Functional Extreme Partial Least-Squares

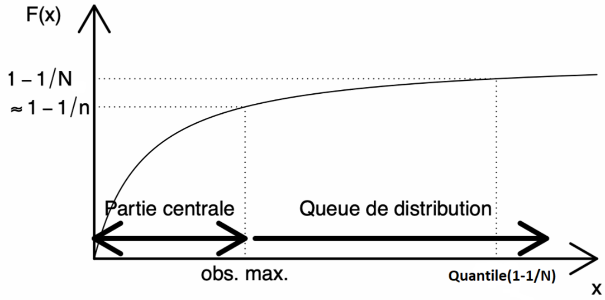
Cambyse Pakzad, Stéphane Girard (Inria Grenoble, team Statify), https://cpakzad.github.io

June 28th, 2024 - ISNPS Braga, Portugal

Extreme risk

Let Y be a random variable with cdf $F(y) = \mathbb{P}(Y \leq y)$ and $\bar{F} = 1 - F$ its survival function.

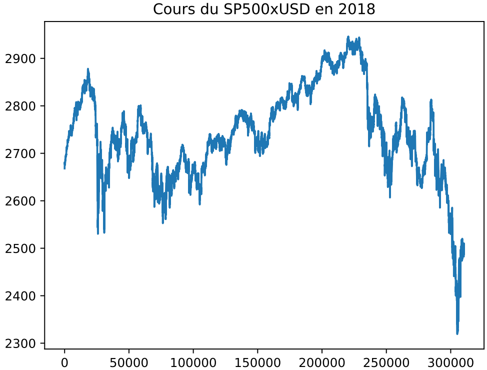
- Y models the loss of a financial asset or the rainfall height.
- Risk measure : cover against a large increase of the response Y values.
- Quantile at level $\alpha \in (0,1)$ satisfies $\mathbb{P}(Y \leq q(\alpha)) = \alpha$, i.e., $q(\alpha) = F^{-1}(\alpha)$ (generalized inverse).
- Extremes : $\alpha \to 1$ and Y heavy-tailed i.e. $\lim_{t\to+\infty} \bar{F}(tx)/\bar{F}(t) = x^{-1/\gamma}$ (regular variation).



Functional covariate

$Y = \max_{t \in \mathcal{T}} \log(p_t/p_{t-1})$ where p_t is the price value of an asset at time t and $\mathcal{T}$ large time domain.

- Include a massive auxiliary information $\rightarrow$ functional covariate.
- $\boldsymbol{X} \in H$ with H separable Hilbert space, e.g., $L^2([0,1])$.
- In practice: $60 \times 24 = 1440$ and $\boldsymbol{X} \in \mathbb{R}^{1440}$ stockprice per minute during even days.



Dimension reduction

Let Y with cdf $F = 1 - \bar{F}$ and $\bar{F}$ its survival function. Let $\boldsymbol{X} \in H$ with H separable Hilbert space.

Classical goal: Statistical inference of the risk measure (e.g., quantile) of Y conditionally to $\boldsymbol{X}$ for large threshold ($\alpha \to 1$).

- ✖ Hindrances: Computational cost; Double sparsity: Curse of dimension + Extremes.
- ➥ Substitute the covariate $\boldsymbol{X} \in H$ by a projection $\langle \boldsymbol{w}, \boldsymbol{X} \rangle_H \in \mathbb{R}$.
- ➥ FEPLS method: find a direction in H that best explains the extreme behaviour of Y according to $\boldsymbol{X}$. PLS is adaptated to the case where $\boldsymbol{X}$ is functional and $\bar{F}$ is regularly varying.

Theoretical FEPLS

Assume that Y is heavy-tailed (but not too much) with tail index $\gamma < 1$ so that Y is integrable.

✍ **Tail-moment** : $m_W(y) := \mathbb{E}(W 1_{\{Y>y\}})$ for large $y > 0$ and W generical random variable.

☞ **FEPLS method**: $\underset{\|\boldsymbol{w}\|_H=1}{\arg\max} \operatorname{Cov}\left(\langle \boldsymbol{w}, \boldsymbol{X}\rangle, Y \mid Y \geq y\right)$ with $y \to +\infty$.

✔ **Unique explicit solution**: $\boldsymbol{v}(y)/\|\boldsymbol{v}(y)\|$ with $\boldsymbol{v}(y) = \bar{F}(y)\boldsymbol{m}_{XY}(y) - \boldsymbol{m}_X(y)m_Y(y)$.

▷ We stay inside Span$(\boldsymbol{X}) \subset H$. Hence we consider, for $\varphi : \mathbb{R} \to \mathbb{R}$ test regularly varying,

$$\boldsymbol{v}_\varphi(y) = m_{\boldsymbol{X}\varphi(Y)}(y) = \mathbb{E}_{\mathcal{B}}(\boldsymbol{X}\varphi(Y)1_{\{Y>y\}}).$$

Here, $\mathbb{E}_{\mathcal{B}}$ denotes the integral in the sense of Bochner/Pettis (for Banach-valued rvs).

FEPLS inference

Let an iid sample $(\boldsymbol{X}_i, Y_i) \subset H \times \mathbb{R}$. We seek to estimate the FEPLS: $\boldsymbol{f}_\varphi = \boldsymbol{v}_\varphi / \|\boldsymbol{v}_\varphi\|$.

- **Theoretical target**: $\boldsymbol{v}_\varphi(y) = \mathbb{E}_\mathcal{B}(\boldsymbol{X}\varphi(Y)1_{\{Y>y\}})$ with large $y > 0$.
- **Empirization**: $\hat{\boldsymbol{v}}_\varphi(y_n) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i\varphi(Y_i)1_{\{Y_i>y_n\}}$ with deterministic $y_n \gg 1$.
- **Interpretation**: empirical FEPLS = linear comb. of $\boldsymbol{X}_i$ with associated Y_i in the distribution tail + weights to each extreme observations through φ.
- **Threshold choice**: Assume $n\mathbb{P}(Y > y_n) \gg 1$ so that the average number of extreme observations $Y_i > y_n$ increases with the sample size (the threshold must not grow too fast so that we dispose of data for the inference).

Theoretical guarantees

We express our results under an inverse regression model. Denote $\boldsymbol{f}_\varphi = \boldsymbol{v}_\varphi / \|\boldsymbol{v}_\varphi\|$.

- **Inverse model**: $\boldsymbol{X} = g(Y)\beta + \varepsilon$ with:
  - $g : \mathbb{R} \to \mathbb{R}$ link function regularly varying, e.g., $g(t) = t^\kappa$, $\kappa > 0$.
  - β a deterministic unit vector in H. Span(β) is the space of dimension reduction.
  - ε is a random noise in H, e.g., Brownian motion etc...
- Inspired from Sliced Inverse Regression (SIR).
- Against the philo. of Fisher : Cook (2007) "Fisher Lecture: Dimension Reduction in Regression".

Heuristic I : If $\varepsilon \perp Y$ and $\mathbb{E}_{\mathcal{B}}(\varepsilon) = 0$, then $\boldsymbol{f}_\varphi(y) = \beta$ for any test function φ regularly varying.

Heuristic II : More generally, if ε has small contributions in the extremes of Y. Then,

$$\boldsymbol{f}_\varphi(y) \xrightarrow[y \to +\infty]{H} \beta.$$

Limit results

Let $y_n \gg 1$ deterministic with $n\bar{F}(y_n) \gg 1$ and $\hat{\boldsymbol{f}}_\varphi := \hat{\boldsymbol{v}}_\varphi / \|\hat{\boldsymbol{v}}_\varphi\|_H$. Under the model $\boldsymbol{X} = g(Y)\boldsymbol{\beta} + \boldsymbol{\varepsilon}$,

Consistency I : For some speed rate $\delta_n \gg 1$,

$$\delta_n \cdot \left\| \hat{\boldsymbol{f}}_\varphi(y_n) - \boldsymbol{\beta} \right\|_H \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

- The rate δ_n depends on the threshold y_n, on the tail-index of Y, on the link function g, on the integrability order of the noise ε but not on the test function/weight φ.

Consistency II : Moreover,

$$\delta_n \left| \frac{\operatorname{Cov}(Y, \langle \hat{\boldsymbol{f}}_\varphi(y_n), \boldsymbol{X} \rangle \mid Y \geq y_n)}{\operatorname{Cov}(Y, \langle \boldsymbol{\beta}, \boldsymbol{X} \rangle \mid Y \geq y_n)} - 1 \right| \xrightarrow[n\to+\infty]{} 0.$$

- Projecting onto $\mathrm{Span}(\hat{\boldsymbol{f}}_\varphi(y_n))$ instead of $\mathrm{Span}(\boldsymbol{\beta})$ asymptotically preserves the same quantity of extrême information.

Sketch of the proof

Let $y_n \gg 1$ deterministic with $n\bar{F}(y_n) \gg 1$ and $\hat{f}_\varphi := \hat{v}_\varphi / \|\hat{v}_\varphi\|_H$ the empirical FEPLS direction.

Model $X = g(Y)\beta + \varepsilon$ where $\|\beta\|_H = 1$, $\bar{F} \in \mathrm{RV}_{-\frac{1}{\gamma}}(+\infty)$, $g \in \mathrm{RV}_\kappa(+\infty)$. Let $\varphi \in \mathrm{RV}_\tau(+\infty)$.

Goal: $\|\hat{f}_\varphi - \beta\|_H^2 = 2(1 - \langle \hat{v}_\varphi, \beta\rangle^2 / \|\hat{v}_\varphi\|_H^2) \xrightarrow[n\to+\infty]{} 0$ where $\hat{v}_\varphi := \hat{m}_{X\varphi(Y)} = \hat{m}_{\varphi g(Y)}\beta + \hat{m}_{\varphi(Y)\varepsilon}$.

After some calculations, this boils down to: $\hat{m}_{\langle\varphi(Y)\varepsilon,\beta\rangle_H} / m_{\varphi g(Y)} \to 0$ and $\|\hat{m}_{\varphi(Y)\varepsilon}\|_H / m_{\varphi g(Y)} \to 0$.

Tools: Chebychev's ineq. + appropriate speed rates. In particular, one needs to control the variance.

- Univariate regular variation results such as Karamata representation. For instance, yielding:

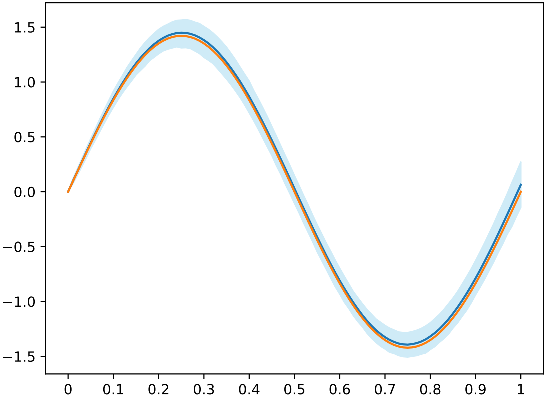
$$\sqrt{n\bar{F}(y_n)} \left(\frac{m_{\varphi g(Y)}(y_n)}{m_{\varphi g(Y)}(y_n)} - 1 \right) \text{ is asymptotically normal.}$$

- Functional part: if $W_1, W_2 \in H$ independent, then $\mathbb{E}(\langle W_1, W_2\rangle_H) = \langle \mathbb{E}_{\mathcal{B}}(W_1), \mathbb{E}_{\mathcal{B}}(W_2)\rangle_H$.

Illustration on synthetic data

Estimation of the FEPLS with data generated under the inverse model and $H = L^2([0,1])$:

$$\boldsymbol{X} = g(Y)\boldsymbol{\beta} + \varepsilon, \text{ with } \boldsymbol{\beta} \in H \text{ deterministic to be estimated and } \varepsilon \in H \text{ random noise.}$$



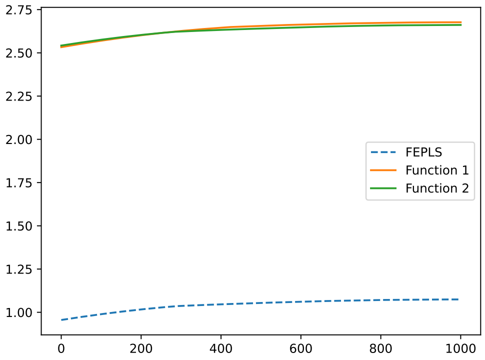
- $y_n = Y_{n-k+1:n}$ order stat. and choice of $k \geq 5$.
- Blue : $\boldsymbol{\beta}(t) := \sqrt{2}\sin(2\pi t)$, $t \in [0,1]$.
- Orange : estimation of $\boldsymbol{\beta}$.
- Blue area : confidence band; top $5 - 95\%$ of the values among 500 Monte-Carlo iterations.

Choice of the model:

- Y with Burr distribution.
- Link function: g polynomial.
- Noise: ε fractional Brownian motion depending on Y (Hurst parameter $= 1/3$).

Illustration on financial data

Comparison of quantiles: $Y|X = x$ vs $Y|\langle X, \hat{f}_\varphi \rangle = \langle x, \hat{f}_\varphi \rangle$.



- Estimation of quantiles with Nadaraya-Watson weights at some function point $x \in H$.
- Consider $x' \neq \hat{f}_\varphi$ and $x' \in \{\text{function } 1, \text{function } 2\}$.
- Projecting on $\hat{f}_\varphi$ should give 'low' error.
- Projecting on x' instead of $\hat{f}_\varphi$ should give higher error.
- Relative error in percentage.