Nonparametric estimation of conditional Value-at-Risk and Expected Shortfall based on Extreme Value Theory

Carlos Martins-Filho, Feng Yao and Maximo Torero

University of Colorado and IFPRI, West Virginia University and IFPRI, IFPRI

University of Bonn - ZEF
Paper is at:
http://spot.colorado.edu/∼martinsc/Research.html

Preliminaries

- Understanding and modeling price volatility
- Pushing the research frontier
- Identifying periods of excessive price volatility

In empirical finance there is often an interest in stochastic models for log returns

$$Y_t = log\frac{P_t}{P_{t-1}} \text{ where } t \in \{0, \pm 1, \cdots\}.$$

Motivation

- $\{Y_t\}_{t\in\mathbb{Z}}$ be a stochastic process
- $F_{Y_t|\mathbf{X}_t=\mathbf{x}}$, $\mathbf{X}_t \in \mathbb{R}^d$.
- Normally, $\mathbf{X}_t' = \begin{pmatrix} Y_{t-1} & \cdots & Y_{t-m} & W_t' \end{pmatrix}$ for $m \in \mathbb{N}$.

For $a \in (0,1)$,

- a-CVaR($\mathbf{x}$) is the a-quantile associated with $F_{Y_t|\mathbf{X}_t=\mathbf{x}}$,
- a-CES($\mathbf{x}$) is the $E(Y_t|Y_t > a\text{-CVaR}(\mathbf{x}))$.

These are frequently used as synthetic measures of risk by regulators, portfolio managers, etc.

How does Y_t evolve through time?

We consider the following conditional location-scale model

$$Y_t = m(\mathbf{X}_t) + h^{1/2}(\mathbf{X}_t)\varepsilon_t, \text{ where } t = 1, \cdots, n.$$

- $m, h : \mathbb{R}^d \to \mathbb{R}$ are suitably restricted real valued functions
- $E(\varepsilon_t|\mathbf{X}_t = \mathbf{x}) = 0$ and $V(\varepsilon_t|\mathbf{X}_t = \mathbf{x}) = 1$
- ε_t has a strictly increasing absolutely continuous distribution F which belongs to the domain of attraction of an extremal distribution [Leadbetter (1983), Resnick (1987)].

A result of Gnedenko (1943)

- Let $\{X_t\}_{t \geq 1}$ be a sequence of iid random variables with distribution F and let $m_n = max\{X_1, \cdots, X_n\}$. Then,

  $$P(m_n \leq x) = P(X_t \leq x, \forall t) = F(x)^n$$

  Suppose there exists $a_n > 0$, $b_n \in \mathbb{R}$ such that as $n \to \infty$

  $$P\left(\frac{m_n - b_n}{a_n} \leq x\right) = F(a_n x + b_n)^n \to E(x)$$

  then $E(x)$ is either

  1. $\Phi_\alpha(x) = e^{-x^{-\alpha}}$ for $x \geq 0$ (Fréchet)
  2. $\Psi_\alpha(x) = e^{-(-x)^\alpha}$ for $x < 0$ (reverse Weibull) and 1 for $x \geq 0$,
  3. $\Lambda(x) = e^{-e^{-x}}$ for $x \in \mathbb{R}$ (Gumbel).

- There are F's that are not in the domain of attraction of E but they constitute rather pathological cases (Leadbetter et al., 1983).

Some restrictions of the location scale model

1. AR(m), ARCH(p): $\mathbf{X}_t = \begin{pmatrix} 1 & Y_{t-1} & \cdots & Y_{t-m} \end{pmatrix}$

   $m(\mathbf{X}_t) = \mathbf{X}_t'\mathbf{b}$

   $h(\mathbf{X}_t) = \begin{pmatrix} 1 & Y_{t-1}^2 & \cdots & Y_{t-p}^2 \end{pmatrix} \mathbf{a}$
2. CHARN Model of Diebolt and Guègan (1993), Härdle and Tsybakov (1997), Hafner (1998).
   $\mathbf{X}_t = \begin{pmatrix} Y_{t-1} \end{pmatrix}$

   $m(\mathbf{X}_t) = m(Y_{t-1})$

   $h(\mathbf{X}_t) = h(Y_{t-1})$
3. Nonaparametric autoregression of Fan and Yao (1998, Biometrika)

CVaR and CES

For $a \in (0,1)$,

$$a - \text{CVaR}(\mathbf{x}) = q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a) = m(\mathbf{x}) + h^{1/2}(\mathbf{x})q(a)$$

and

$$a{-}\text{CES}(\mathbf{x}) = E(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)) = m(\mathbf{x}){+}h^{1/2}(\mathbf{x})E(\varepsilon_t|\varepsilon_t > q(a))$$

where $q(a)$ is the a-quantile associated with F.

The sequence $\{\varepsilon_t\}$ is not observed.

Motivation: McNeill and Frey (2000), Martins-Filho and Yao (2006).

Estimation

Given a sample $\{(Y_t, \mathbf{X}_t^T)\}_{t=1}^n$ and estimators $\hat{m}(\mathbf{x})$ and $\hat{h}(\mathbf{x})$ it is possible to obtain a sequence of standardized nonparametric residuals

$$\hat{\varepsilon}_t = \frac{Y_t - \hat{m}(\mathbf{X}_t)}{\hat{h}^{1/2}(\mathbf{X}_t)} \chi_{\{\hat{h}(\mathbf{X}_t)>0\}} \text{ for } t = 1, \cdots, n,$$

These can be used to construct

$$\hat{q}_{Y|X_i=x}(a) = \hat{m}(\mathbf{x}) + \hat{h}^{1/2}(\mathbf{x})\hat{q}(a)$$

$$\hat{E}(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)) = \hat{m}(\mathbf{x}) + \hat{h}^{1/2}(\mathbf{x})\hat{E}(\varepsilon_t|\varepsilon_t > q(a))$$

Pickands' result

- We are interested in the case where a is in the vicinity of 1.
- The restriction that a is in a neighborhood of 1 is useful in estimation. The result is due to Pickands (1975).

$F(x) \in D(E)$ if, and only if,

$$lim_{\xi \to u_\infty} sup_{0<u<u_\infty - \xi} |F_\xi(u) - G(u; 0, \sigma(\xi), k)| = 0,$$

where

- $F_\xi(u) = \frac{F(u+\xi) - F(\xi)}{1 - F(\xi)}$,
- G is a generalized Pareto distribution (GPD), i.e.,

$$G(y; \mu, \sigma, k) = \begin{cases} 1 - (1 - k(y-\mu)/\sigma)^{1/k} & \text{if } k \neq 0, \sigma > 0 \\ 1 - exp(-(y-\mu)/\sigma) & \text{if } k = 0, \sigma > 0 \end{cases}$$

with $\mu \leq y < \infty$ if $k \leq 0$, $\mu \leq y \leq \mu + \sigma/k$ if $k > 0$

A restriction on F

- Index of regular variation: If for $x > 0$, $lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{\alpha}$ we say that $1 - F$ is regularly varying at ∞ with index α.If $\alpha = 0$ we say that $1 - F$ is slowly varying at ∞.
- $F \in \Phi_{\alpha}(x) \Leftrightarrow lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\alpha} \Leftrightarrow x^{\alpha}(1 - F(x))$ is slowly varying at ∞.
- If $F \in D(\Psi_{\alpha})$ its endpoint u_{∞} is finite.
- If $F \in D(\Lambda)$ and its endpoint u_{∞} is not finite, $1 - F$ is rapidly varying, a situation we will (must?) avoid.
- If F belongs to the domain of attraction of a Fréchet distribution (Φ_{α}) with parameter α, then $k = -\frac{1}{\alpha}$ and $\sigma(\xi) = \xi/\alpha$.
- An estimator for $q(a)$ can be obtained from the estimation of the parameters k and $\sigma(\xi)$.

Estimation procedure

First stage: a) We consider the local linear (LL) estimator $\hat{m}(\mathbf{x}) \equiv \hat{\beta}_0$ where

$$(\hat{\beta}_0, \hat{\beta}) \equiv \underset{\beta_0, \beta}{\operatorname{argmin}} \sum_{t=1}^{n} \left(Y_t - \beta_0 - (\mathbf{X}_t^T - \mathbf{x}^T)\beta \right)^2 K_1 \left(\frac{\mathbf{X}_t - \mathbf{x}}{h_{1n}} \right),$$

$K_1(\cdot)$ is a multivariate kernel function and $h_{1n} > 0$ is a bandwidth.

b) We obtain $\{\hat{U}_t \equiv Y_t - \hat{m}(\mathbf{X}_t)\}_{t=1}^n$ and define $\hat{h}(\mathbf{x}) \equiv \hat{\eta}$ where $(\hat{\eta}, \hat{\eta}_1) \equiv \underset{\eta, \eta_1}{\operatorname{argmin}} \sum_{t=1}^{n} \left(\hat{U}_t^2 - \eta - (\mathbf{X}_t^T - \mathbf{x}^T)\eta_1 \right)^2 K_2 \left(\frac{\mathbf{X}_t - \mathbf{x}}{h_{2n}} \right)$,
$K_2(\cdot)$ is a multivariate kernel function and $h_{2n} > 0$ is a bandwidth.

Estimation procedure

Second stage:

- We use $\{\hat{\varepsilon}_t\}_{t=1}^n$ to estimate F as

$$\tilde{F}(u) = \frac{1}{nh_{3n}} \sum_{t=1}^{n} \int_{-\infty}^{u} K_3\left(\frac{\hat{\varepsilon}_t - y}{h_{3n}}\right) dy \qquad (1)$$

  where $K_3(\cdot)$ is a univariate kernel and $h_{3n} > 0$ is a bandwidth.
- Let $\tilde{q}(a)$ be the solution for $\tilde{F}(\tilde{q}(a)) = a$. Letting $0 < a_n < a < 1$ be such that $a_n \to 1$ as $n \to \infty$ we use N_s residuals that exceed $\tilde{q}(a_n)$ to form

$$\{\tilde{Z}_i\}_{i=1}^{N_s} = \{\hat{\varepsilon}_{(n-N_s+i)} - \tilde{q}(a_n)\}_{i=1}^{N_s}$$

  where $\{\hat{\varepsilon}_{(t)}\}_{t=1}^n$ denotes the order statistics associated with $\{\hat{\varepsilon}_t\}_{t=1}^n$.

Estimation procedure

- $\{\tilde{Z}_i\}_{i=1}^{N_s}$ is used to obtain maximum likelihood estimators for σ and k based on $g(z;\sigma,k) = \frac{1}{\sigma}\left(1 - \frac{kz}{\sigma}\right)^{1/k-1}$. That is,

$$\frac{\partial}{\partial\sigma}\frac{1}{N_s}\sum_{i=1}^{N_s} log\, g(\tilde{Z}_i;\tilde{\sigma}_{\tilde{q}(a_n)},\tilde{k}) = 0 \tag{2}$$

$$\frac{\partial}{\partial k}\frac{1}{N_s}\sum_{i=1}^{N_s} log\, g(\tilde{Z}_i;\tilde{\sigma}_{\tilde{q}(a_n)},\tilde{k}) = 0. \tag{3}$$

- Based on Pickands approximation

$$F_{\tilde{q}(a_n)}(y) = \frac{F(y+\tilde{q}(a_n)) - F(\tilde{q}(a_n))}{1 - F(\tilde{q}(a_n))} \approx 1 - \left(1 - \frac{ky}{\sigma_{\tilde{q}(a_n)}}\right)^{1/k}$$

Estimation procedure

- For $a \in (a_n, 1)$, $q(a) = \tilde{q}(a_n) + y_{\tilde{q}(a_n),a}$ where by construction $F(\tilde{q}(a_n) + y_{\tilde{q}(a_n),a}) = a$.
  Then,
  $$\frac{1-a}{1-F(\tilde{q}(a_n))} \approx \left(1 - \frac{k\ y_{\tilde{q}(a_n),a}}{\sigma_{\tilde{q}(a_n)}}\right)^{1/k}.$$
- If F is estimated by $\tilde{F}$, and noting that $1-\tilde{F}(\tilde{q}(a_n)) = 1-a_n$, we have
  $$y_{\tilde{q}(a_n),a} \approx \frac{\sigma_{\tilde{q}(a_n)}}{k}\left(1 - \left(\frac{1-a}{1-a_n}\right)^k\right).$$
- Thus,
  $$\hat{q}(a) = \tilde{q}(a_n) + \frac{\tilde{\sigma}_{\tilde{q}(a_n)}}{\tilde{k}}\left(1 - \left(\frac{1-a}{1-a_n}\right)^{\tilde{k}}\right).$$

Estimation procedure

- If $\varepsilon_t - q(a)$ were distributed *exactly* as $g(z; \sigma, k)$, then integration by parts gives

$$E(\varepsilon_t | \varepsilon_t > q(a)) = q(a) + \frac{\sigma}{1+k}$$

- If $F \in D(\Phi_\alpha)$ and restrictions are placed on how fast $x^\alpha(1 - F(x))$ is slowly varying, it is easy to show that

$$E(\varepsilon_t | \varepsilon_t > q(a)) = \frac{q(a)}{1+k} + o(1)$$

  This motivates the estimator

$$\widehat{E}(\varepsilon_t | \varepsilon_t > q(a)) = \frac{\hat{q}(a)}{1+\tilde{k}}.$$

Assumptions

There are a number of regularity conditions.

- Assumption 1 deals with the properties of the three kernels used in estimation. The order s for K_1 and K_2 are needed to establish that the biases for $\hat{m}$ and $\hat{h}$ are, respectively, of order $O(h_{in}^s)$ for $i = 1, 2$ in Lemmas 2 and 3. The order m_1 for K_3 is necessary in the proof of Lemma 4.
- Assumption 3 restricts m and h to be sufficiently smooth.
- Assumption A5 is necessary in Lemma 4 and is directly related to the verification of existence of bounds required to use Lemma A.2 in Gao (2007).

Assumptions

- Assumption 2 is important:
  1) $\{(\mathbf{X}_t \; \varepsilon_t)^T\}_{t=1,2,\dots}$ is a strictly stationary α-mixing process with $\alpha(l) \leq C\, l^{-B}$ for some $B > 2$;
  2) The joint density of $\mathbf{X}_t$ and ε_t is given by $f_{\mathbf{X}\varepsilon}(\mathbf{x}, \varepsilon) = f_{\mathbf{X}}(\mathbf{x}) f(\varepsilon)$;
  3) $f_{\mathbf{X}}(\mathbf{x})$ and all of its partial derivatives of order $< s$ are differentiable and uniformly bounded on $\mathbb{R}^d$;
  4) $0 < \inf_{\mathbf{x}\in\mathcal{G}} f_{\mathbf{X}}(\mathbf{x})$ and $\sup_{\mathbf{x}\in\mathcal{G}} f_{\mathbf{X}}(\mathbf{x}) \leq C$.
  A2 1) implies that for some $\delta > 2$ and $a > 1 - \frac{2}{\delta}$, $\sum_{j=1}^{\infty} j^a \alpha(j)^{1-\frac{2}{\delta}} < \infty$, a fact that is needed in our proofs. We note that α-mixing is the weakest of the mixing concepts Doukhan (1994) and its use here is only possible due to Lemma A.2 in Gao (2007), which plays a critical role in the proof of Lemma 4.

Lemmata

Lemma 2: Assume that the kernel K_1 used to define $\hat{m}$ satisfies assumption A1 and assumptions A2 and A3 are holding. Assume also that the bandwidth h_{1n} used to define $\hat{m}$ satisfies equations

$$n^{1-\frac{2}{a}-2\theta} h_n^d \longrightarrow \infty$$

and

$$n^{(B+1.5)(\frac{1}{a}+\theta)-\frac{B}{2}+0.75+\frac{d}{2}} h_n^{-1.75d-\frac{d}{2}(d+B)} (log\ n)^{0.25+0.5(B-d)} \longrightarrow 0.$$

Then, if $E(|\varepsilon_t|^a) < \infty$, $E(h^{1/2}(\mathbf{X}_t)^a) < \infty$ for some $a > 2$ and condition c) in Lemma 1 is holding

$$\sup_{\mathbf{x}\in\mathcal{G}} |\hat{m}(\mathbf{x}) - m(\mathbf{x})| = O_p\left(L_{1n}\right), \qquad (4)$$

where $L_{1n} = \left(\frac{log\ n}{nh_{1n}^d}\right)^{1/2} + h_{1n}^s$.

Lemmata

Lemma 3: Assume that the kernel K_2 used to define $\hat{h}$ satisfies assumption A1 and assumptions A2 and A3 are holding. Assume also that the bandwidth h_{2n} used to define $\hat{h}$ satisfies equations

$$n^{1-\frac{2}{a}-2\theta} h_n^d \longrightarrow \infty$$

and

$$n^{(B+1.5)(\frac{1}{a}+\theta)-\frac{B}{2}+0.75+\frac{d}{2}} h_n^{-1.75d-\frac{d}{2}(d+B)} (log\ n)^{0.25+0.5(B-d)} \longrightarrow 0.$$

Then, under the assumptions in Lemma 2, if $E(|\varepsilon_t^2 - 1|^a) < \infty$ and $E(h(\mathbf{X}_t)^a) < \infty$ for some $a > 2$,

$$\sup_{\mathbf{x}\in\mathcal{G}} |\hat{h}(\mathbf{x}) - h(\mathbf{x})| = O_p\left(L_{1n} + L_{2n}\right), \quad (5)$$

where $L_{1n} = \left(\frac{log\ n}{nh_{1n}^d}\right)^{1/2} + h_{1n}^s$ and $L_{2n} = \left(\frac{log\ n}{nh_{2n}^d}\right)^{1/2} + h_{2n}^s$.

Lemmata

Corollary: Under the assumptions of Lemma 3,

$$\sup_{\mathbf{x}\in\mathcal{G}} |\hat{h}^{1/2}(\mathbf{x}) - h^{1/2}(\mathbf{x})| = O_p\left(L_{1n} + L_{2n}\right)$$

and

$$\sup_{\mathbf{x}\in\mathcal{G}} |\chi_{\{\hat{h}(\mathbf{x})>0\}} - 1| = O_p\left(L_{1n} + L_{2n}\right),$$

where $L_{1n} = \left(\frac{\log n}{nh_{1n}^d}\right)^{1/2} + h_{1n}^s$ and $L_{2n} = \left(\frac{\log n}{nh_{2n}^d}\right)^{1/2} + h_{2n}^s$.

- FR1': Assume that for some $\alpha > 0$ we have $\lim_{x\to\infty} \frac{xf(x)}{1-F(x)} = \alpha$.

Lemmata

Lemma 4: Under assumptions A1-A5 and conditions FR1' and FR2, if $\alpha \geq 1$ we have

$$N^{1/2}\left(\frac{\tilde{q}\left(a_n\right)-q_n\left(a_n\right)}{q\left(a_n\right)}\right)=O_p(1),\ \text{where } a_n = 1-\frac{N}{n}.$$

provided that a) $h_{1n} \propto n^{-\frac{1}{2s+d}}$, $h_{2n} \propto n^{-\frac{1}{2s+d}}$, $h_{3n} \propto n^{-\frac{s}{2(2s+d)}+\delta}$, $N \propto n^{\frac{2s}{2s+d}-\delta}$ for some $\delta > 0$ with $s \geq 2d$ and b) $E(|\varepsilon_t^2 - 1|^a) < \infty$ and $E(h(\mathbf{x})^a) < \infty$ for some $a > 2$.

Theorem 1

Assume that FR1' with $\alpha > 1$, FR2 and assumptions A1-A5 are holding. In addition, assume that a) $h_{1n} \propto n^{-\frac{1}{2s+d}}$, $h_{2n} \propto n^{-\frac{1}{2s+d}}$, $h_{3n} \propto n^{-\frac{s}{2(2s+d)}+\delta}$, $N \propto n^{\frac{2s}{2s+d}-\delta}$ for some $\delta > 0$ and $s \geq 2d$, b) $E(|\varepsilon_t^2 - 1|^a) < \infty$ and $E(h(\mathbf{x})^a) < \infty$ for some $a > 2$. Let $\tau_1, \tau_2 \in \mathbb{R}$, $0 < \delta_N \to 0$, $\delta_N N^{1/2} \to \infty$ as $N \to \infty$ and denote arbitrary σ and k by $\sigma = \sigma_N(1 + \tau_1\delta_N)$ and $k = k_0 + \tau_2\delta_N$. We define the log-likelihood function

$$\tilde{L}_{TN}(\tau_1, \tau_2) = \frac{1}{N}\sum_{i=1}^{N_s} log\, g(\tilde{Z}_i; \sigma_N(1 + \tau_1\delta_N), k_0 + \tau_2\delta_N),$$

where $\tilde{Z}_i = \hat{\varepsilon}_{(n-N_s+i)} - \tilde{q}(a_n)$, $a_n = 1 - \frac{N}{n}$. Then, as $n \to \infty$, $\frac{1}{\delta_N^2}\tilde{L}_{TN}(\tau_1, \tau_2)$ has, with probability approaching 1, a local maximum (τ_1^*, τ_2^*) on $S_T = \{(\tau_1, \tau_2) : \tau_1^2 + \tau_2^2 < 1\}$ at which $\frac{1}{\delta_N^2}\frac{\partial}{\partial\tau_1}\tilde{L}_{TN}(\tau_1^*, \tau_2^*) = 0$ and $\frac{1}{\delta_N^2}\frac{\partial}{\partial\tau_2}\tilde{L}_{TN}(\tau_1^*, \tau_2^*) = 0$.

Comments on Theorem 1

- The vector (τ_1^*, τ_2^*) implies a value $\tilde{\sigma}_{\tilde{q}(a_n)}$ and $\tilde{k}$ which are solutions for the likelihood equations

$$\frac{\partial}{\partial \sigma} \frac{1}{N} \sum_{j=1}^{N_s} log\, g(\tilde{Z}_j; \tilde{\sigma}_{\tilde{q}(a_n)}, \tilde{k}) = 0, \frac{\partial}{\partial k} \frac{1}{N} \sum_{j=1}^{N_s} log\, g(\tilde{Z}_j; \tilde{\sigma}_{\tilde{q}(a_n)}, \tilde{k}) = 0.$$

- There exists, with probability approaching 1, a local maximum $(\tilde{\sigma}_N = \sigma_N(1 + t^*\delta_N),\ \tilde{k} = k_0 + \tau^*\delta_N)$ on $S_R = \{(\sigma,\ k) : \|(\frac{\sigma}{\sigma_N} - 1,\ k - k_0)\| < \delta_N\}$ that satisfy the first order conditions
- Theorem 1 states that the solutions for the first order conditions correspond to a local maximum of the likelihood associated with the GPD in a shrinking neighborhood of the arbitrary point (σ_N, k_0).

Comments on Theorem 1

The proof of Theorem 1 depends critically on:

- $$\sup_{\mathbf{x}\in\mathcal{G}}|\hat{m}(\mathbf{x})-m(\mathbf{x})| = O_p\left(L_{1n}\right) \text{ and } \sup_{\mathbf{x}\in\mathcal{G}}|\hat{h}(\mathbf{x})-h(\mathbf{x})| = O_p\left(L_{2n}\right),$$

  where $L_{1n} = \left(\frac{\log n}{nh_{1n}^d}\right)^{1/2} + h_{1n}^s$ and $L_{2n} = \left(\frac{\log n}{nh_{2n}^d}\right)^{1/2} + h_{2n}^s$.
- These orders are sufficient for

  $$|\hat{\varepsilon}_t - \varepsilon_t| = O_p(L_{1n}) + (O_p(L_{1n}) + O_p(L_{2n}))|\varepsilon_t|$$

  uniformly in $\mathcal{G}$.
- Lemma 4 shows that $\tilde{q}(a_n)$ is asymptotically close to $q_n(a_n)$ by satisfying $\frac{\tilde{q}(a_n)-q_n(a_n)}{q_n(a_n)} = O_p(N^{-1/2})$

  It is here that the stochasticity of the threshold ($\tilde{q}$) is handled and FR1', FR2 and $\alpha > 1$ is used.

Asymptotic normality of $\tilde{\gamma}' = (\tilde{\sigma}_N, \tilde{k})$ - Theorem 2

Suppose FR1' with $\alpha > 1$, FR2, A1-A5 hold and that $\frac{C}{\alpha-\rho} N^{1/2}\phi(q(a_n)) \to \mu \in \mathbb{R}$. In addition, assume that conditions a) and b) in Theorem 1 are holding. Then, the local maximum $(\tilde{\sigma}_{\tilde{q}(a_n)}, \tilde{k})$ of the GPD likelihood function, is such that for $k_0 = -\frac{1}{\alpha}$ and $\sigma_N = \frac{q(a_n)}{\alpha}$

$$\sqrt{N} \begin{pmatrix} \frac{\tilde{\sigma}_{\tilde{q}(a_n)}}{\sigma_N} - 1 \\ \tilde{k} - k_0 \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \frac{\mu(1-k_0)(1+2k_0\rho)}{1-k_0+k_0\rho} \\ \frac{\mu(1-k_0)k_0(1+\rho)}{1-k_0+k_0\rho} \end{pmatrix}, H^{-1}V_2H^{-1} \right)$$

where $V_2 = \begin{pmatrix} \frac{k_0^2-4k_0+2}{(2k_0-1)^2} & \frac{-1}{k_0(k_0-1)} \\ \frac{-1}{k_0(k_0-1)} & \frac{2k_0^3-2k_0^2+2k_0-1}{k_0^2(k_0-1)^2(2k_0-1)} \end{pmatrix}.$

Comments on Theorem 1

- It is easy to show that $H^{-1}V_2V^{-1} - H^{-1}$ is positive definite.
- Any additional bias resulting from the use of $\tilde{Z}_i$ is of second order.
- The fact that $\tilde{Z}_i$ is *not* iid as Z_i does not require the use of a CLT for dependent processes as justified in Lemma 5.

Asymptotic normality of $\hat{q}(a)$ - Theorem 3

Suppose FR1' with $\alpha > 1$, FR2, A1-A5 and $\frac{C}{\alpha-\rho} N^{1/2}\phi(q(a_n)) \to \mu$ with $k_0 = -\frac{1}{\alpha}$ and $\sigma_N = q(a_n)/\alpha$. In addition, assume that conditions a) and b) in Theorem 1 are holding. Then, if $n(1-a) \propto N$, for some $z_a > 0$

$$\sqrt{n(1-a)}\left(\frac{\hat{q}(a)}{q(a)} - 1\right) \xrightarrow{d}$$

$$N\left((-k_0)\left(-\frac{(z_a^{\rho}-1)\mu(\alpha-\rho)}{\rho} - c_b^T H^{-1} \lim_{n\to\infty} \sqrt{N} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right)\right),$$

$$k_0^2 \left(c_b^T H^{-1} V_2 H^{-1} c_b + 2 c_b^T \begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1\right)\Bigg).$$

where $c_b^T = \begin{pmatrix} -k_0^{-1}(z_a^{-1}-1) & k_0^{-2} log(z_a) + k_0^{-2}(z_a^{-1}-1) \end{pmatrix}$, $b_\sigma = E\left(\frac{\partial}{\partial\sigma} log\, g(Z_i;\sigma_N,k_0)\sigma_N\right)$ and $b_k = E\left(\frac{\partial}{\partial k} log\, g(Z_i;\sigma_N,k_0)\right)$.

Asymptotic normality of $\hat{E}(\varepsilon_t|\varepsilon_t > q(a))$ - Theorem 4

Suppose FR1' with $\alpha > 1$, FR2, A1-A5 and $\frac{C}{\alpha-\rho} N^{1/2}\phi(q(a_n)) \to \mu$ with $k_0 = -\frac{1}{\alpha}$ and $\sigma_N = q(a_n)/\alpha$. In addition, assume that conditions a) and b) in Theorem 1 are holding. Then, if $n(1-a) \propto N$, for some $z_a > 0$

$$\sqrt{n(1-a)} \left(\frac{\hat{E}(\varepsilon_t|\varepsilon_t > q(a))}{\frac{q(a)}{1+k_0}} - 1 \right) \xrightarrow{d}$$

$$N \left(k_0 \frac{(z_a^\rho - 1)\mu(\alpha - \rho)}{\rho} + k_0 c_b^T H^{-1} \lim_{n\to\infty} \sqrt{N} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \right.$$

$$\left. - \frac{1}{1+k_0} \lim_{n\to\infty} \sqrt{N} \begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}, \Sigma \right),$$

where c_b, b_σ, b_k are as defined in Theorem 3,

$$\Sigma = k_0^2 \left(c_b^T H^{-1} V_2 H^{-1} c_b + \right.$$

$$\left. 2 c_b^T \begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1 \right) + 2 \frac{k_0}{1+k_0} \eta^T V_3 \theta + \frac{1}{(1+k_0)^2} \theta^T V_1 \theta,$$

with

$$\eta^T = \left(\begin{array}{ccc} -c_b^T H^{-1} & -c_b^T H^{-1} \left(\begin{array}{c} b_\sigma \\ b_k \end{array} \right) & 1 \end{array} \right),$$

$$\theta^T = \left(\begin{array}{ccc} \left(\begin{array}{cc} 0 & 1 \end{array} \right) H^{-1} & \left(\begin{array}{cc} 0 & 1 \end{array} \right) H^{-1} \left(\begin{array}{c} b_1 \\ b_2 \end{array} \right) & 0 \end{array} \right),$$

$$V_3 = \left(\begin{array}{cccc} \frac{1}{1-2k_0} & -\frac{1}{(k_0-1)(2k_0-1)} & 0 & 0 \\ -\frac{1}{(k_0-1)(2k_0-1)} & \frac{2}{(k_0-1)(2k_0-1)} & 0 & 0 \\ 0 & 0 & k_0^2 & -k_0 \\ 0 & 0 & -k_0 & 1 \end{array} \right),$$

$b_1 = \frac{1-k_0}{k_0(2k_0-1)}$ and $b_2 = \frac{1}{k_0^2}\left(\frac{k_0-1}{2k_0-1} - \frac{1}{k_0-1}\right)$.

Consistency

Given Lemmas 2, 3, Theorems 3 and 4 we have that for all $a \in (0,1)$,
$\hat{q}_{Y_t|\mathbf{X}_t=\mathbf{x}}(a) = \hat{m}(\mathbf{x}) + \hat{h}^{1/2}(\mathbf{x})\hat{q}(a) \xrightarrow{p} m(\mathbf{x}) + h^{1/2}(\mathbf{x})q(a) = a$-CVaR($\mathbf{x}$)
and
$\hat{E}(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)) = \hat{m}(\mathbf{x}) + \hat{h}^{1/2}(\mathbf{x})\hat{E}(\varepsilon_t|\varepsilon_t > q(a)) \xrightarrow{p} m(\mathbf{x}) + h^{1/2}(\mathbf{x})E(\varepsilon_t|\varepsilon_t > q(a)) = a$-CES($\mathbf{x}$).

Comments

- Estimation of $q_{Y|X=x}(a)$ when a is in the vicinity of 0 has been considered by Chernozhukov (2005) when $q_{Y|X=x}(a) = x\beta(a)$, $\beta(a) \in \Re^d$.
- He provides a complete asymptotic characterization of the quantile regression estimator of $\beta(a)$.
- Here $q_{Y|X=x}(a)$ is nonparametric, it is in this sense more general than the one considered by Chernozhukov. a approaches 1 at a speed that is slower then the sample size ($n(1-a) \propto N \to \infty$ in Theorem 2).
- Furthermore, similar to Smith (1987) and Hall (1982), our proofs require the specification of the speed at which the tail $1 - F(x)$ behaves asymptotically as a power function. Theorem 1 specifies this speed to be proportional to $\sqrt{N}$.

Table 7 Backtest results for $a-$conditional Value-at-Risk (q) and expected shortfall(E) on $m-n=500$ observations, expected violations $=(m-n)(1-a)$. q: Number of violations and p-value (in brackets). E: p-value for exceedance residuals to have zero mean.

	q			E		
	$a=0.95$	$a=0.99$	$a=0.995$	$a=0.95$	$a=0.99$	$a=0.995$
	Expected violations					
	25	5	2.5			
Maize	18 (.151)	5(1)	2(.751)	0	.161	.735
Rice	29(.412)	4(.653)	2(.751)	0	.081	.248
Soybean	21(.412)	3(.369)	2(.751)	0	.302	.244
Wheatcbot	30(.305)	6(.653)	2(.751)	.001	.339	.273
Wheatkcbt	25(1)	5(1)	2(.751)	0	.082	.239

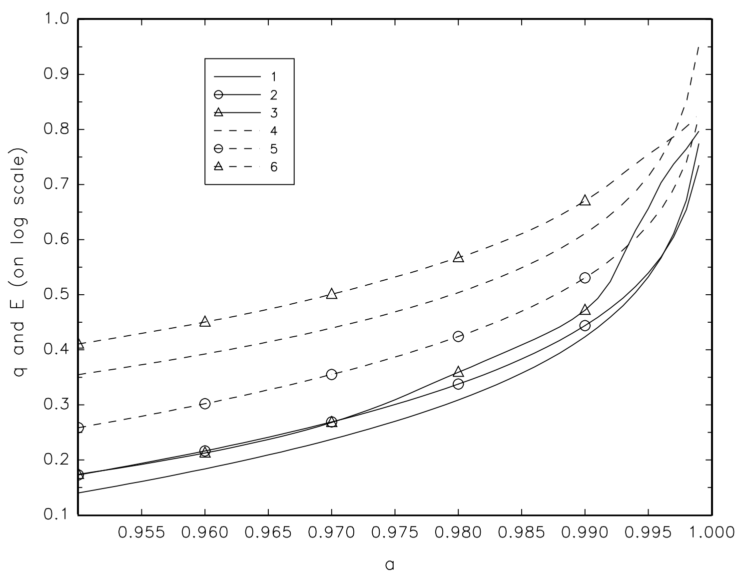

Figure 1: Plot of conditional value-at-risk (q) and expected shortfall (E) estimates evaluated at sample mean across different a, with $n=1000$, $h_1(Y_{t-1})=1+0.01Y_{t-1}^2+0.5sin(Y_{t-1})$, $\theta=0$ and student-t distributed ε_t with $v=3$. 1 : true q, $2:\hat{q}$, $3:\dot{q}$, 4 : true E, $5:\hat{E}$, and $6:\dot{E}$.