Principal components in an asymmetric norm

Ngoc Mai Tran
Petra Burdejova
Maria Osipenko
Wolfgang Karl Härdle

Ladislaus von Bortkiewicz Chair of Statistics
School of Business and Economics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

Instructive dependent extremes

"All situations in which the interrelationships between extremes are involved are the most interesting and instructive."

Wilhelm von Humboldt

Quantiles and Expectiles

- Quantiles and Expectiles are tail measures.
- Capture tail behavior of conditional distributions.
- Applications in
  - Finance: VaR and Expected Shortfall
  - Weather: Energy, Agriculture, Drought, Rainfall
  - Neuroscience: Risk aversion

(Functional) Principal Component Analysis (FPCA)

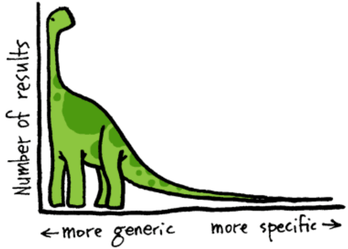
- captures high dimensional data (curves), Ramsey & Silverman (2008),
- dimension reduction for complex data over space and time,
- interpretability of principal components (PC),
- identification of similarities /differences via PC scores,
- possibility to forecast.

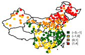
"Principal Components" for expectiles

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



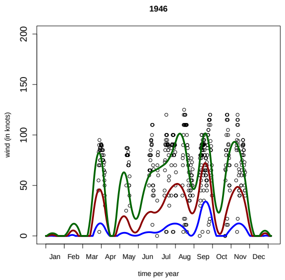


Trend of storm extremes

- Hurricane curves
- Burdejova et al. (2016)
- different linear trend for every τ-level

$$e_n^\tau(t) = \alpha_\tau(t) + n\beta_\tau(t) + \varepsilon_\tau(t)$$

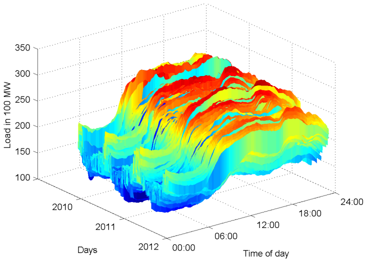
- trend only in upper levels



Annual expectiles for $\tau = 0.25, 0.5, 0.75$

Expectile demand models

- **Electricity demand**
  - Quarter-hourly
  - Jan.2010 - Dec.2012
  - distributional forecast
  - Schulz & Lopez-Cabrera (2016)



- **Water demand**
- **Gas demand**

"Principal Components" for expectiles

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

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

Outline

1. Motivation ✓
2. Quantiles and Expectiles
3. Algorithms for "PCA" in an asymmetric norm
4. Simulations
5. Application - fMRI brain data
6. Application - Chinese Temperature data

Quantiles and Expectiles

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

τ-quantile:

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

τ-expectile

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

where for $\alpha = 1, 2$

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

Quantiles and Expectiles

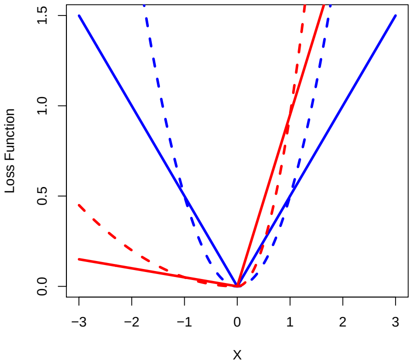


Figure 1: Loss functions for $\tau = 0.9$; $\tau = 0.5$; $\alpha = 1$ (solid); $\alpha = 2$ (dashed)

Quantiles vs. Expectiles

- Newey and Powel (1987) introduced expectiles:
  - simpler to compute
  - efficient estimators
  - for asym. cov. matrix for need to compute density
- expectiles sensitive to extreme values:
  - preferred in the calculation of risk measures
    Kuan et al. (2009): VaR vs. EVaR

Appendix- Expectile as Quantile
Appendix-Expected shortfall

PCA geometry

- PCA: minimize error vs. maximize variance

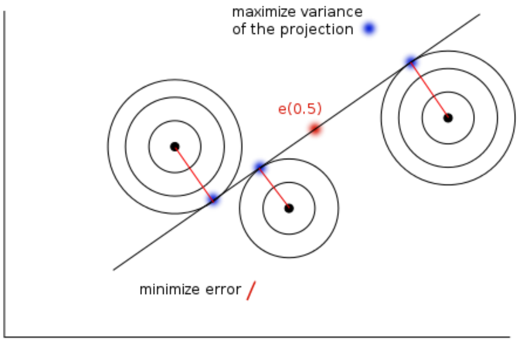


Figure 2: Best one dimensional approximation of two-dimensional variables

"PEC" geometry

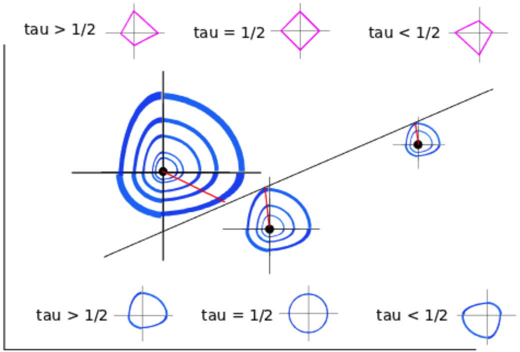


Figure 3: One dimensional approximation of two-dimensional variables in an asymmetric L_1 (magenta) and L_2 (blue) norm

"PEC" as error minimizers

Find best k-dimensional approximation Ψ_k^*:

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

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.

"PEC" 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)

$$\phi^*_\tau = \underset{\phi \in \mathbb{R}^p, \phi^\top \phi = 1}{\operatorname{argmax}} \mathrm{Var}_\tau(\phi^\top Y_i)$$

$$\hat{\phi}^*_\tau = \underset{\phi \in \mathbb{R}^p, \phi^\top \phi = 1}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} (\phi^\top Y_i - \mu_\tau)^2 w_i,$$

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

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

$n^+ = |\mathcal{I}_\tau^+|$ and $n^- = |\mathcal{I}_\tau^-|$, then the estimator of τ-expectile is:

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

ϕ_τ^* is the eigenvector for largest eigenvalue of C_τ where

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

PEC is constrained PC!

Theorem

Suppose the true weights are given and $\hat{e}_\tau$ and C_τ defined as before. Then ϕ^*_τ is the solution to the following optimization problem:

$$
\begin{aligned}
\text{maximize } & \phi^\top C_\tau \phi \\
\text{subject to } & \phi^\top Y_i > \phi^\top \hat{e}_\tau \text{ for } i \in \mathcal{I}^+_\tau \\
& \phi^\top \phi = 1.
\end{aligned} \tag{1}
$$

Algorithm for computing PEC

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

- Compute $\hat{e}_\tau$, 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

Algorithm for computing PEC

Theorem

The LAWS algorithm is well-defined, and is a gradient descent algorithm. Thus it converges to a critical point of the defined optimization problem.

Theorem

If $Y_1, \ldots, Y_n \in \mathbb{R}$ are n real numbers, LAWS finds their τ-expectile e_τ in $\mathcal{O}\{\log(n)\}$ iterations.

LAWS_expectile

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: $\phi_\tau^*(Y_n) \xrightarrow{P} \phi_\tau^*(Y)$.

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].

$$\begin{aligned}
\mu(t) &= 1 + t + \exp\{-(t-0.6)^2/0.05\} \\
f_1(t) &= \sqrt{2}\sin(2\pi t); \quad f_2(t) = \sqrt{2}\cos(2\pi t) \\
\alpha_{r,i} &\sim \mathrm{N}(0, \sigma_r^2),
\end{aligned}$$

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.

PEC_sim_setup

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

PEC_sim_setup

MSE against sample

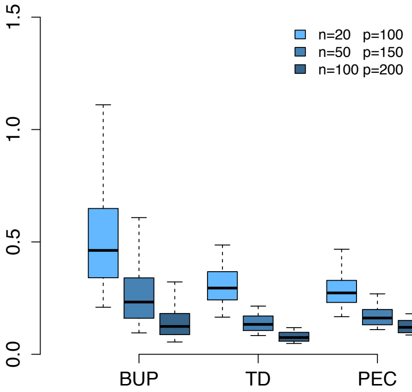


Figure 4: average MSE of BUP, TD and PEC by 500 simulations back

MSE against scenarios

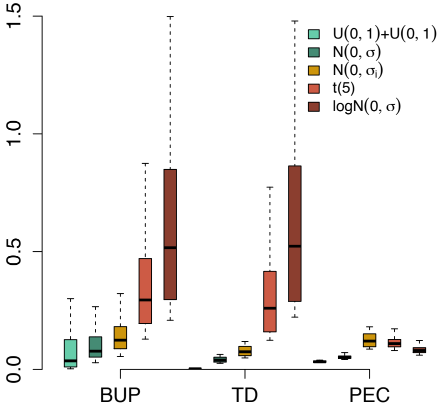


Figure 5: average MSE of BUP, TD and PEC by 500 simulations ▸ back

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

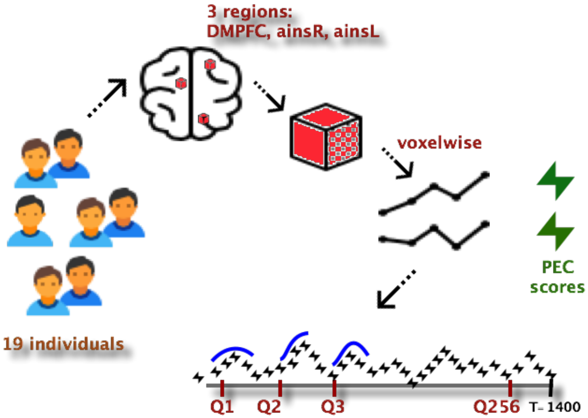
Table 2: Convergence rates (ratio of converged to unconverged cases by 30 iterations) of the algorithms by 500 simulation runs

fMRI data

- 19 volunteers
- 256 Risk investment task (stimulus)
- 1400 scans (every 2s)
- measure Blood Oxygenation Level Dependent Effect

- take data "voxel-wise", use all information

fMRI data



Free icons obtained from: icons8.com

fMRI data - Risk attitude

Following common Markowitz mean-variance model
Majer et al. (2014), Mohr and Nagel (2010)

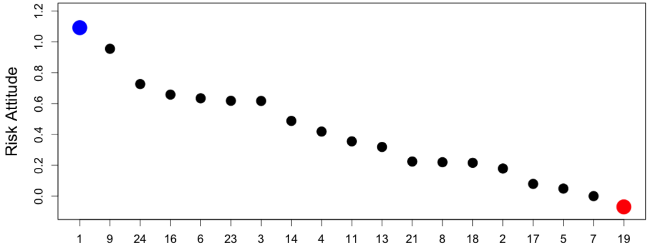


Figure 6: Risk attitude of 19 individuals

Application to fMRI data

$risk.att =$
$\beta_0 + \beta_1 \psi_{1,\tau}^{ainsL} + \beta_2 \psi_{1,\tau}^{ainsR} + \beta_3 \psi_{1,\tau}^{DMPCF} + \beta_4 \psi_{2,\tau}^{ainsL} + \beta_5 \psi_{2,\tau}^{ainsR} + \beta_6 \psi_{2,\tau}^{DMPCF}$

where $\psi_{k,\tau}^{\cdot}$ is the score of k-th PEC

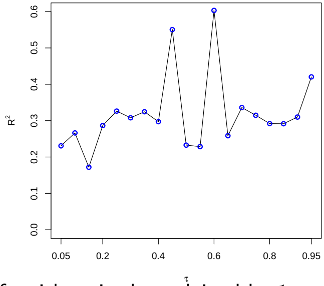


Figure 7: R^2 for risk attitude explained by 1st and 2nd PEC scores

PEC_fmri

Application to fMRI data

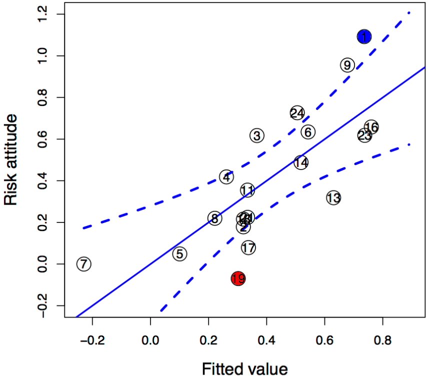


Figure 8: R^2 for risk attitude explained by PEC scores $\tau = 0.6$.

PEC_fmri

Application to Chinese Temperature

Daily average temperatures in 159 stations in China in period 1957-2009.

Chinese temperature data

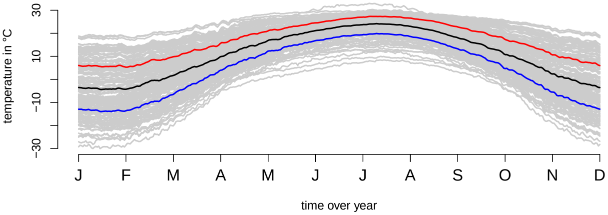


Figure 9: Figure 10: Averaged (over years) temperature curves (gray) and the estimated average expectiles by PEC for τ=0.1, 0.9

1st and 2nd PECs

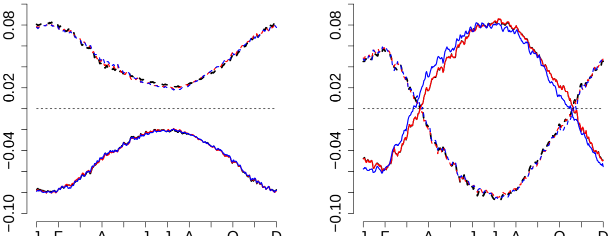


Figure 10: The estimated first PEC (left) and 2nd PEC (right) for $\tau = 0.1$ (dashed) and $\tau = 0.9$ (solid) computed with three proposed algorithms TopDown, BottomUp and PrincipileExpectile.

PEC_temperature

1st and 2nd PECs

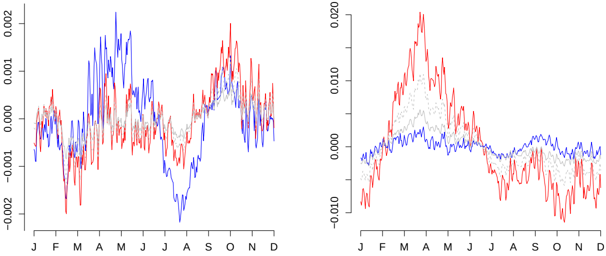


Figure 11: The differences of estimated PECs for $\tau = 0.1$ and $\tau = 0.9$ from estimated PEC for $\tau = 0.5$, computed with `PrincipileExpectile` algorithm. Differences for 1st component are shown in left, for 2nd component in right

Interpretation

- Indicate changes in distribution from lighter to heavier tails and vice-versa.
- Scores indicates the periodic change over years.
- Positive score on PC_1 – heavier tails in spring and winter.
- Positive score on PC_2 – heavier tails in summer (January-March).

北京 - Dimension reduction

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

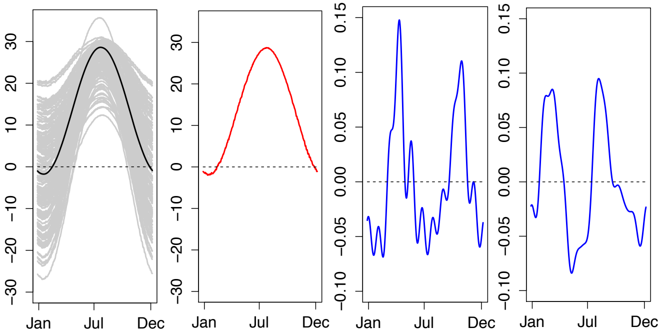


Figure 12: Approximation via PEC for the temperature expectile curve of Beijing for τ=0.95

PEC scores

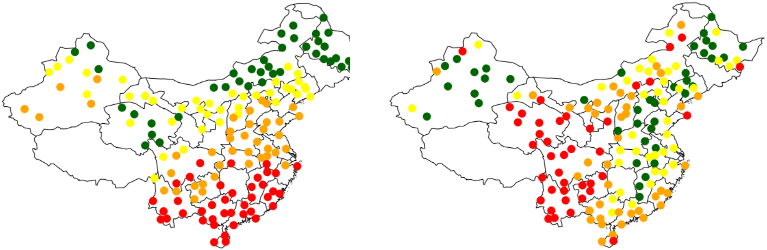


Figure 13: Scores on 1st PEC (left) and 2nd PEc (right) for τ=0.9

PEC_temperature

1st PEC scores

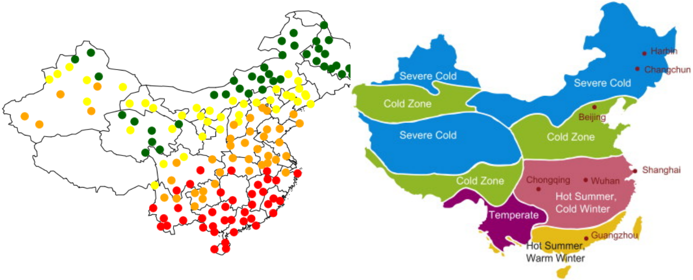


Figure 14: Scores on 1st PEC (left) and climate zones

PEC_temperature

2nd PEC scores

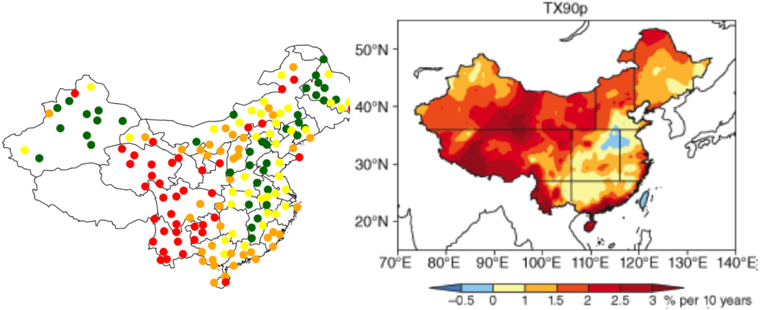


Figure 15: Scores on 2nd PEC (left) and index TX90p

PEC_temperature

Note: TX90p - warm days indicator, the core indicator by WMO.

Conclusion

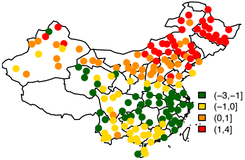
- Dimension reduction technique for tail event curves.
- Two ways to define PC for τ-expectiles: minimize error in the τ-norm (BUP and TD), and maximize the τ-variance.
- Maximize τ-variance (PEC) is a version of weighted PCA.

Conclusion

- PEC outperforms BUP and TD in simulations.
- PEC robust to 'fattails' and skewness of the data distribution.
- In practice the outputs of BUP, TD, and PEC do not differ much.
- fMRI: $\tau = 0.6$ provides the best explanation of risk attitude.
- Temperature: clarified seasonal and long-term component.

Principal components in an asymmetric norm

Ngoc Mai Tran
Petra Burdejova
Maria Osipenko
Wolfgang Karl Härdle



Ladislaus von Bortkiewicz Chair of Statistics
School of Business and Economics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

Literature

- N.M. Tran, P. Burdejova, M. Osipenko and W.K. Härdle
  Principal Component Analysis in an Asymmetric Norm
  Discussion Paper 2016-040, CRC 649: *Economic Risk*.

- P. Burdejova, W. K. Härdle , P. Kokoszka, Q. Xiong
  Change point and trend analysis of annual expestile curves of tropical storms
  Econometrics and Statistics, 2016.

- B. López-Cabrera, F. Schulz
  Forecasting Generalized Quantiles of Electricity Demand: A Functional Data Approach
  Journal of the American Statistical Association, 2016.

Literature

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

P. Majer, P. Mohr, H. R. Heekeren, and W. K. H‘d’rdle
Portfolio Decisions and Brain Reactions via the CEAD method
Psychometrika, 1987.

S. Schnabel
Expectile smoothing: new perspectives on asymmetric least squares.
PhD Thesis, Utrecht University 2011.

Literature

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

- C. M. Kuan, J. H. Yeh, Y. C. Hsu
  Assessing value at risk with CARE, the Conditional Autoregressive Expectile models
  Journal of Econometrics, 150 2009, p.261-270.

- P. N. Mohr, I. E. Nagel
  Variability in Brain Activity as an Individual Difference Measure in Neuroscience
  Journal of Neuroscience, 30(23) 2010, p.7755-7757.

Expectile-quantile correspondence

$$\tau(s) = \frac{s q_s(Y) - \int_{-\infty}^{q_s(Y)} y dF(y)}{\mathsf{E}(Y) - 2\int_{-\infty}^{q_s(Y)} y dF(y) - (1-2s) q_s(Y)} \qquad (2)$$

s-quantile corresponds to expectile with transformation $\tau(s)$.

Back to Expectiles

Expectile-quantile correspondence

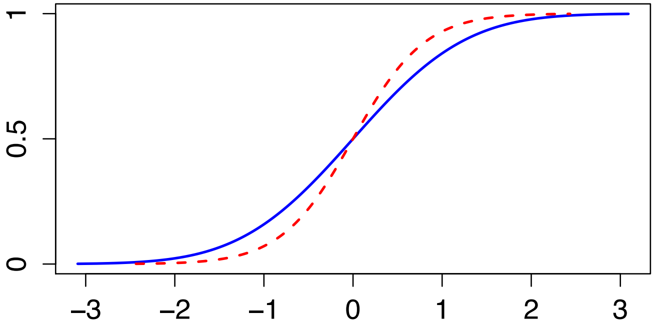


Figure 16: Quantiles (solid) and expectiles (dashed) of a normal N(0,1)

Back to Expectiles

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 to Expectiles

Skorokhod space $\mathbf{D}([0, 1])$

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. C.d.f is càtlàg

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

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

Then:

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

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

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

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

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

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

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