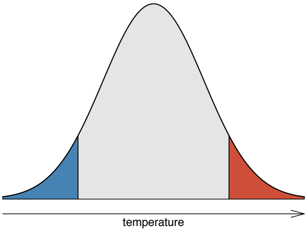
Principal components in an asymmetric norm

Ngoc Mai Tran
Maria Osipenko
Wolfgang Karl Härdle

Ladislaus von Bortkiewicz
Chair of Statistics
C.A.S.E. Centre for Applied Statistics and Economics
School of Business and Economics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

Tail events, emotions, investments



'If you hold a cat by the tail you learn things you can not learn any other way.' An english proverb.

Quantiles and Expectiles

- Quantiles and Expectiles are tail measures.
- Capture tail behavior of conditional distributions.
- Applications in
  - Finance: VaR and Expected Shortfall
  - Weather: Energy, Tourism, Agriculture
  - ...
- Some applications involve MANY curves.

Temperature Data

- Daily average temperatures
- 29 Provinces, 159 stations in China,
- from 19510101 to 20121231.

Model for temperature

- Temperature T_{it} on day t for city i:

$$T_{it} = X_{it} + \Lambda_{it}$$

- The seasonal effect Λ_{it}:

$$\Lambda_{it} = a_{i0} + a_{i1} t + a_{i2} \sin(2\pi t/365) + a_{i3} \cos(2\pi t/365)$$

- X_{it} follows an AR(p) process:

$$X_{it} = \sum_{j=1}^{p} \beta_{ij} X_{i,t-j} + \varepsilon_{it}$$

$$\hat{\varepsilon}_{it} = X_{it} - \sum_{j=1}^{p} \hat{\beta}_{ij} X_{i,t-j}$$

back

Risk factors

- Pricing weather derivatives:

$$\varepsilon_{it} \sim \mathrm{N}(0, \sigma_{it}^2)$$

- Change from "light" to "heavy" tails within a typical year
- Regions with high (low) variability of temperature extremes

(Functional) Principal Component Analysis (FPCA)

- a common tool to capture high dimensional data (curves), Ramsey & Silverman (2008),
- dimension reduction for complex data over space and time:
    - implied vola, correlation, temperature, rain, snowfall...,
- interpretability of principal components (PC),
- identification of similarities /differences via PC scores.

~~Functional~~ PCA

- Curves discretized on a regular grid of length p are vectors in $\mathbb{R}^p$ – usual PCA.

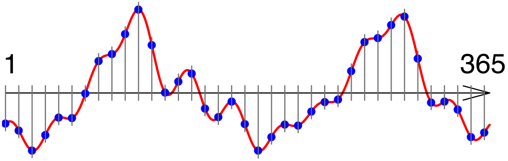
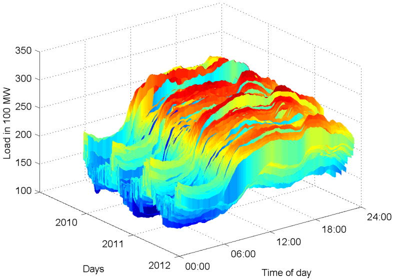


Figure 1: Average temperature curve discretized on a grid

PCA: best L_2 approximation by a K-dimensional subspace.
What about τ-quantile or τ-expectile approximation?

Applications:

- Weather derivatives / weather extremes
- Extreme events / risk modeling
- Electricity load



Quantiles and Expectiles

For X an $\mathbb{R}$-valued rv:
τ-quantile: $q(\tau) = F^{-1}(\tau)$ can also be defined as

$$q_\tau(X) = \operatorname*{argmin}_{q \in \mathbb{R}} \mathsf{E} \| X - q \|_{\tau,1}^{1},$$

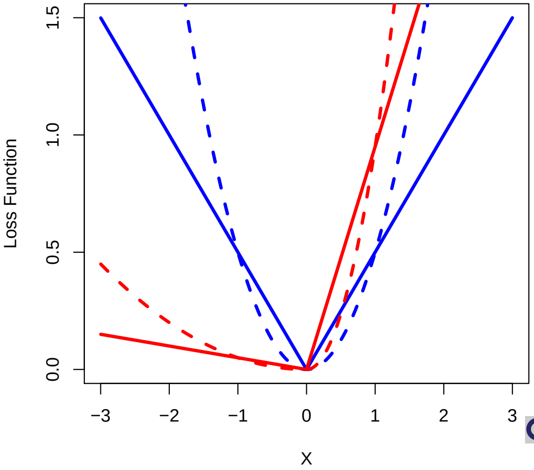
τ-expectile q vs.e:

$$e_\tau(X) = \operatorname*{argmin}_{e \in \mathbb{R}} \mathsf{E} \| X - e \|_{\tau,2}^{2}.$$

where for $\alpha = 1, 2$

$$\|x\|_{\tau,\alpha}^{\alpha} = |x|^{\alpha} \cdot \left\{ \tau \, \mathbf{I}_{\{x \geq 0\}} + (1-\tau) \, \mathbf{I}_{\{x < 0\}} \right\}.$$

For $\tau \neq 1/2$, these norms are asymmetric.



LQRcheck

Figure 2: Loss functions for $\tau = 0.9$; $\tau = 0.5$; $\alpha = 1$ (solid); $\alpha = 2$ (dashed)

‣ Expected shortfall

"Principal Components" for expectiles

- naive approach: (usual PCA on the estimated expectile curves): loose efficiency
- Principal components in an asymmetric norm:

$$\text{PCA} + \text{Expectiles} = \|\mathbf{PCA}\|_{\tau,\alpha}^{\alpha}$$

Outline

1. Motivation ✓
2. Quantiles and Expectiles
3. Algorithms for "PCA" in an asymmetric norm
4. Simulation
5. Chinese Temperature data
6. Outlook

Quantiles and Expectiles

For Y an $\mathbb{R}^p$-valued rv:

τ-quantile:

$$q_\tau(Y) = \underset{q \in \mathbb{R}^p}{\operatorname{argmin}} \mathsf{E} \| Y - q \|_{\tau,1}^1,$$

τ-expectile

$$e_\tau(Y) = \underset{e \in \mathbb{R}^p}{\operatorname{argmin}} \mathsf{E} \| Y - e \|_{\tau,2}^2.$$

where for $\alpha = 1, 2$

$$\|y\|_{\tau,\alpha}^\alpha = \sum_{j=1}^{p} |y_j|^\alpha \cdot \left\{ \tau \, \mathbf{I}_{\{y_j \geq 0\}} + (1-\tau) \, \mathbf{I}_{\{y_j < 0\}} \right\}.$$

Properties

For Y an $\mathbb{R}^p$-valued rv holds coordinatewise:

- $e_\tau(Y+t) = e_\tau(Y) + t$ for $t \in \mathbb{R}^p$.
- $e_\tau(sY) = \begin{cases} se_\tau(Y) & \text{for} \quad s \in \mathbb{R}, s > 0 \\ -se_{1-\tau}(Y) & \text{for} \quad s \in \mathbb{R}, s < 0 \end{cases}$
- $e_\tau(Y)$ is the τ-quantile of the cdf T, where

$$T(y) = \frac{G(y) - xF(y)}{2\{G(y) - yF(y)\} + \{y - \int_{-\infty}^{\infty} u\, dF(u)\}}, \tag{1}$$

$$G(y) = \int_{-\infty}^{y} u\, dF(u). \tag{2}$$

Asymptotic normality

F – differentiable cdf of a distribution with $\mu = 0$ and σ^2;
$e : (0,1) \to \mathbb{R}, \tau \mapsto e_\tau$ – the expectile function;
F_n, e_n are the empirical versions. Then for any $0 < \delta < 1$,

$$\sqrt{n}(e_n - e) \xrightarrow{\mathcal{L}} \mathcal{E},$$

over $D([\delta, 1-\delta])$, and $\mathcal{E}$ is a stochastic process on $[\delta, 1-\delta]$ with normal zero-mean marginals and variance

$$\mathrm{Var}\{\mathcal{E}(\tau)\} = \frac{\mathsf{E}\{\tau(Y - e_\tau)_+ + (1-\tau)(e_\tau - Y)_+\}^2}{[\tau\{1 - F(e_\tau)\} + (1-\tau)F(e_\tau)]^2} \tag{3}$$

for $\tau \in [\delta, 1-\delta]$.

▸ Skorokhod

PCA geometry

- PCA: minimize error vs. maximize variance

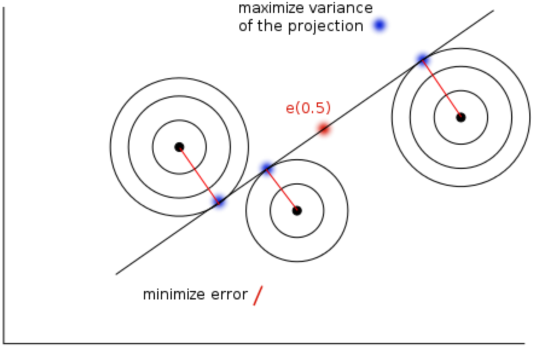


Figure 3: Best one dimensional approximation of two-dimensional variables

"PCA" as error minimizers

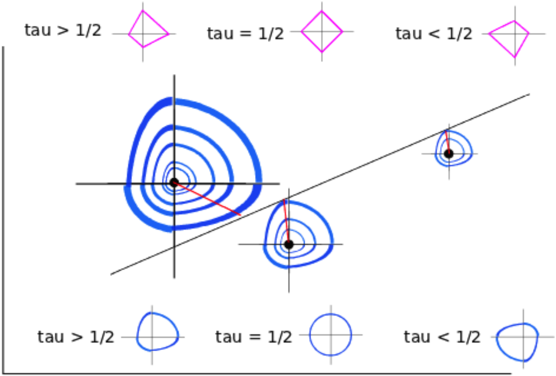


Figure 4: One dimensional approximation of two-dimensional variables in an asymmetric L_1 (magenta) and L_2 (blue) norm

"PCA" as error minimizers

Find best k-dimensional approximation Ψ_k^*:

$$\Psi_k^* = \underset{\Psi_k \in \mathbb{R}^{n \times p}:\mathrm{rank}(\Psi_k)=k+1}{\operatorname{argmin}} ||Y - \Psi_k^\top \Psi_k Y||_\tau^2$$

BUT $e_\tau(X+Y) \neq e_\tau(X) + e_\tau(Y)$ and $\Psi_k^* \not\supseteq \Psi_{k-1}^*$, thus no basis for Ψ_k^*.

Solution (via asymmetric weighted least squares: LAWS)

- Top Down (TD): first find Ψ_k^*, then find $\hat{\Psi}_1$, the best 1-D subspace contained in Ψ_k^*, and so on.
- Bottom Up (BUP): first find Ψ_1^*, then find $\hat{\Psi}_2$, the best 2-D subspace which contains Ψ_1^*, and so on.

"PCA" as variance maximizers

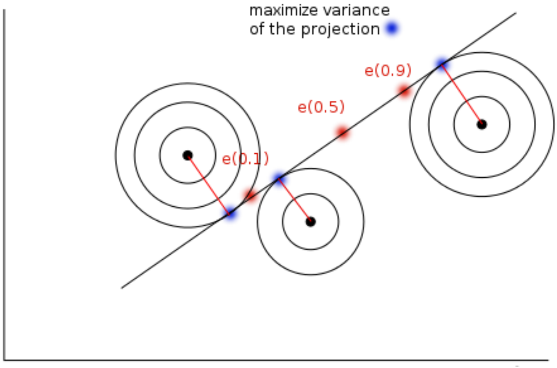


Figure 5: One dimensional approximation of two-dimensional variables in an asymmetric norm

"PCA" as variance maximizers

Define the τ-variance for $X \in \mathbb{R}$

$$\mathrm{Var}_\tau(X) = \mathrm{E}\|X - e_\tau(X)\|^2_{\tau,2}$$

The principal expectile component(PEC)

$$\begin{aligned}\phi^*_\tau &= \operatorname*{argmax}_{\phi \in \mathbb{R}^p, \phi^\top \phi = 1} \mathrm{Var}_\tau(\phi^\top Y_i, i = 1, \ldots, n) \\ &= \operatorname*{argmax}_{\phi \in \mathbb{R}^p, \phi^\top \phi = 1} \frac{1}{n} \sum_{i=1}^{n} (\phi^\top Y_i - \mu_\tau)^2 w_i,\end{aligned}$$

where $\mu_\tau \in \mathbb{R}$ is the τ-expectile of $\phi^\top Y_1, \ldots \phi^\top Y_n$, and

$$w_i = \begin{cases} \tau & \text{if } \sum_{j=1}^{p} Y_{ij}\phi_j > \mu_\tau, \\ 1 - \tau & \text{otherwise.} \end{cases}$$

PEC is weighted PC!

Given the true weights w_i and

$\mathcal{I}_\tau^+ = \{i \in \{1, \dots, n\} : w_i = \tau\}, \mathcal{I}_\tau^- = \{i \in \{1, \dots, n\} : w_i = 1-\tau\}$,

$n^+ = |\mathcal{I}_\tau^+|$ and $n^- = |\mathcal{I}_\tau^-|$, then the τ-expectile $e_\tau = e_\tau(Y) \in \mathbb{R}^p$ is:

$$e_\tau = \frac{\tau \sum_{i \in \mathcal{I}_\tau^+} Y_i + (1-\tau) \sum_{i \in \mathcal{I}_\tau^-} Y_i}{\tau n_+ + (1-\tau) n_-}.$$

ϕ_τ^* is the largest eigenvector of C_τ where

$$C_\tau = \frac{\tau}{n} \left\{ \sum_{i \in \mathcal{I}_\tau^+} (Y_i - e_\tau)(Y_i - e_\tau)^\top \right\} + \frac{1-\tau}{n} \left\{ \sum_{i \in \mathcal{I}_\tau^-} (Y_i - e_\tau)(Y_i - e_\tau)^\top \right\}.$$

Algorithm for computing PEC

Idea: start with randomly generated w_i and iterate between the following two steps.

- Compute e_τ, ϕ^*_τ and μ_τ as above,
- Update the weights w_i via:

$$w_i = \begin{cases} \tau & \text{if } \sum_{j=1}^{p} Y_{ij}\phi_j > \mu_\tau, \\ 1-\tau & \text{otherwise.} \end{cases},$$

- stop if there is no change in w_i.

▸ LAWS estimation

Properties of PEC

Random variable $Y \in \mathbb{R}^p$. Assume the PEC $\phi^*_\tau(Y)$ is unique.

- Invariance under translation: $\phi^*_\tau(Y + c) = \phi^*_\tau(Y)$ for all $c \in \mathbb{R}^p$.
- Rotational invariance: $\phi^*_\tau(BY) = B\phi^*_\tau(Y)$ for all orthogonal matrix $B \in \mathbb{R}^{p \times p}$.
  If the distribution of Y is elliptical, $\phi^*_\tau(Y) =$ classical PCA of Y for any $\tau \in (0, 1)$.
- Consistency: $\phi^*_\tau(Y_n) \to \phi^*_\tau(Y)$.

Finite sample analysis

- TopDown, BottomUp - consistency? show
- Robustness: skewness, fat tails, heteroscedasticity? show
- Relative speed, convergence rate show

Simulation

$$Y_i(t_j) = \mu(t_j) + f_1(t_j)\alpha_{1i} + f_2(t_j)\alpha_{2i} + \varepsilon_{ij}$$

with $i = 1, \ldots, n$, $j = 1, \ldots, p$ and t_j equi-spaced in [0,1].

$$\mu(t) = 1 + t + \exp\{-(t-0.6)^2/0.05\}$$

$$f_1(t) = \sqrt{2}\sin(2\pi t); \quad f_2(t) = \sqrt{2}\cos(2\pi t)$$

$$\alpha_{r,i} \sim \mathrm{N}(0, \sigma_r^2),$$

with setup (1): $\sigma_1^2 = 36$, $\sigma_2^2 = 9$ and (2): $\sigma_1^2 = 16$, $\sigma_2^2 = 9$.
Estimate k=2 components in 500 simulation runs.

Scenarios

Errors:

- $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma_\epsilon^2)$,
- $\varepsilon_{ij} \sim \mathrm{N}(0, \mu(t_j)\sigma_\epsilon^2)$,
- $\varepsilon_{ij} \sim t(5)$,
- $\varepsilon_{ij} \sim U(0, \sigma_\epsilon^2) + U(0, \sigma_\epsilon^2)$
- $\varepsilon_{ij} \sim \mathrm{logN}(0, \sigma_\epsilon^2)$

with $\sigma_\epsilon^2 = 0.5$ for setup (1) and $\sigma_\epsilon^2 = 1$ for (2).

- small sample: n=20, p=100
- medium sample: n=50, p=150
- large sample: n=100, p=200

MSE against sample

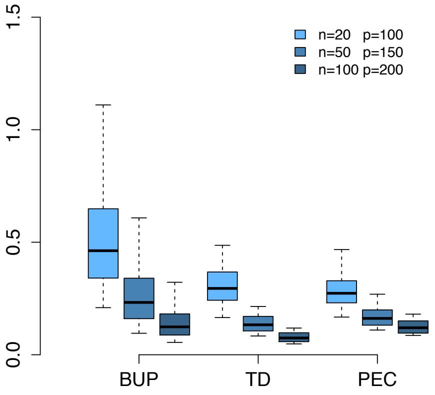


Figure 6: average MSE of BUP, TD and PEC by 500 simulations back

MSE against scenarios

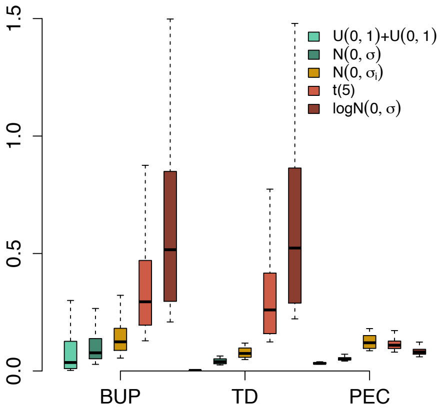
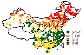


Figure 7: average MSE of BUP, TD and PEC by 500 simulations ▸ back

Computational time

sample		small			medium			large	
τ/sec	BUP	TD	PEC	BUP	TD	PEC	BUP	TD	PEC
0.90	1.24	0.70	0.57	2.91	1.59	1.39	7.53	4.02	2.71
0.95	1.64	1.13	0.55	4.01	2.68	1.57	10.53	6.88	3.03
0.98	2.36	2.05	0.56	5.56	4.59	1.56	14.62	10.96	3.54

Table 1: Average time in seconds for convergence of the algorithms (unconverged cases excluded) by 500 simulations



Convergence rate

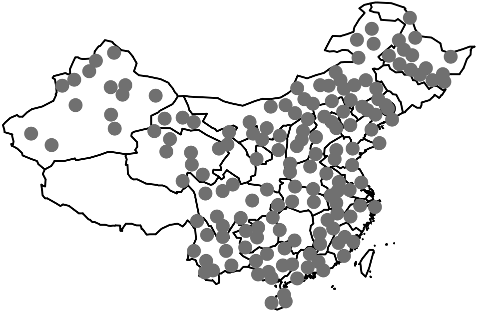
sample		small			medium			large	
τ/rate	BUP	TD	PEC	BUP	TD	PEC	BUP	TD	PEC
0.90	0.02	0.00	0.24	0.01	0.00	0.23	0.00	0.00	0.20
0.95	0.18	0.03	0.22	0.05	0.00	0.26	0.06	0.00	0.21
0.98	0.43	0.22	0.21	0.23	0.04	0.25	0.17	0.00	0.24

Table 2: Convergence rates (ratio of converged to unconverged cases by 30 iterations) of the algorithms by 500 simulation runs

back

Data

Daily average temperatures in 159 stations in China from 19510101 to 20121231 remind.



Chinese temperature data

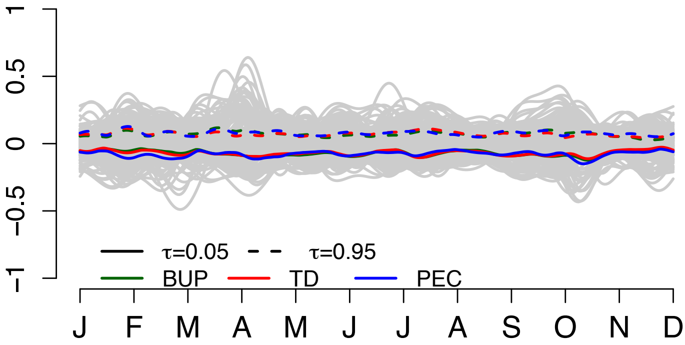


Figure 8: Averaged and smoothed temperature residual curves (gray) and the estimated average expectiles by BUP, TD and PEC for τ=0.05, 0.95

TD, BUP PCs, and PEC

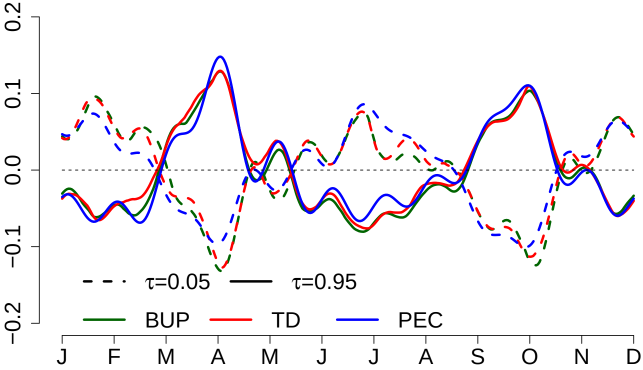


Figure 9: 1st PCs by BUP (green), TD (red) and PEC (blue), τ=0.05, 0.95

Back to interpretation

TD, BUP PCs, and PEC

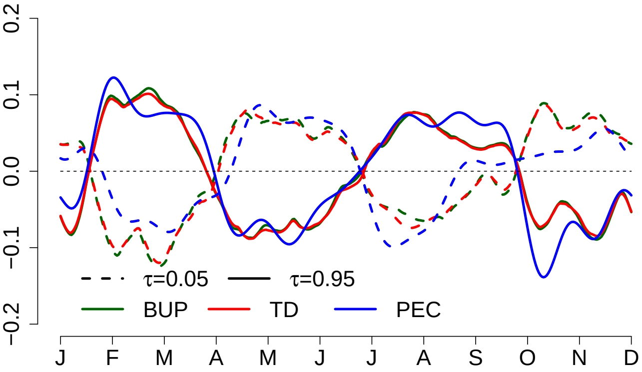


Figure 10: 2nd PCs by BUP (green), TD (red) and PEC (blue), τ=0.05, 0.95

Back to interpretation

Comparison

- Outputs of BUP, TD, and PEC are similar.
- BUP and TD PCs are particulary close to each other.
- PC(τ) $\approx$ –PC($1 - \tau$).

Interpretation

- Indicate changes in distribution from lighter to heavier tails and vice versa.
- Positive score on PC_1 – heavier tails in spring and fall, lighter in winter and summer.
- Positive score on PC_2 – heavier tails in Feb., Mar., Apr., Jul., Aug., and Sep., and lighter otherwise.

Figure - PCs

Figure - Scores

PECs for Chinese temperature data

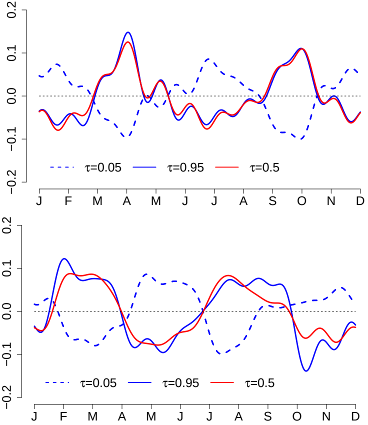


Figure 11: 1st (top), 2nd (bottom) PEC for τ=0.05 (dashed), 0.95 (solid) and 0.5

Normality?

- Observe departure of PEC from usual PC – non-normality.

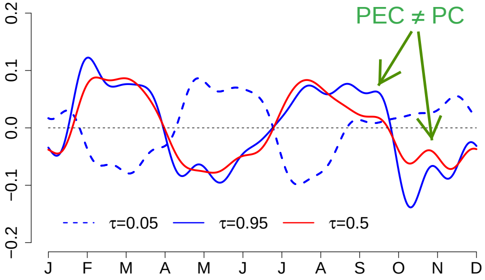
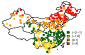


Figure 12: 2nd PEC for τ=0.05 (dashed), 0.95 (solid) and 0.5 . The latter corresponds to classical PC.

Details for PEC≠PCA



北京 - Dimension reduction

$$\mathrm{Exp}_{\mathrm{BEI},0.95} \approx \overline{\mathrm{Exp}}_{0.95} + 3.3 \times \mathrm{PEC}_{1,0.95} + 0.6 \times \mathrm{PEC}_{2,0.95}$$

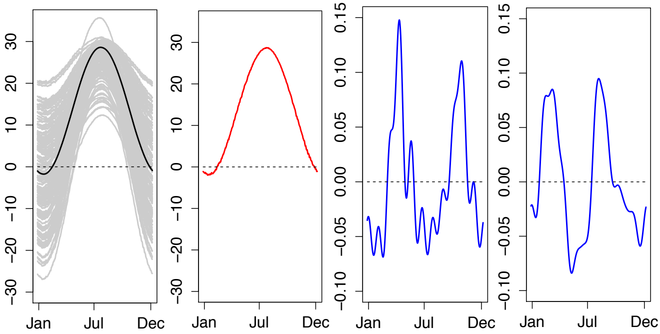


Figure 13: Approximation via PEC for the temperature expectile curve of Beijing for τ=0.95

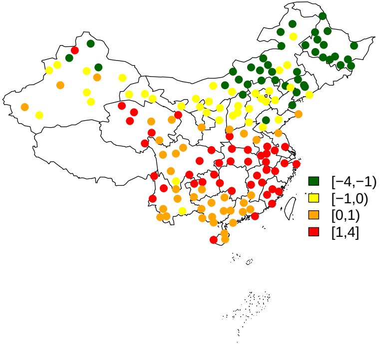
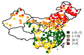


Figure 14: Scores on 1st PEC for τ=0.05 (left) and 0.95 (right) classified in four intervals

Back to interpretation



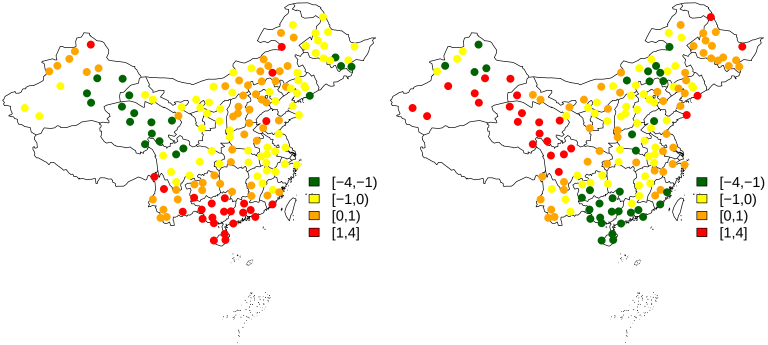


Figure 15: Scores on 2nd PEC for τ=0.05 (left) and 0.95 (right) classified in four intervals

▸ Back to interpretation

Outlook

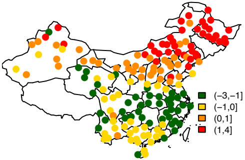
- Dimension reduction technique for tail index curves.
- Two ways to define PC for τ-expectiles: minimize error in the τ-norm (BUP and TD), and maximize the τ-variance.
- Maximize τ-variance (PEC) is a version of weighted PCA.
- PEC outperforms BUP and TD in simulations.

Outlook

- In practice the outputs of BUP, TD, and PEC do not differ much.
- Nice to study extremes of multivariate temperature data:
    - interpretability
    - normality check
    - dimension reduction

Principal components in an asymmetric norm

Ngoc Mai Tran
Maria Osipenko
Wolfgang Karl Härdle



Ladislaus von Bortkiewicz
Chair of Statistics
C.A.S.E. Centre for Applied Statistics and Economics
School of Business and Economics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

Literature

N.M. Tran, M. Osipenko and W.K. Härdle
Principal Component Analysis in an Asymmetric Norm
Discussion Paper 2014-001, CRC 649: *Economic Risk*.

M. Jones
Expectiles and M-quantiles are Quantiles
Statistics & Probability Letters, 20:149-153, 1994.

M. Guo and W. Härdle
Simultaneous Confidence Bands for Expectile Functions
Advances in Statistical Analysis, 2011,
DOI:10.1007/s10182-011-0182-1.

Literature

M. Guo, L. Zhou, J. Huang, and W. Härdle
Functional Data Analysis of Generalized Quantile Regressions
Statistics and Computing, DOI 10.1007/s11222-013-9425-1.

W. Newey and J. Powell
Asymmetric least squares estimation and testing
Econometrica, 1987, p. 819-847.

S. Schnabel and P. Eilers
Optimal Expectile Smoothing
Computational statistics and data analysis, 53(12) 2009, p.4168-4177.

Literature

F.E Benth and J.S. Benth and S. Koekebakker (2007)
Putting a price on temperature
Scandinavian Journal of Statistics

J.O. Ramsay, B.W. Silverman
Functional Data Analysis
Springer Verlag, Heidelberg, 2008

Ș. Cobzaș
Functional analysis in asymmetric normed spaces
Springer Verlag, Heidelberg, 2013

Expectile-quantile correspondence

$$\tau(s) = \frac{sq_s(Y) - \int_{-\infty}^{q_s(Y)} ydF(y)}{\mathsf{E}(Y) - 2\int_{-\infty}^{q_s(Y)} ydF(y) - (1-2s)q_s(Y)} \tag{4}$$

s-quantile corresponds to expectile with transformation $\tau(s)$ (Guo and Härdle, 2011). BACK

Expectile-quantile correspondence

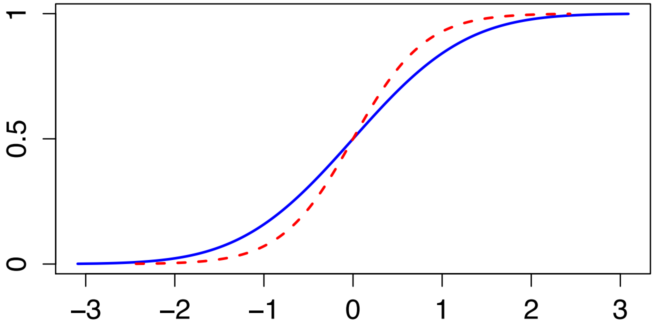


Figure 16: Quantiles (solid) and expectiles (dashed) of a normal N(0,1)

BACK

Expectile-quantile correspondence

		$\tau(s)$	
s	N(0, 1)	$t(5)$	logN(0, 1)
1%	0.12%	0.35%	0.01%
5%	1.21%	2.13%	0.09%
10%	3.31%	5.08%	0.31%
90%	96.41%	95.06%	53.57%
95%	98.65%	97.97%	66.97%
99%	99.81%	99.71%	83.53%

Table 3: s-quantile correspondence to expectile with transformation $\tau(s)$ for different distributions ▸ BACK

Relating Expectiles and Expected Shortfall

Newey and Powell (1987):

$$e_\tau = \arg\min_e \mathsf{E}\left\{\left|\tau - \mathbf{I}_{\{Y<e\}}\right|(Y-e)^2\right\}$$

$$\frac{1-2\tau}{\tau}\,\mathsf{E}\left\{(Y-e_\tau)\,\mathbf{I}_{\{Y<e_\tau\}}\right\} = e_\tau - \mathsf{E}(Y)$$

Taylor (2008):

$$\mathsf{E}\left(Y|Y<e_\tau\right) = e_\tau + \frac{\tau\left\{e_\tau - \mathsf{E}(Y)\right\}}{(1-2\tau)F(e_\tau)}$$

Back

Skorokhod space $D([0,1])$

space of real functions $f\colon [0,1] \to \mathbb{R}$
(also known as "càtlàg" functions) which

- are right-continuous
- have left limits everywhere

E.g. Cumulative distribution functions are càtlàg functions

Back

LAWS estimation

Schnabel and Eilers (2009):

$$\min \sum_{i=1}^{n} w_i(\tau)(y_i - \mu_i)^2$$

where

$$w_i(\tau) = \begin{cases} \tau & \text{if } y_i > \mu_i \\ 1-\tau & \text{if } y_i \le \mu_i, \end{cases}$$

μ_i expected value according to some model.

Iterations:

- fixed weights, closed form solution of weighted regression
- recalculate weights

until convergence criterion met.

Back to PEC

LAWS estimation

Example:

Classical linear regression model

$$Y = X\beta + \varepsilon$$

where $\mathsf{E}(\varepsilon \,|X) = 0$ and $\mu = \mathsf{E}(Y\,|X) = X\beta$.

$$\widehat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} w_i (y_i - \mu_i)^2$$

Then:

$$\widehat{\beta} = (X^\top W X)^{-1} XWY$$

with W diagonal matrix of fixed weights w_i.

Back to PEC

PEC $\neq$ PCA

Coordinate-wise $Y_{i,j}^t$ i.i.d. with some distribution of Y

$$e_{\tau,i}\{\mathsf{E}_j(Y_{ij}^t)\} \xrightarrow{\mathcal{L}} e_\tau(\bar{Y})$$

$$\mathsf{E}_i\{e_{\tau,j}(Y_{ij}^t)\} \xrightarrow{\mathcal{L}} e_\tau(Y)$$

where Y_j are i.i.d. copies of Y and $\bar{Y} = \frac{1}{J}\sum_{j=1}^{J} Y_j$

$$\text{PEC} = \text{PCA} \quad \text{iff} \quad \bar{Y} \overset{\mathcal{L}}{=} Y$$

It holds for Cauchy or $Y \overset{a.s.}{=}$ constant

Back