COVARIATE EFFECTS IN EXTREMES – REMARKS AND THEORY[1]

Jan Dienstbier

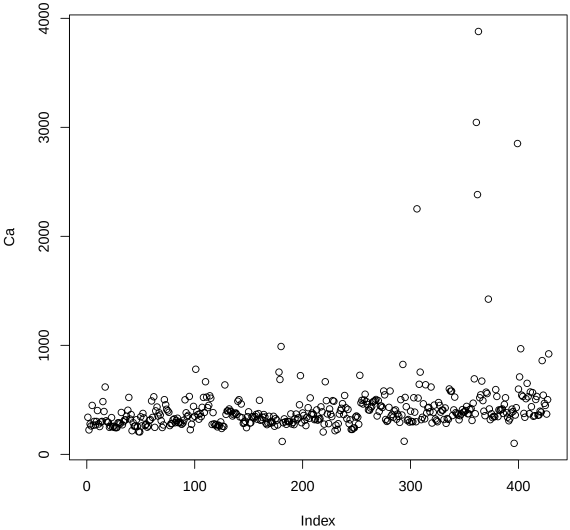
e-mail: dienstbier.jan@gmail.com

Technical University in Liberec

Němčičky, 9.9.2012

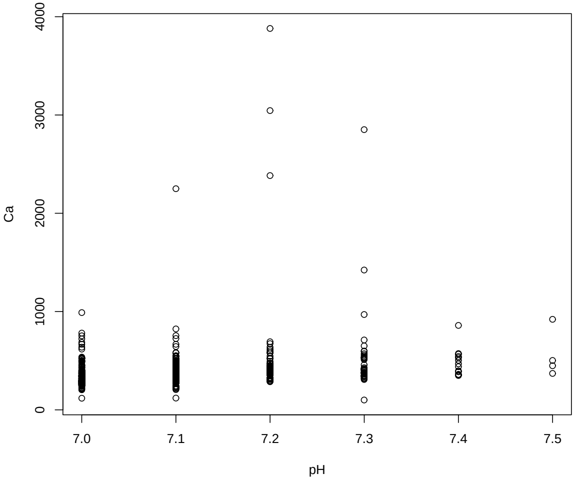
[1]The author and the research team KLIMATEXT benefited from project CZ.1.07/2.3.00/20.0086 co-financed by the European Social Fund and the state budget of Czech Republic.

Motivation – sample data



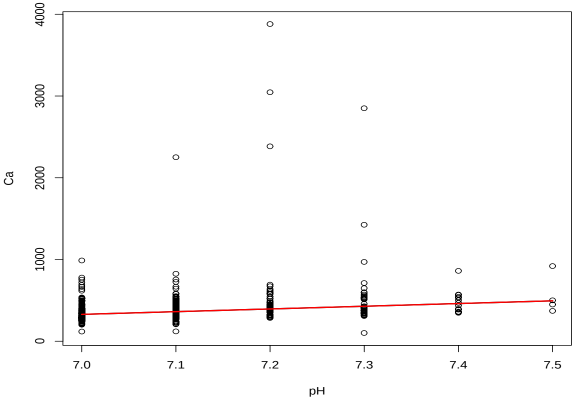
Condroz data

Dataset of Calcium content vs. pH in soil in Condroz region in Belgium.



Condroz data

Dataset of Calcium content vs. pH in soil in Condroz region in Belgium.



See Goegebeur et al. (2005), Vandewalle, Beirlant, Hubert (2006), Beirlant et al. (2004).

Heavy-tailed data – univariate case

Have E_i i.i.d. random variables. We say

$$E_i \in \mathcal{D}(G_\gamma), G_\gamma = \exp\left(-\left(1+\gamma x\right)^{-1/\gamma}\right)$$

i.e. there exists $a_n > 0$ and $b_n \in \mathbb{R}$

$$P\left(X_{n:n} \leq a_n x + b_n\right) \to G_\gamma(x).$$

for all $x \in \mathbb{R}$.

We are chiefly interested in the heavy-tailed errors ($\gamma > 0$), i.e. $F^{-1}(1-x)$ is regularly varying at zero (RV^0_γ).

$$\lim_{t \searrow 0} \frac{F^{-1}(1-tx)}{F^{-1}(1-t)} = x^{-\gamma}$$

…and as usual (to get more precise asymptotic), suppose we have a constant signed $A(t)$ and the second order approximation with some $\rho \in \mathbb{R}^+$

$$\lim_{t \searrow 0} \frac{\frac{F^{-1}(1-tx)}{F^{-1}(1-t)} - x^{-\gamma}}{A(t)} = x^{-\gamma} \cdot \frac{1-x^\rho}{\rho} =: K_{\gamma,\rho}(x).$$

Heavy-tailed data – univariate case

Have E_i i.i.d. random variables. We say

$$E_i \in \mathcal{D}(G_\gamma), G_\gamma = \exp\left(-(1+\gamma x)^{-1/\gamma}\right)$$

i.e. there exists $a_n > 0$ and $b_n \in \mathbb{R}$

$$P\left(X_{n:n} \leq a_n x + b_n\right) \to G_\gamma(x).$$

for all $x \in \mathbb{R}$.

We are chiefly interested in the heavy-tailed errors ($\gamma > 0$), i.e. $F^{-1}(1-x)$ is regularly varying at zero (RV_γ^0).

$$\lim_{t \searrow 0} \frac{F^{-1}(1-tx)}{F^{-1}(1-t)} = x^{-\gamma}$$

…and as usual (to get more precise asymptotic), suppose we have a constant signed $A(t)$ and the second order approximation with some $\rho \in \mathbb{R}^+$

$$\lim_{t \searrow 0} \frac{\frac{F^{-1}(1-tx)}{F^{-1}(1-t)} - x^{-\gamma}}{A(t)} = x^{-\gamma} \cdot \frac{1-x^\rho}{\rho} =: K_{\gamma,\rho}(x).$$

Heavy-tailed data – Drees (1998)

Suppose that E_i, $i = 1, \ldots, n$ are i.i.d. random variables fulfilling the second order condition for some $\gamma, \rho > 0$ and $k = k(n)$ is an intermediate sequence. Then we can define a sequence of Wiener processes $W_n(t), t \in [0,1]$ such that for $\varepsilon > 0$ sufficiently small

$$\sup_{0 \le t \le 1} t^{\gamma+1/2+\varepsilon} \left| k^{1/2} \left(\frac{E_{n-[kt],n}}{F^{-1}\left(1 - \frac{k}{n}\right)} - t^{-\gamma} \right) - \gamma t^{-\gamma-1} W_n(t) \right.$$
$$\left. - k^{1/2} A\left(\frac{k}{n}\right) t^{-\gamma} \frac{1 - t^{\rho}}{\rho} \right| \xrightarrow[n \to \infty]{\mathrm{P}} 0.$$

- wide range of applications – consider functional $T(E_{n-[kt],n})$ with T being location and scale invariant smooth functional
- a more complicated version for $\gamma \in \mathbb{R}$ exists
- Q: can something similar be established for linear models?

Simple linear model with heavy tails

$$\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times d}\boldsymbol{\beta}_{d\times 1} + \mathbf{E}_{n\times 1}$$

- $\mathbf{X}_{n\times d}$ known covariate matrix
- $\mathbf{E}_{n\times 1}$ i.i.d. errors
- $E_i \in \mathcal{D}(G_\gamma)$, $G_\gamma = \exp\left(-\left(1+\gamma x\right)^{-1/\gamma}\right)$

  i.e. there exists $a_n > 0$ and $b_n \in \mathbb{R}$

  $$P\left(X_{n:n} \leq a_n x + b_n\right) \to G_\gamma(x).$$

  for all $x \in \mathbb{R}$.
- $\gamma > 0$.

Regression quantiles

regression quantiles for $\alpha \in (0,1)$ and loss $\rho_\alpha(u) = u(\alpha - I(u<0))$ are defined

$$\widehat{\boldsymbol{\beta}}_n(\alpha) = \widehat{\boldsymbol{\beta}}_n\left(\alpha \middle| \mathbf{Y}, \mathbf{X}\right) := \arg\min_{b \in \mathbb{R}^d} \sum_{i=1}^{n} \rho_\alpha\left(Y_i - \mathbf{x}_i b\right).$$

Extreme regression quantiles

- the largest regression quantile

$$\widehat{\boldsymbol{\beta}}_n(1) = \widehat{\boldsymbol{\beta}}_n\left(1 \middle| \mathbf{Y}, \mathbf{X}\right) := \arg\min_{b \in \mathbb{R}^d} \sum_{i=1}^{n} \left(Y_i - \mathbf{x}_i b\right)^+,$$

cf. Smith (1994), Portnoy and Jurečková (1999), Jurečková (2000), Knight (2002).

- $\alpha_n^* \to 1$ with a given order
  - extreme order regression quantiles $(1-\alpha)n \to k > 0$, $n \to \infty$,
  - intermediate order regression quantiles $(1-\alpha)n \to \infty$, $\alpha \to 0$,

cf. Chernozhukov (2005).

Example: asymptotic for intermediate regression quantiles by Chernozhukov (2005)

$$\frac{\sqrt{\alpha n}}{\mu_{\mathbf{X}}^{\top}(\beta(\alpha) - \beta(m\alpha))}\left(\hat{\beta}(\alpha) - \beta(\alpha)\right) \xrightarrow[n \to \infty]{\mathcal{D}} \mathcal{N}(0, \Omega(\gamma))$$

where $\mu_{\mathbf{X}} = E\mathbf{x}$, $\boldsymbol{\beta}(\alpha) = (\beta_1 + F^{-1}(\alpha), \beta_2, \ldots, \beta_d)$, $m < 1$.

Tail quantile function vs. Regression quantile process

- the tail quantile function

$$Q_{n,k}(t) := F_n^{-1}\left(1 - \frac{kt}{n}\right) = E_{n-[k_n t]:n}, \quad t \in [0,1].$$

- the sample quantile process

$$q_n(\alpha) = n^{1/2}(F_n^{-1}(\alpha) - F^{-1}(\alpha)), \quad 0 < \alpha \leq 1.$$

- the tails of regression quantiles

$$\hat{\mathbf{Q}}_{n,k}(t) := \widehat{\boldsymbol{\beta}}_n\left(1 - \frac{tk}{n}\right), \qquad t \in [0,1],$$

- the process of regression quantiles

$$\hat{\mathbf{q}}_n(\alpha) := n^{\frac{1}{2}} f(F^{-1}(\alpha))\left(\widehat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right), \qquad 0 < \alpha < 1,$$

where $\boldsymbol{\beta}(\alpha) := (\beta_1 + F^{-1}(\alpha), \beta_2, \dots, \beta_d)$.

Tail quantile function vs. Regression quantile process

- the tail quantile function

$$Q_{n,k}(t) := F_n^{-1}\left(1 - \frac{kt}{n}\right) = E_{n-[k_n t]:n}, \quad t \in [0,1].$$

- the sample quantile process

$$q_n(\alpha) = n^{1/2}(F_n^{-1}(\alpha) - F^{-1}(\alpha)), \quad 0 < \alpha \leq 1.$$

- the tails of reparametrized regression quantiles

$$\hat{Q}_{n,k}(t) := \overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n\left(1 - \frac{tk}{n}\right), \qquad t \in [0,1],$$

- the process of reparametrized regression quantiles

$$\hat{q}_n(\alpha) := n^{\frac{1}{2}} f(F^{-1}(\alpha))\overline{\mathbf{x}}^\top\left(\widehat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right), \qquad 0 < \alpha < 1,$$

where $\boldsymbol{\beta}(\alpha) = (\beta_1 + F^{-1}(\alpha), \beta_2, \ldots, \beta_d)$ and $\overline{\mathbf{x}} := \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$.

Main results – an outline

1. Approximate $\mathbf{q}_n(\alpha)$, on $[\alpha_n^*, 1-\alpha_n^*]$, $\alpha_n^* \to 0$.
2. Approximate $\hat{q}_n(\alpha)$ on $[1-\alpha_n^*, 1-1/n]$.
3. Approximate $\hat{Q}_{n,k}(t)$ in the same way as $Q_{n,k}(t)$, cf. Drees (1998).
4. Describe estimators of γ as functionals of $Q_{n,k}(t)$.
5. The functionals have same properties on $\hat{Q}_{n,k}(t)$.

1. Approximation of regression quantile process

Under suitable conditions it holds

$$\sup_{\alpha_n^* \le \alpha \le 1-\alpha_n^*} \left| \sigma_\alpha^{-1} (\widehat{\boldsymbol{\beta}}_n\,(\alpha |\, Y, \mathbf{x}) - \boldsymbol{\beta}(\alpha)) \right| = O_P(n^{-1/2} (\log \log n)^{\frac{1}{2}}),$$

and

$$n^{1/2} \sigma_\alpha^{-1} \left(\widehat{\boldsymbol{\beta}}_n\,(\alpha |\, \mathbf{Y}, \mathbf{X}) - \boldsymbol{\beta}(\alpha) \right) =$$

$$n^{-1/2} (\alpha(1-\alpha))^{-1/2} \mathbf{D}_n^{-1} \sum_{i=1}^{n} \mathbf{x}_i \left(\alpha - I[E_i - F^{-1}(\alpha) < 0] \right) + o_P(1)$$

where $\sigma_\alpha := (\alpha(1-\alpha))^{1/2} / f(F^{-1}(\alpha))$ and $\alpha_n^* = (\frac{1}{n} \log^{2+\delta} n)$ for any $\delta > 0$

cf. Gutenbrunner et al. (1993) and Jurečková (1999), where $\alpha_n^* = n^{-1+\varepsilon}$ is used.

Assumptions

Distribution function

(F.1) F is absolutely continuous with the positive density on (x_*, x^*). There exists f', the derivative of density f.

(F.2) There exists some $0 < K_\gamma < \infty$ such that

$$\sup_{x_* < x < x^*} F(x)(1 - F(x)) \left| \frac{f'(x)}{f^2(x)} \right| \le K_\gamma .$$

(F.3)

$$\limsup_{x \uparrow x^*} \frac{(1 - F(x)) f'(x)}{f^2(x)} = -1 - \gamma^* .$$

for some $\gamma^* > -1/2$ (lower tail index γ_* similarly).

Covariance matrix

(X.1) $x_{i1} = 1, \quad i = 1, \ldots, n$.

(X.2) $\lim_{n\to\infty} \mathbf{D}_n = \mathbf{D}$, where $\mathbf{D}_n = n^{-1} \mathbf{X}_n^\top \mathbf{X}_n$ and $\mathbf{D}$ is a positive definite $(d \times d)$ matrix.

(X.3) $n^{-1} \sum_{i=1}^n |\mathbf{x}_{ni}|^4 = O(1)$ as $n \to \infty$.

(X.4) $\max_{1\le i \le n} |\mathbf{x}_{ni}| = O((\log\log n)^{1/2})$ as $n \to \infty$.

Proof

- Prove that $\sup\left\{|r_n(\mathbf{t},\alpha)| : \alpha_n^* \le \alpha \le 1-\alpha_n^*, \|\mathbf{t}\| \le (\log\log n)^{1/2}\right\} = o_P(1)$,

$$r_n(\mathbf{t},\alpha) := (\alpha(1-\alpha))^{-1/2}\sigma_\alpha^{-1}\sum_{i=1}^n\left[\rho_\alpha\left(E_{i\alpha}-n^{-1/2}\sigma_\alpha\mathbf{x}_i^\top\mathbf{t}\right)-\rho_\alpha(E_{i\alpha})\right]$$
$$+n^{-1/2}(\alpha(1-\alpha))^{-1/2}\mathbf{t}^\top\sum_{i=1}^n\mathbf{x}_i\psi_\alpha(E_{i\alpha})-\frac{1}{2}\mathbf{t}^\top\mathbf{D}_n\mathbf{t}$$

  and $E_{i\alpha} := E_i - F^{-1}(\alpha)$, $i=1,\ldots,n$, $\quad 0<\alpha<1$, $\psi_\alpha(u) := \alpha - I(u<0)$.

  - approximate the mean of $r_n(\mathbf{t},\alpha)$ for any suitable α and $\mathbf{t}$.
  - Bernstein inequality gives a probabilistic bound for any α and $\mathbf{t}$.
  - Chaining arguments give the uniform bound.

- $n^{1/2}\sigma_\alpha^{-1}(\hat{\boldsymbol{\beta}}_n(\alpha)-\boldsymbol{\beta}(\alpha))$ minimizes the convex function

$$G_{n\alpha}(\mathbf{t}) = (\alpha(1-\alpha))^{-1/2}\sigma_\alpha^{-1}\sum_{i=1}^n\left[\rho_\alpha(E_{i\alpha}-n^{-1/2}\sigma_\alpha\mathbf{x}_i^\top\mathbf{t})-\rho_\alpha(E_{i\alpha})\right]$$

- use the properties of $r_n(\alpha,\mathbf{t})$ to calculate the solution for $\|\mathbf{t}\| \le (\log\log n)^{1/2}$.
- convexity of $G_{n\alpha}(\mathbf{t})$ implies that the minimum cannot be attained elsewhere.

2. Regression quantile process at the tails

Suppose that $\gamma^* > 0$. Then

$$\sup_{1-\alpha_n^* \le \alpha \le \frac{n-1}{n}} \left|\overline{\mathbf{x}}^\top \mathbf{q}_n(\alpha)\right| = \sup_{1-\alpha_n^* \le \alpha \le \frac{n-1}{n}} \left| n^{1/2} f\left(F^{-1}(\alpha)\right) \overline{\mathbf{x}}^\top \left(\widehat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right)\right|$$
$$= \mathcal{O}_P(n^{-1/2}(\log n)^{(2+\delta)(1\vee\gamma^*)}) = o_P(1).$$

and if $\gamma_* > 0$ is the tail index of the lower tail it holds also

$$\sup_{1/n \le \alpha \le \alpha_n^*} \left|\overline{\mathbf{x}}^\top \mathbf{q}_n(\alpha)\right| = \sup_{1/n \le \alpha \le \alpha_n^*} \left| n^{1/2} f\left(F^{-1}(\alpha)\right) \overline{\mathbf{x}}^\top \left(\widehat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right)\right|$$
$$= \mathcal{O}_P(n^{-1/2}(\log n)^{(2+\delta)(1\vee\gamma_*)}) = o_P(1),$$

Proof

- $\overline{\mathbf{x}}^\top \boldsymbol{\beta}_n(\alpha_1) \leq \overline{\mathbf{x}}^\top \boldsymbol{\beta}_n(\alpha_2)$ iff $\alpha_1 \leq \alpha_2$,
- similarly as in Portnoy and Jurečková (1999) get

$$P_{\boldsymbol{\beta}}\left(\sum_{i=1}^{n} \mathbf{x}_i^\top \left(\boldsymbol{\beta}(1) - \boldsymbol{\beta}\right) \geq nt\right) \quad \leq \quad P\left(E_{n:n} \geq t\right),$$

- assuming $\gamma = \gamma^* > 0$ it follows

$$P\left(\frac{E_{n:n}}{F^{-1}\left(1 - 1/n\right)} \geq \zeta\right) \xrightarrow[n\to\infty]{\mathcal{D}} 1 - \exp\left(-\zeta^{-\frac{1}{\gamma}}\right),$$

- use von Mises condition and Lemma 4.5.2. of Csörgő and Révész (1977) for transition from $f(F^{-1}(1 - kt/n))$ to $f(F^{-1}(1 - k/n))$.

3. Tails of regression quantiles

Assume

- model with i.i.d. errors fulfilling the second order condition, $\gamma, \rho > 0$,
- $k = k(n) \to \infty$, $k/n \to 0$ and $\sqrt{k}A(k/n) = \lambda$,
- $k \geq \log^{\Delta(1\vee\gamma)}(n)$, $\Delta > 4 + 2\delta$.
- $\|z\|_{\gamma,\varepsilon} := \sup_{t\in[0,1]} \left|t^{1/2+\gamma+\varepsilon} z(t)\right|$, $z \in D[0,1]$.

There are Wiener processes $W_n(t)$, $\tilde{W}_n(t)$, and $\mathbf{W}(t)$ such that for any $\varepsilon > 0$.

$$\left\| k^{1/2} \left(\frac{\overline{\mathbf{x}}^\top \left(\widehat{\boldsymbol{\beta}}_n \left(1 - \frac{kt}{n}\right) - \boldsymbol{\beta} \right)}{F^{-1}\left(1 - \frac{k}{n}\right)} - t^{-\gamma} \right) - \gamma t^{-\gamma-1} W_n(t) \right.$$
$$\left. - k^{1/2} A\left(\frac{k}{n}\right) t^{-\gamma} \frac{1 - t^\rho}{\rho} \right\|_{\gamma,\varepsilon}$$
$$\leq \left\| \gamma t^{-\gamma} \overline{\mathbf{x}}^\top \mathbf{D}^{-1} \mathbf{W}(t) \right\|_{\gamma,\varepsilon} + \left\| \gamma t^{-\gamma} \tilde{W}_n(t) \right\|_{\gamma,\varepsilon} + o_P(1),$$

...which is an analogy to Drees(1998).

$$\sup_{0\le t\le 1} t^{\gamma+1/2+\varepsilon}\left|k^{1/2}\left(\frac{E_{n-[kt],n}}{F^{-1}\left(1-\frac{k}{n}\right)}-t^{-\gamma}\right)-\gamma t^{-\gamma-1}W_n(t)\right.$$
$$\left.-k^{1/2}A\left(\frac{k}{n}\right)t^{-\gamma}\frac{1-t^{\rho}}{\rho}\right|\xrightarrow[n\to\infty]{\mathrm{P}} 0.$$

Proof

- combination of the previous results on approximations of regression quantiles,
- von Mises condition and Lemma 4.5.2. of Csörgő and Révész (1981) used for transition from $f(F^{-1}(1-kt/n))$ to $f(F^{-1}(1-k/n))$,
- direct procedure – just a rough approximation.

4. Functionals of tail quantile functions

Have $\gamma \in \mathbb{R}$ and a functional $T : span(\mathcal{H}_\mathcal{M}, 1) \to \mathbb{R}$ satisfying

1. $\mathcal{H}_\mathcal{M}$ is semimetric space, where the tail quantile function and its relatives live
2. $T(az+b) = T(z), \qquad$ for all $z \in \mathcal{H}_\mathcal{M}$, $a > 0$, $b \in \mathbb{R}$,
3. $T(z_\gamma) = T(\frac{x^{-\gamma}-1}{\gamma}) = \gamma$
4. $T_{|\mathcal{H}_\mathcal{M}}$ is Hadamard differentiable tangentially to suitable continuous $\mathcal{C}_\mathcal{M} \subset \mathcal{H}_\mathcal{M}$, at z_γ with a derivative T'_γ, i.e. for some signed measure $\nu_{T,\gamma}$ it holds for all $0 < \varepsilon_n \to 0$ and all $y_n \in \mathcal{H}_\mathcal{M}$ such that $y_n \to y \in \mathcal{C}_\mathcal{M}$

$$\lim_{\varepsilon_n \to 0} \frac{T(z_\gamma - \varepsilon_n y_n) - T(z_\gamma)}{\varepsilon_n} = T'_\gamma(y) = \int_0^1 y \mathrm{d}\nu_{T,\gamma}.$$

Then $T(Q_{n,k}) \to \gamma$ and for intermediate sequence k_n with the rate parameter $\lambda = func.(\gamma, \rho, k_n)$ it holds

$$\mathcal{L}(k_n^{1/2}(T(Q_{n,k}) - \gamma)) \to \mathcal{N}(\lambda\nu_{T,\gamma,\rho}, \sigma_{T,\gamma}),$$

c.f. Drees (1998).

Variance and bias of the estimation

Provided that $\sqrt{k}A(k/n) \to \lambda$ it holds

(i) $T(Q_{n,k}) \to \gamma$

(ii) $\mathcal{L}(k_n^{1/2}(T(Q_{n,k}) - \gamma)) \to \mathcal{N}(\lambda\nu_{T,\gamma,\rho}, \sigma_{T,\gamma})$, where

$$
\begin{aligned}
\mu_{T,\gamma,\rho} &:= \int_0^1 t^{-\gamma}\frac{1-t^\rho}{\rho}\mathrm{d}\,\nu_{T,\gamma} \\
\sigma_{T,\gamma} &:= Var\left(\int_0^1 t^{-\gamma-1}W(t)\,\mathrm{d}\nu_{T,\gamma}(t)\right) \\
&= \int_0^1\int_0^1 (st)^{\gamma-1}\min(s,t)\mathrm{d}\,\nu_{T,\gamma}(s)\,\mathrm{d}\nu_{T,\gamma}(t)
\end{aligned}
$$

5. $T(\hat{Q}_{n,k}(t))$ is the estimator

Suppose that T fulfills the given assumptions. For $T(\hat{Q}_{n,k}(t))$ it follows:

1. consitency:
   - follows immediately from continuity of T and approximations given previously.
2. asymptotic normality:
   - requires Hadamard differentiability,
   - as we have an extra random remainder (with zero mean), asymptotic variance can be only roughly estimated,
   - asymptotic bias is the same one as in the i.i.d. case.

Available estimators

Examples:

- Pickands estimator

$$T_{\text{Pick}}(z) := \frac{1}{\log 2} \log\left(\frac{z(1/4) - z(1/2)}{z(1/2) - z(1)}\right) I\left[\frac{z(1/4) - z(1/2)}{z(1/2) - z(1)} > 0\right].$$

- Probability weighted moments estimator

$$T_{\text{PWM}}(z) := \frac{\int_0^1 (z(t) - z(1))(1 - 4t)dt}{\int_0^1 (z(t) - z(1))(1 - 2t)dt} I\left[\int_0^1 (z(t) - z(1))(1 - 2t)dt > 0\right].$$

- Maximum likelihood estimator – generated by an implicitly given functional, see Drees (1998).

Functionals on $\overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n(1-tk/n)_{t\in[0,1]}$

$T(\overline{\mathbf{x}}^\top \hat{\mathbf{Q}}_{n,k}) = T(\overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n(1-tk/n)_{t\in[0,1]})$ are consistent and asymptotically normal estimators of γ.

- ML-estimator of γ based on the k largest unique estimates of $\overline{\mathbf{x}}^\top \hat{\boldsymbol{\beta}}_n(\tau), \quad \tau \in (0,1)$, i.e. the estimator fits generalized Pareto distribution (GPD) on the exceedances of $\{\overline{\mathbf{x}}^\top \hat{\boldsymbol{\beta}}_n(\tau_j)\}_{j=m-k,\ldots,m}$ over $\overline{\mathbf{x}}^\top \hat{\boldsymbol{\beta}}_n(\tau_{m-k-1})$.
- Probability weighted moments estimator (PWM)

$$\hat{\gamma}_{m,k}^{\text{RQ,PWM}} = \frac{\frac{1}{k}\sum_{j=1}^{k}\left(4\frac{j}{k+1}-3\right)\overline{\mathbf{x}}^\top \hat{\boldsymbol{\beta}}_n(\tau_{m-i+1})}{\frac{1}{k}\sum_{j=1}^{k}\left(2\frac{j}{k+1}-1\right)\overline{\mathbf{x}}^\top \hat{\boldsymbol{\beta}}_n(\tau_{m-i+1})}$$

- Pickands estimator

$$\hat{\gamma}_{m,k}^{\text{RQ,P}} = \frac{1}{\log 2}\log\left(\frac{\overline{\mathbf{x}}^\top \hat{\boldsymbol{\beta}}_n(\tau_{m-[k/4]}) - \overline{\mathbf{x}}^\top \hat{\boldsymbol{\beta}}_n(\tau_{m-[k/2]})}{\overline{\mathbf{x}}^\top \hat{\boldsymbol{\beta}}_n(\tau_{m-[k/2]}) - \overline{\mathbf{x}}^\top \hat{\boldsymbol{\beta}}_n(\tau_{m-k})}\right)$$

Where $\tau_1,\ldots,\tau_m$ are such that $\widehat{\boldsymbol{\beta}}_n\tau_i, i=1,\cdots,m$ are m unique solution of minimazation problem $\arg\min_{b\in\mathbb{R}^d}\sum_{i=1}^{n}\rho_\alpha(Y_i-\mathbf{x}_i\mathbf{b})$ for $\alpha\in[0,1]$.

Reparametrization

Have

$$\begin{aligned}\tilde{x}_{i,1} &= 1, \\ \tilde{x}_{i,j} &= x_{i,j} - \frac{1}{n}\sum_{i=1}^{n} x_{i,j}, \qquad j = 2, \ldots, p.\end{aligned}$$

Hence, after reparametrization $\overline{\mathbf{x}} = (1, 0, \ldots, 0)$.

Functionals on $\hat{\beta}_{n,1}(1-tk/n)_{t\in[0,1]}$

$T(\hat{Q}_{n,k}) = T(\hat{\beta}_{n,1}(1-tk/n)_{t\in[0,1]})$ are consistent and asymptotically normal estimators of γ.

- ML-estimator of γ based on the k largest unique estimates of $\hat{\beta}_{n,1}(\tau), \quad \tau \in (0,1)$, i.e. the estimator fits generalized Pareto distribution (GPD) on the exceedances of $\{\hat{\beta}_{n,1}(\tau_j)\}_{j=m-k,\dots,m}$ over $\hat{\beta}_{n,1}(\tau_{m-k-1})$.
- Probability weighted moments estimator (PWM)

$$\hat{\gamma}_{m,k}^{\text{RQ,PWM}} = \frac{\frac{1}{k}\sum_{j=1}^{k}\left(4\frac{j}{k+1}-3\right)\hat{\beta}_{n,1}(\tau_{m-i+1})}{\frac{1}{k}\sum_{j=1}^{k}\left(2\frac{j}{k+1}-1\right)\hat{\beta}_{n,1}(\tau_{m-i+1})}$$

- Pickands estimator

$$\hat{\gamma}_{m,k}^{\text{RQ,P}} = \frac{1}{\log 2}\log\left(\frac{\hat{\beta}_{n,1}(\tau_{m-[k/4]}) - \hat{\beta}_{n,1}(\tau_{m-[k/2]})}{\hat{\beta}_{n,1}(\tau_{m-[k/2]}) - \hat{\beta}_{n,1}(\tau_{m-k})}\right)$$

Where $\tau_1,\dots,\tau_m$ are such that $\hat{\beta}_{n,1}(\tau_i), i = 1,\cdots,m$ are m unique intercepts of regression quantiles in reparametrized model.

Summary of previous

Achievements:

- improvements of older approximations of regression quantiles
  - wider interval $[\alpha_n^*, 1-\alpha_n^*]$
  - at least a rough approximation for $[1-\alpha_n^*, 1-1/n]$
- general approximation methodology of γ based on regression quantiles

Open questions:

- further improvements of approximations
  - use Hungarian construction instead of Bahadur representation
  - improve approximation of regression quantile process in $[1-\alpha_n^*, 1-1/n]$
- dependency of errors

Two-step regression quantiles

Step 1: Calculate R-estimate of the slope i.e. invert the rank statistics in Hodges-Lehmann manner. Have

- $R_{ni}(\mathbf{Y} - \mathbf{X}\mathbf{b})$ be the rank of $Y_i - \mathbf{x}_i^\top \mathbf{b}$ among $(Y_1 - \mathbf{x}_1^\top \mathbf{b}, \ldots, Y_n - \mathbf{x}_n^\top \mathbf{b}), \mathbf{b} \in \mathbb{R}^p$
- $\varphi_\alpha = \alpha - I[x < 0], x \in \mathbb{R}$
- $\mathbf{x}_i$ be the i-th row of the $\mathbf{X}_{n \times d}$

Minimize the Jaeckel's measure of rank dispersion.

$$\widehat{\boldsymbol{\beta}}_{nR} = \operatorname{argmin}_{\mathbf{b} \in \mathbb{R}^p} \sum_{i=1}^{n} (Y_i - \mathbf{x}_i' \mathbf{b}) \varphi_\alpha \left(\frac{R_{ni}(\mathbf{Y} - \mathbf{X}\mathbf{b})}{n+1} \right)$$

Step 2: Get the ordered residuals $\tilde{\beta}_{n0} = Y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{nR}(\alpha)$

$\Rightarrow$ two-step regression quantiles $\left(\tilde{\beta}_{n0}, \hat{\boldsymbol{\beta}}_{nR}(\alpha) \right)$

Two step r.q. process and its tail approximation

$$\hat{E}_{k:n} := \left(\{Y_1 - \mathbf{x}_1^\top \widehat{\boldsymbol{\beta}}_{nR}, \ldots, Y_n - \mathbf{x}_n^\top \widehat{\boldsymbol{\beta}}_{nR}\}\right)_{k:n}$$

and

$$\tilde{Q}_{n,k}(t) := F_n^{-1}\left(1 - \frac{k_n}{n} t\right) = \hat{E}_{n-[k_n t]:n}, \quad t \in [0,1],$$

Have again model with $E_i \sim F$, with F satisfying the second order condition for some $\gamma \in \mathbb{R}$ and $\rho \leq 0$. Then under suitable conditions on F and $\mathbf{X}$ we can define a sequence of Wiener processes $\{W_n(t)\}_{t \geq 0}$ such that for suitable chosen functions A and a and each $\varepsilon > 0$,

$$\sup_{t \in (0,1]} t^{\gamma + \frac{1}{2} + \varepsilon} \left| \frac{\tilde{Q}_{n,k}(t) - F^{-1}\left(1 - \frac{k}{n}\right) - \beta_0}{a(k/n)} - \left(z_\gamma(t) - k^{-\frac{1}{2}} t^{-(\gamma+1)} W_n(t) \right.\right.$$
$$\left.\left. + A\left(\frac{k}{n}\right) H(t)\right) \right| = o_P\left(k^{-1/2} + |A(k/n)|\right)$$

$n \to \infty$, provided $k = k(n) \to \infty$, $k/n \to 0$ and $\sqrt{k} A(k/n) = O(1)$ and $z_\gamma(t) = \frac{t^{-\gamma} - 1}{\gamma}$, cf. Picek and Dienstbier (2010).

Remark

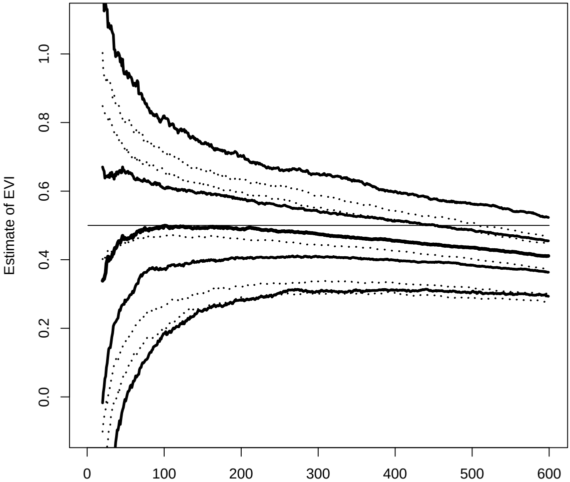
- under suitable condition, the method of proof can be used for any convergent estimate of β and its ordered residuals
- however, is not “suitable condition” = “neglecting real data structures”?

Remark

- under suitable condition, the method of proof can be used for any convergent estimate of β and its ordered residuals
- however, is not “suitable condition” = “neglecting real data structures”?

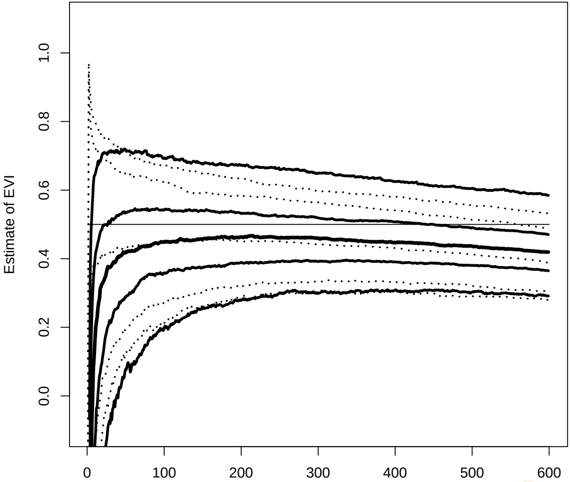
Simulations

Example: $Y_i = 1 + 5x_i + e_i$, where $x_i \sim U(0,1)$, e_i have Burr distribution with shape $\gamma = 0.5$



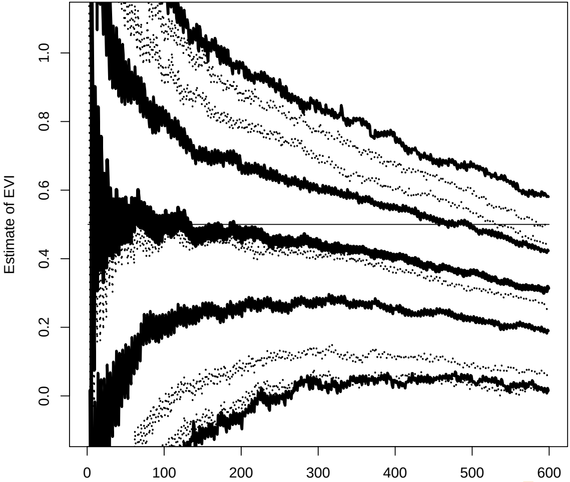
Simulations

Example: $Y_i = 1 + 5x_i + e_i$, where $x_i \sim U(0,1)$, e_i have Burr distribution with shape $\gamma = 0.5$



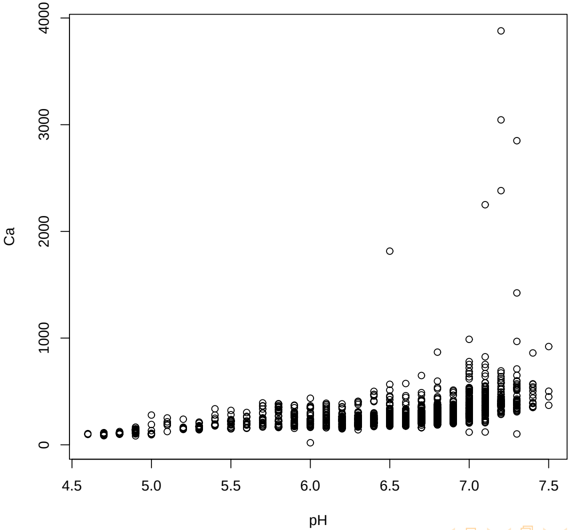
Simulations

Example: $Y_i = 1 + 5x_i + e_i$, where $x_i \sim U(0,1)$, e_i have Burr distribution with shape $\gamma = 0.5$



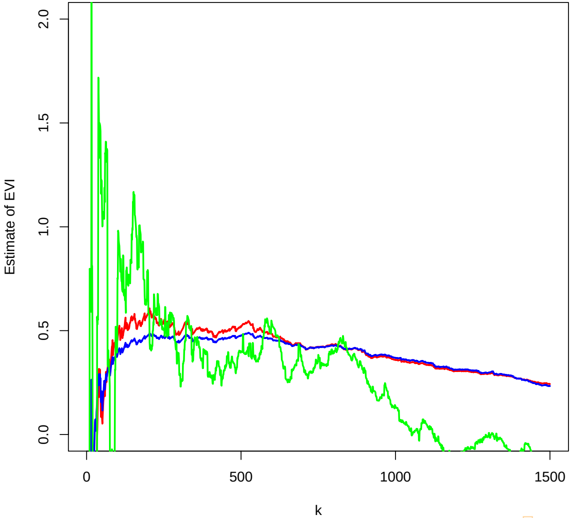
Condroz

Example: Condroz dataset again



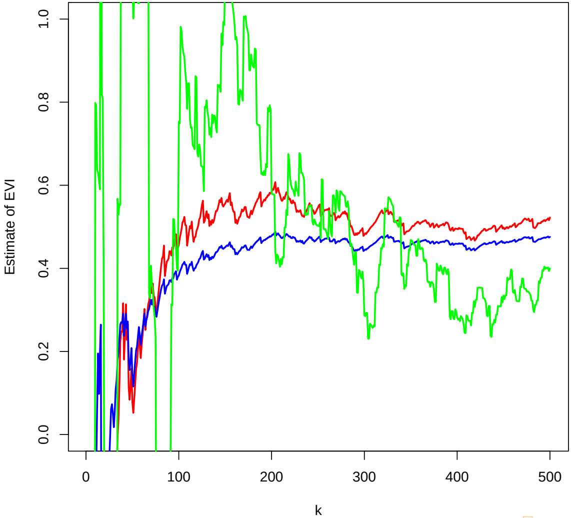
Condroz

Example: Condroz dataset again, estimator plots



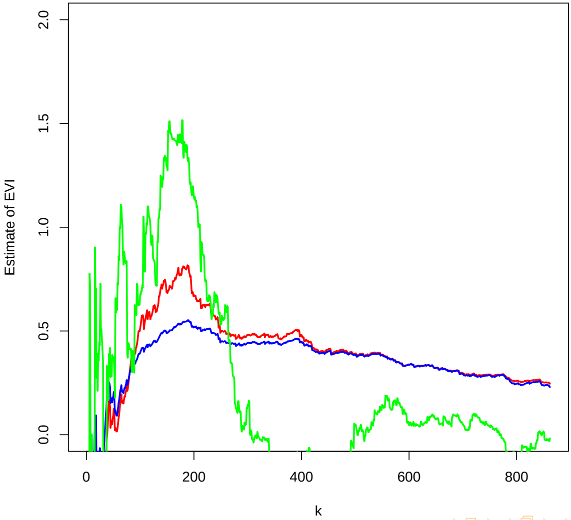
Condroz

Example: Condroz dataset again, estimator plots



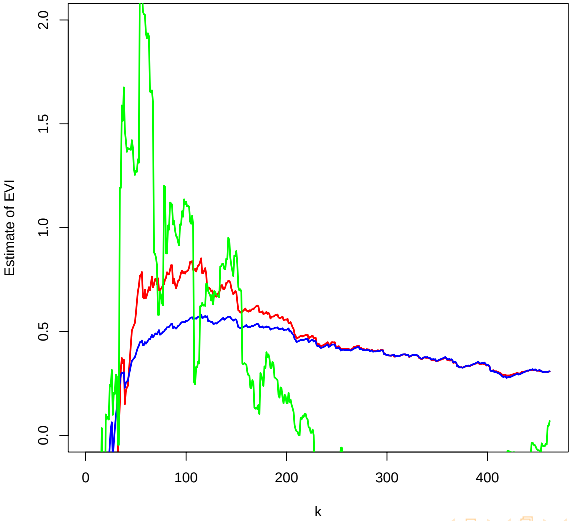
Condroz

Example: Condroz dataset again, estimator plots $6.6 \le pH \le 7.5$



Condroz

Example: Condroz dataset again, estimator plots $6.6 \leq pH \leq 7.3$



Remarks

- theory shows that if "thinks goes well" estimates based on quantile regression works
  - i.e. if the model is same (or simpler) as we suppose
- one additional interpretation of Condroz data *(hurray!)*

Something structural?

- extremes in predictors matters
- we often do not have the same number of observations for different predictors
  - problem, if nice model for all responses desired
  - get a nice model = root out enough data as outliers!
- linear models tend not to be linear
  - at least, if we want to work with all responses
- other possible problems. . .

Remarks

- theory shows that if “thinks goes well” estimates based on quantile regression works
  - i.e. if the model is same (or simpler) as we suppose
- one additional interpretation of Condroz data *(hurray!)*

Something structural?

- extremes in predictors matters
- we often do not have the same number of observations for different predictors
  - problem, if nice model for all responses desired
  - get a nice model = root out enough data as outliers!
- linear models tend not to be linear
  - at least, if we want to work with all responses
- other possible problems. . .

Further remarks & hypocrisy continued

- in EVT linear models, the choice of model matters, not the data
- which model is better than others depends strictly on the exact data settings and not the theory
- EVT can be a dangerous drug (do not abuse)

cf.

hypocrite *n. One who, professing virtues that he does not respect, secures the advantage of seeming to be what he despises.*

story *n. A narrative, commonly untrue.*

politeness *n. The most acceptable hypocrisy.*

– Ambrose Bierce, *Devil's Dictionary*

COVARIATE EFFECTS IN EXTREMES – REMARKS AND THEORY[1]

Jan Dienstbier

e-mail: dienstbier.jan@gmail.com

Technical University in Liberec

Němčičky, 9.9.2012

[1]The author and the research team KLIMATEXT benefited from project CZ.1.07/2.3.00/20.0086 co-financed by the European Social Fund and the state budget of Czech Republic.