymbol{x})\frac{J(J-1)(2J-1)}{6\log^2(J!)}\right).$$

This relies on the asymptotic normality of $\widehat{\gamma}^{(J)}_{\tau_n}(\boldsymbol{x})$, which uses the joint asymptotic normality of J intermediate conditional smoothed quantiles.

Some models where our assumptions are satisfied

Our conditions hold in:

- Location-scale models $Y_t = m(\boldsymbol{X}_t) + \sigma(\boldsymbol{X}_t)\varepsilon_t$ where m and $\sigma > 0$, and (ε_t) is a stationary and centered sequence of unobserved heavy-tailed innovations independent from the sequence $(\boldsymbol{X}_t)$.
- Nonlinear regression models $Y_t = q(U_t, \boldsymbol{\theta}(\boldsymbol{X}_t))$ where (U_t) is a stationary sequence of unobserved, uniformly distributed innovations independent from the sequence $(\boldsymbol{X}_t)$, and $q(\cdot, \boldsymbol{\theta})$ is a smooth parametric family of heavy-tailed quantile functions.
- Autoregressive models with heavy-tailed innovations, with the covariate made of lags of the response.

Regularity conditions (*e.g.* on m, σ, $\boldsymbol{\theta}$...) are of course necessary!

Back to the tornado loss data analysis

It was checked using local Generalized Pareto QQ-plots that the response Y is heavy-tailed conditional on $\boldsymbol{X}$.

We represent (a bias-corrected version of) $\widehat{q}^{W}_{n,\tau_n}(0.995|\boldsymbol{x})$ at each location $\boldsymbol{x}$, along with the Nadaraya-Watson estimator.

The extreme quantile estimate is calculated using data-driven selection rules of τ_n and h_n geared towards MSE minimization.

We also provide pointwise asymptotic Gaussian 95% confidence intervals, with a couple of corrections to try to improve coverage.

Two questions:

- Are the regression and extreme value analyses different?
- Is kernel smoothing able to smooth over the state-by-state analysis?

Tornado loss data - Regression analysis

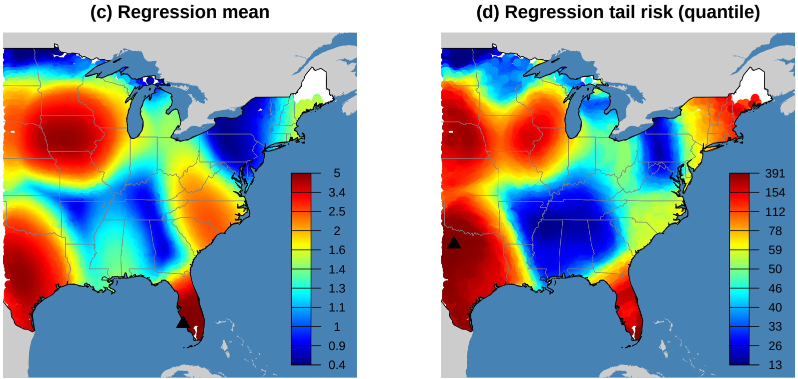


Figure: Tornado loss data. Left: Estimated conditional mean of losses per squared yard, right: Extrapolated conditional quantile estimate of those losses at level $\tau_n' = 0.995$. Cities with the highest estimated conditional average loss and extreme loss are marked with a black triangle in the left and right panels.

Location $\boldsymbol{x}$	$N_{h_n,\star}(\boldsymbol{x})$	$\widehat{\gamma}^{(J,BR)}_{1-k_{n,\star}/n}(\boldsymbol{x})$	$\widehat{q}_n(0.995 \mid \boldsymbol{x})$	$\widehat{q}^{W,BR}_{n,1-k_{n,\star}/n}(0.995 \mid \boldsymbol{x})$
New York (NY)	413	1.11	16.71 [2.46, 113.35]	63.53 [18.13, 222.70]
Charleston (SC)	839	0.96	48.08 [12.54, 184.28]	56.15 [22.58, 139.64]
Nashville (TN)	2,317	0.95	16.57 [8.31, 33.04]	24.51 [14.62, 41.08]
Captiva (FL)	205	0.93	236.74 [36.08, 1553.36]	144.67 [43.50, 481.13]
New Orleans (LA)	1,427	0.98	27.46 [11.44, 65.88]	31.01 [16.42, 58.58]
Woodson (TX)	958	1.53	118.37 [16.53, 847.72]	390.59 [100.95, 1511.26]
Kansas City (MO)	1,326	1.30	45.99 [11.29, 187.43]	69.00 [25.57, 186.21]
Minneapolis (MN)	620	0.93	34.09 [9.54, 121.79]	40.49 [17.01, 96.40]
Harrisville (MI)	472	0.79	24.86 [4.76, 129.71]	29.32 [10.28, 83.67]

Table: Tornado loss data. Results at selected cities, with the number of neighboring observations, along with 95% asymptotic confidence intervals. Captiva, FL is the city with maximal estimated average loss.

Conclusion

We revisit nonparametric extremal regression estimation via kernel smoothing for the conditional distribution function.

We find optimal rates akin to those of central nonparametric regression, under appropriate twice differentiability conditions w.r.t. $\boldsymbol{x}$.

Our assumptions hold for a variety of regression models and time series.

I did not talk about: Particular challenges of expectile regression (bias/variance correction), tuning parameter selection, simulation results for $\boldsymbol{X} \in \mathbb{R}$ or $\mathbb{R}^2$, a stock market data analysis...

For more, see Daouia, S. & Usseglio-Carleve (2023). Inference for extremal regression with dependent heavy-tailed data, *Annals of Statistics* **51**(5): 2040-2066.

Thank you for your attention!