Factorizable Sparse Tail Event Curves with Expectiles

Shih-Kang Chao
Wolfgang K. Härdle
Chen Huang

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt–Universität zu Berlin
Department of Statistics, Purdue University
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
http://irtg1792.hu-berlin.de
http://www.stat.purdue.edu

Holding on the two ends...

FASTEC: FActorisable Sparse Tail Event Curves

- Common structure
  - High-dimensional time series with factors
  - Sparse penalization
- Individual variety
  - Tail behaviour
  - Spread analysis on factor loadings

Chao et al. (2015)

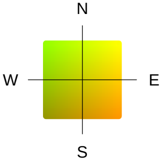
fMRI Application

- High-dimensional, high frequency & large data set
  - 19 volunteers, 256 investment decisions tasks
  - Around 100^3 voxels' data points, Blood Oxygenation Level Dependent (BOLD) effect every 2 sec
- Investment decisions and brain reactions
- Economics, Psychology and Statistics
- Spectral clustering identifies active zones

Majer et al. (2015)

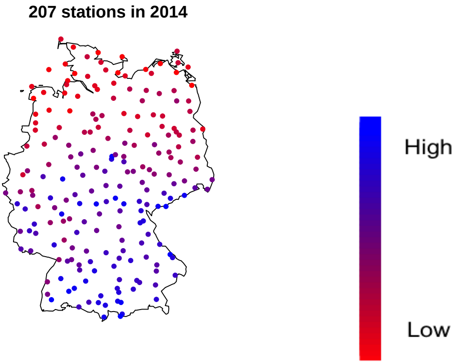
Chinese Temperature

- Daily data from 1957 to 2013
- 159 Chinese weather stations
- Temperature distribution and extreme weather forecasting
- Weather derivatives in financial industry



DWD Climate Data

- Daily wind speed data for German stations from 1964 to 2014



Aging and Growing over the World

- About 40 countries or areas, 1921-2011
- Mortality trend over ages
- Extremes and expectiles, tail events

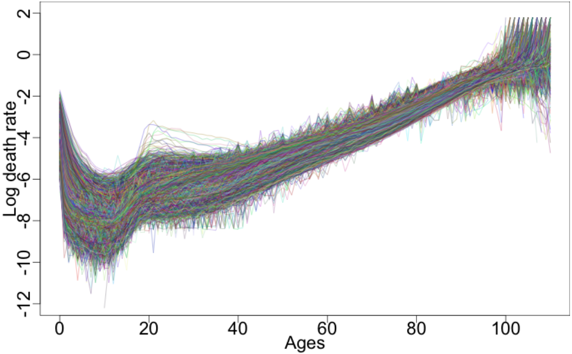


Figure 1: Log death rates over ages

Challenges

- Dimension reduction
- Multivariate tail event regression
- Fast Iterative Shrinkage Thresholding Algorithm?
- Oracle inequalities for the estimator
- How do Tail Event Curves vary in time?

Outline

1. Motivation ✓
2. FASTEC with Expectiles
3. fMRI data & risk perception
4. Empirical Results
5. Conclusions

Tail Event Curve

- Quantile
  - Ratio of areas
  - Local influence

$$\frac{\tau}{1-\tau} = \frac{\int_{-\infty}^{q_\tau} dF(y)}{\int_{q_\tau}^{\infty} dF(y)}$$

VaR and ES

- Expectile
  - Ratio of weighted averaged distances
  - Capture the tail moments, not robust

$$\frac{\tau}{1-\tau} = \frac{\int_{-\infty}^{e_\tau} |y - e_\tau|\, dF(y)}{\int_{e_\tau}^{\infty} |y - e_\tau|\, dF(y)}$$

History of Expectiles

Example

When $\tau = 0.5$, quantile = median, expectile = mean.

Quantile and Expectile

- Loss function
  $\rho_{\tau,\alpha}(u) = |\tau - \mathbf{I}\{u < 0\}|\,|u|^{\alpha}$, with $\alpha = 1, 2$, $\tau \in (0, 1]$
  - Quantile
    $q_\tau = \arg\min_\theta \mathsf{E}\,\rho_{\tau,1}(Y - \theta)$
  - Expectile
    $e_\tau = \arg\min_\theta \mathsf{E}\,\rho_{\tau,2}(Y - \theta)$

Note: the MLE of the location parameter of an ALD/AND correspond to the quantile/expectile regression estimator

Quantile and Expectile

- Loss function

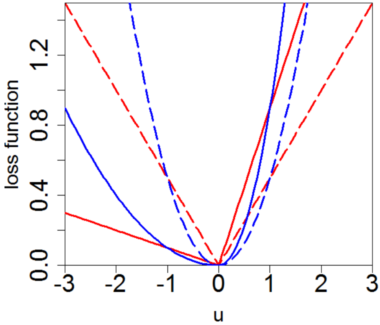


Figure 2: Expectile and quantile loss functions at $\tau = 0.5$ (dashed), $\tau = 0.9$ (solid).

LQRcheck

Model Specification

- $\{\boldsymbol{Y}_i\}_{i=1}^{n} \in \mathbb{R}^m$: multivariate curves to be jointly modelled
- $\{\boldsymbol{X}_i\}_{i=1}^{n} \in \mathbb{R}^p$: p increases with n, B-spline basis or other regression variables

Example

Chinese temperature: for each year, $m = 159$ (stations), $n = 365$ (days), $p = n^{0.6} \approx 34$.
fMRI: $m = 19$ (individuals) $\times 256$ (questions) $= 4864$, $n = 50$ (data points), $p = n^{0.8} \approx 23$.
Demographic data: for each year, $m = 38$ (countries), $n = 111$ (ages), $p = n^{0.8} \approx 43$.

Model Specification - ctd

- Conditional expectile function $e(\tau|\boldsymbol{X}_i)$ is approximated by linear factor model:

$$e(\tau|\boldsymbol{X}_i) = \sum_{k=1}^{r} \psi_k(\tau) f_k^\tau(\boldsymbol{X}_i), \tag{1}$$

  where $f_k^\tau(\boldsymbol{X}_i)$ is the kth factor, r is the number of factors, $\psi_k(\tau)$ are the factor loadings.

- Dimension reduced from p to r

Model Specification - ctd

- Factors are constructed by linear combination of $\boldsymbol{X}_i$:

$$f_k^{\tau}(\boldsymbol{X}_i) = \boldsymbol{\varphi}_{\boldsymbol{k}}^{\top}(\tau)\boldsymbol{X}_i \tag{2}$$

- Substituting (2) into (1):

$$e(\tau|\boldsymbol{X}_i) = \gamma^{\top}(\tau)\boldsymbol{X}_i \tag{3}$$

with $\boldsymbol{\gamma}(\tau) = \left(\sum_{k=1}^{r} \psi_k(\tau)\varphi_{k,1}(\tau), \dots, \sum_{k=1}^{r} \psi_k(\tau)\varphi_{k,p}(\tau)\right)^{\top}$, which is one column of the coefficient matrix $\mathbf{\Gamma}$.

Estimation

- Coefficient matrix $\mathbf{\Gamma}$:

$$\widehat{\mathbf{\Gamma}}_\lambda(\tau) = \arg \min_{\mathbf{\Gamma} \in \mathbb{R}^{p \times m}} \left\{ (mn)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho_\tau \left(Y_{ij} - \boldsymbol{X}_i^\top \mathbf{\Gamma}_{\cdot j} \right) + \lambda \|\mathbf{\Gamma}\|_* \right\} \tag{4}$$

  - $\mathbf{\Gamma}_{\cdot j}$ is the jth column of $\mathbf{\Gamma} \in \mathbb{R}^{p \times m}$
  - nuclear norm $\|\mathbf{\Gamma}\|_* = \sum_{j=1}^{\min(p,m)} \sigma_j (\mathbf{\Gamma})$, given the eigenvalues of $\mathbf{\Gamma}$: $\sigma_1 (\mathbf{\Gamma}) \geq \sigma_2 (\mathbf{\Gamma}) \geq \ldots \geq \sigma_{\min(p,m)} (\mathbf{\Gamma})$,
  - # of factors is # of nonzero eigenvalues of $\mathbf{\Gamma}$
  - Solved by fast iterative shrinkage thresholding algorithm
  - Identify factors and loadings ▸ Factorizable

Fast Iterative Shrinkage Thresholding Algorithm

▸ Iterative procedure

Beck and Teboulle (2009)

- Objective: $\min_{\boldsymbol{\Gamma}} \left\{ F(\boldsymbol{\Gamma}) \stackrel{\text{def}}{=} g(\boldsymbol{\Gamma}) + h(\boldsymbol{\Gamma}) \right\}$
- g: smooth convex function with Lipschitz continuous gradient

$$\|\nabla g(\boldsymbol{\Gamma}_1) - \nabla g(\boldsymbol{\Gamma}_2)\|_{\mathsf{F}} \leq L_{\nabla g} \|\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2\|_{\mathsf{F}}, \forall \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2$$

  where $L_{\nabla g}$ is the Lipschitz constant of ∇g
- h: continuous convex function, possibly nonsmooth
- $|F(\boldsymbol{\Gamma}_t) - F(\boldsymbol{\Gamma}^*)| \leq \frac{2L_{\nabla g}\|\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}^*\|_{\mathsf{F}}^2}{(t+1)^2}$

Loss Error Bound and Convergence Analysis

Theorem 1

- Lipschitz continuity of expectile loss gradient:
  $L_{\nabla g} = 2(mn)^{-1} \max(\tau, 1-\tau) \|\boldsymbol{X}\|_F^2$
- In the $t-$th step of the iteration

$$|F(\boldsymbol{\Gamma}_t) - F(\boldsymbol{\Gamma}^*)| \leq \frac{4(mn)^{-1} \max(\tau, 1-\tau) \|\boldsymbol{X}\|_F^2 \|\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}^*\|_F^2}{(t+1)^2} \tag{5}$$

- To achieve $|F(\boldsymbol{\Gamma}_t) - F(\boldsymbol{\Gamma}^*)| \leq \varepsilon$, $\forall \varepsilon > 0$, we need

$$t \geq \frac{2\sqrt{\max(\tau, 1-\tau)} \|\boldsymbol{X}\|_F \|\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}^*\|_F}{\sqrt{mn\varepsilon}} - 1 \tag{6}$$

- Convergence rate $\mathcal{O}(1/\sqrt{\varepsilon})$

Oracle Inequalities

- Upper bounds for $\|\widehat{\mathbf{\Gamma}}_\lambda - \mathbf{\Gamma}^*\|_F^2$ in finite sample
- $\mathbf{\Gamma}^*$ can be exactly sparse or not
- High-dimensional framework: $\operatorname{rank}(\mathbf{\Gamma}^*)$ and $p+m$ are both allowed to tend to infinity (but no quicker than n)
- $\{(\boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i=1}^n \in \mathbb{R}^{p+m}$ are identically distributed observations

Error Bounds for the Estimator

- Unified framework for high-dimensional $M-$Estimators with decomposable regularizers by Negahban et al. (2012)
- Conditions need to be verified:
  - Restricted strong convexity holds for expectile loss function RSC
  - Nuclear norm is decomposable with respect to appropriately chosen subspaces Decomposable

Error Bounds for the Estimator

Theorem 2

Suppose $\{\boldsymbol{X}_i\}_{i=1}^{n} \in \mathbb{R}^p$ are i.i.d. samples from $\mathrm{N}(\mathbf{0}, \mathbf{\Sigma})$, for $n \geq 2\min(m, p)$, any optimal solution $\widehat{\mathbf{\Gamma}}_\lambda$ with a strictly positive tuning parameter $\lambda \geq 2\|\nabla g(\mathbf{\Gamma}^*)\|$ satisfies the bound

$$\|\widehat{\mathbf{\Gamma}}_\lambda - \mathbf{\Gamma}^*\|_{\mathsf{F}}^2 \leq \frac{9^3 m^2 \lambda^2}{\{\min(\tau, 1-\tau)\,\sigma_{\min}(\mathbf{\Sigma})\}^2} \Psi^2\left(\overline{\mathcal{M}}\right) + \frac{36 m\lambda}{\min(\tau, 1-\tau)\,\sigma_{\min}(\mathbf{\Sigma})} \|\mathbf{\Gamma}^*_{\mathcal{M}^\perp}\|_*, \tag{7}$$

with probability greater than $1 - 4\exp(-n/2)$.

Error Bounds for the Estimator

- $\Psi\left(\overline{\mathcal{M}}\right) \stackrel{\text{def}}{=} \sup_{\mathbf{Z}\in\overline{\mathcal{M}}\setminus\{\mathbf{0}\}} \frac{\|\mathbf{Z}\|_*}{\|\mathbf{Z}\|_F}$ is the subspace compatibility constant, which is $\sqrt{\text{rank}\left(\boldsymbol{\Gamma}^*\right)}$ when $\boldsymbol{\Gamma}^*$ is exactly sparse
- $\boldsymbol{\Gamma}^*_{\mathcal{M}^\perp} \stackrel{\text{def}}{=} \arg\min_{\mathbf{Z}\in\mathcal{M}^\perp} \|\mathbf{Z}-\boldsymbol{\Gamma}^*\|_F$
- Best choice of λ ▸ Tuning Parameter

Error Bounds for the Estimator

Corollary

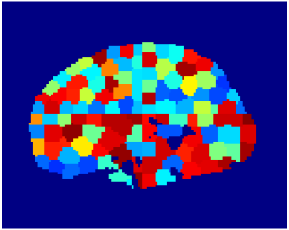
Under the assumptions on sample setting, selecting $\lambda = 2m^{-1} S \max(\tau, 1-\tau) \sqrt{K_u^2 \|\mathbf{\Sigma}\|} \sqrt{\frac{p+m}{n}}$, for $n \geq 2\min(m, p)$, any optimal solution $\widehat{\mathbf{\Gamma}}_\lambda$ satisfies the bound

$$\|\widehat{\mathbf{\Gamma}}_\lambda - \mathbf{\Gamma}^*\|_F^2 \leq \frac{9^3 \cdot \{2S \max(\tau, 1-\tau) K_u\}^2 \|\mathbf{\Sigma}\| (p+m)}{n\{\min(\tau, 1-\tau)\sigma_{\min}(\mathbf{\Sigma})\}^2} \Psi^2(\overline{\mathcal{M}}) + \frac{72S \max(\tau, 1-\tau) \sqrt{K_u^2 \|\mathbf{\Sigma}\|} \sqrt{p+m}}{\sqrt{n} \min(\tau, 1-\tau) \sigma_{\min}(\mathbf{\Sigma})} \|\mathbf{\Gamma}^*_{\mathcal{M}^\perp}\|_*, \tag{8}$$

with probability greater than $1 - 3 \times 8^{-(p+m)} - 4\exp(-n/2)$.

Investment Decisions and Brain Reactions

- How does an individual perceive risk?
- Is risk attitude reflected in brain activity?



Investment Decision Experiment

- Survey by Department of Education and Psychology, FU Berlin
- 19 healthy volunteers payoff

- Investment Decision (ID) task ($\times 256$)
  safe vs. **random** (μ, σ) return
- fMRI images: 2 sec$\times 1400 \approx 48$ min

Investment Decision

Choose between:

A) **Safe**, fixed return 5%

B) **Random**, investment return (3 types)

- ▸ Single Investment
- Portfolio of 2 (perfectly) ▸ correlated investments
- Portfolio of 2 ▸ uncorrelated investments

- Each type of portfolio $\times 64$, single $\times 128$
- Display and decision time: 7 sec

ID Experiment

Figure 3: Decide between **A)** 5% return and displayed **B)** portfolio/investment

fMRI Dynamics

Hemodynamic response (1 voxel) HRF

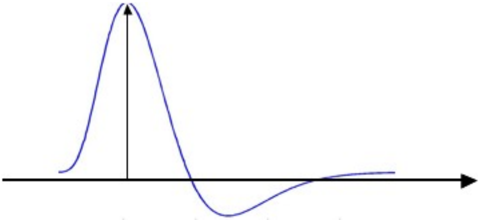


Figure 4: Hemodynamic response of a stimulus signal

Risk Attitude Parameter

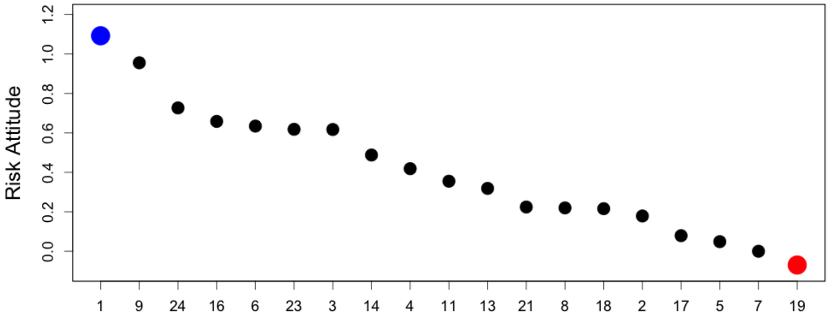


Figure 5: Estimated risk attitude for 19 subjects

▸ Risk attitude parameter

Importance of Tails

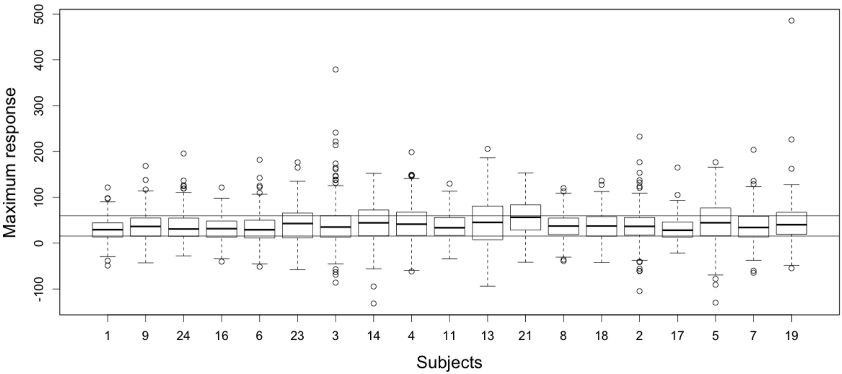


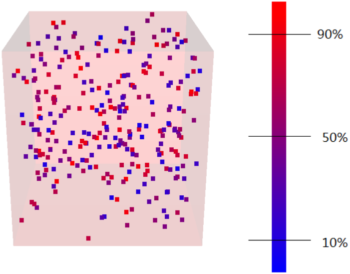
Figure 6: Boxplot of maximum responses over questions for 19 subjects (ordered by risk attitude parameters)

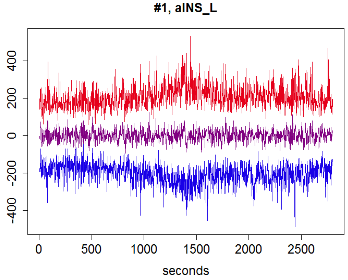
fMRI Data

- Three ID-related active clusters: aINS_Left and aINS_Right, DMPFC. Majer et al. (2015)

▸ aINS ▸ DMPFC

- At each t, take different quantile levels (0.1, 0.5, 0.9) among all voxels in each cluster





Data Smoothing

- 19 individuals, 256 questions, $19 \times 256 = 4864$ curves
- Use 4 scans (6 seconds) after each stimulus
- Linear interpolation, take 50 points from the fitted curve

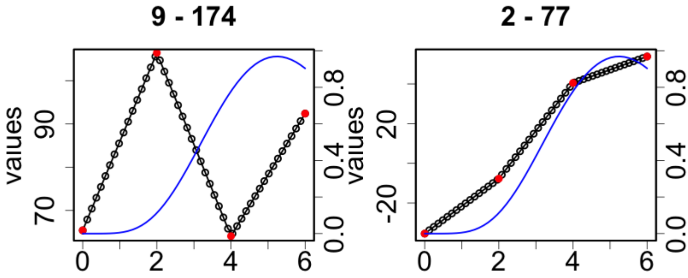


Figure 7: Two examples in aINS_Left cluster at 50% quantile level, with Hemodynamic response.

Factor Analysis

- $n = 50$ observations, $m = 4864$ curves
- $\boldsymbol{X}_i$: B-spline basis (cubic splines) with $p = n^{0.8} \approx 23$, $t = i/n$, $i = 1, \ldots, n$

▸ Cubic spline basis

	aINS_L	aINS_R	DMPFC
1st factor	0.624	0.631	0.612
2nd factor	0.791	0.793	0.779
3rd factor	0.907	0.913	0.898
4th factor	1.000	1.000	1.000

Table 1: Proportion of variance explained by the first four factors under $\tau = 50\%$

Factor Curves

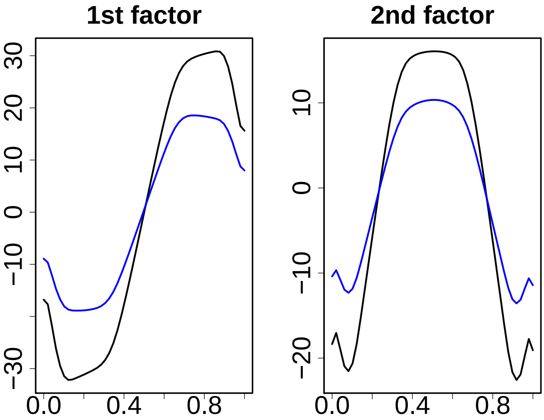


Figure 8: The first 2 factors under $\tau = 99\%$ and $\tau = 1\%$ respectively (1% quantile level in aINS_Left cluster)

Risk attitude - Stimulus Response

- Standard deviation of the factor loadings

$$\beta_i = \alpha_0 + \alpha_1 \cdot \text{sd}\,(\psi_1)^{\tau}_{i,ainsL} + \alpha_2 \cdot \text{sd}\,(\psi_1)^{\tau}_{i,ainsR} + \alpha_3 \cdot \text{sd}\,(\psi_1)^{\tau}_{i,DMPFC} + \varepsilon_i \tag{9}$$

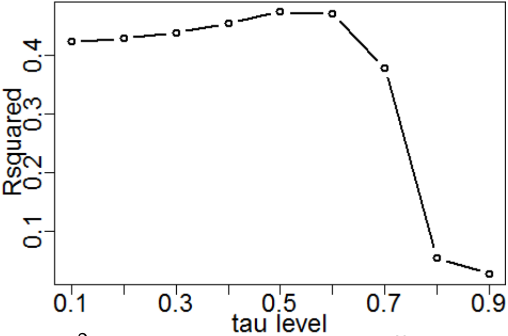


Figure 9: R^2 in the regressions under different τ levels

Risk attitude - Stimulus Response

$$\beta_i = \alpha_0 + \alpha_1 \cdot \mathrm{sd}\,(\psi_1)^{0.1}_{i,ainsL} + \alpha_2 \cdot \mathrm{sd}\,(\psi_1)^{0.1}_{i,ainsR} + \alpha_3 \cdot \mathrm{sd}\,(\psi_1)^{0.1}_{i,DMPFC} + \varepsilon_i \qquad (10)$$

	Estimate	SE	tStat	pValue
α_0	0.961	0.201	4.777	$0.245 \cdot 10^{-3}$
α_1	−34.414	12.250	−2.809	0.013
α_2	−37.571	15.623	−2.405	0.029
α_3	30.984	14.152	2.189	0.044

Table 2: Coefficients estimation results, $R^2 = 0.422$, adj.$R^2 = 0.307$.

Risk attitude - Stimulus Response

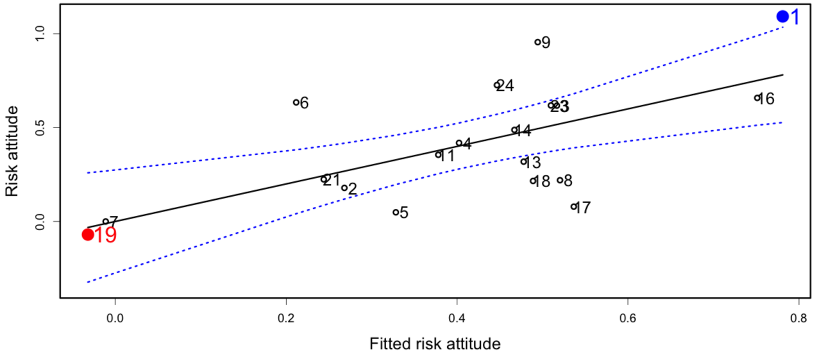


Figure 10: Fitted risk attitude by model given in (10) with $\tau = 0.1$.

Risk attitude - Stimulus Response

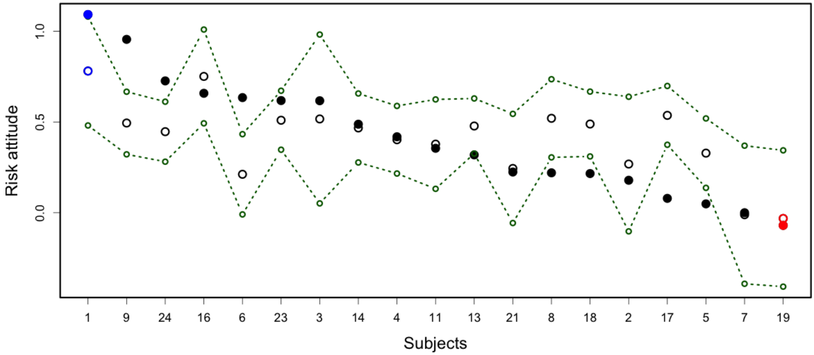


Figure 11: Fitted risk attitude (hollow points) by model given in (10) with $\tau = 0.1$.

Factor Curves

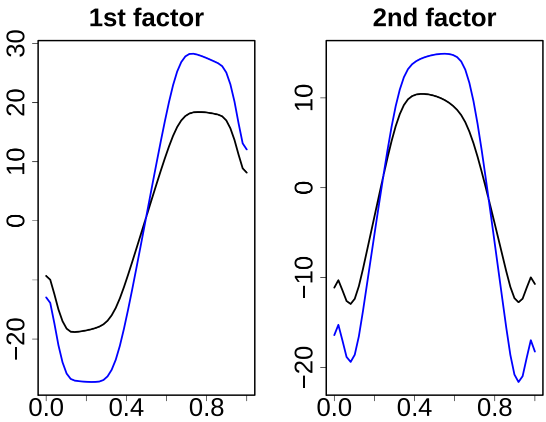


Figure 12: The first 2 factors under $\tau = 99\%$ and $\tau = 1\%$ respectively (99% quantile level in aINS_Left cluster)

Risk attitude - Stimulus Response

- Mean of the factor loadings

$$\beta_i = \alpha_0 + \alpha_1 \cdot (\bar{\psi}_1)^{\tau}_{i,ainsL} + \alpha_2 \cdot (\bar{\psi}_1)^{\tau}_{i,ainsR} + \alpha_3 \cdot (\bar{\psi}_1)^{\tau}_{i,DMPFC} + \varepsilon_i \tag{11}$$

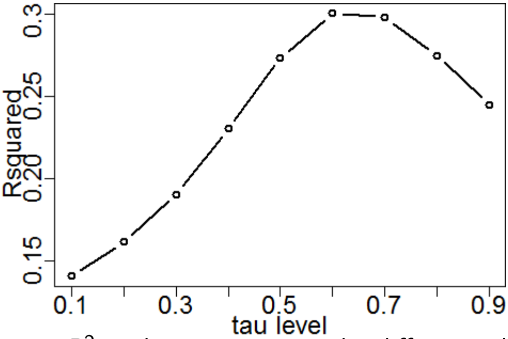


Figure 13: R^2 in the regressions under different τ levels

Risk attitude - Stimulus Response

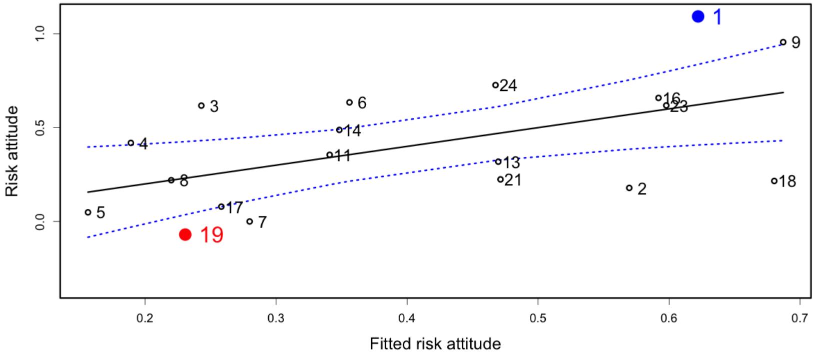


Figure 14: Fitted risk attitude by model given in (11) with $\tau = 0.7$.

Risk attitude - Stimulus Response

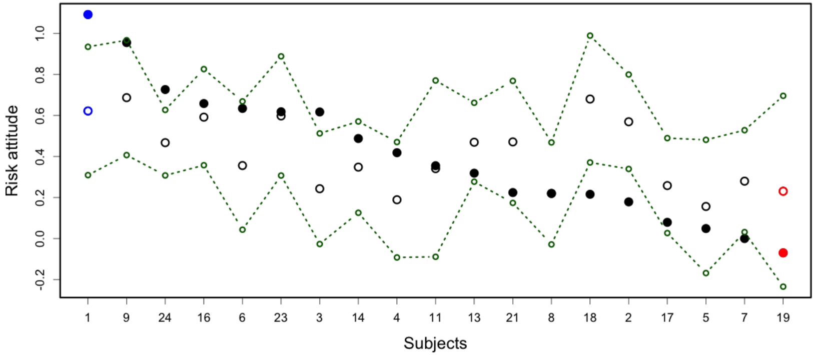


Figure 15: Fitted risk attitude (hollow points) by model given in (11) with $\tau = 0.7$.

Factor Curves

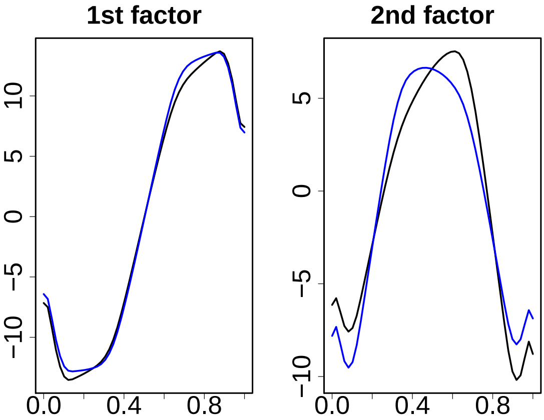


Figure 16: The first 2 factors under $\tau = 99\%$ and $\tau = 1\%$ respectively (50% quantile level in aINS_Left cluster)

Factor Loadings

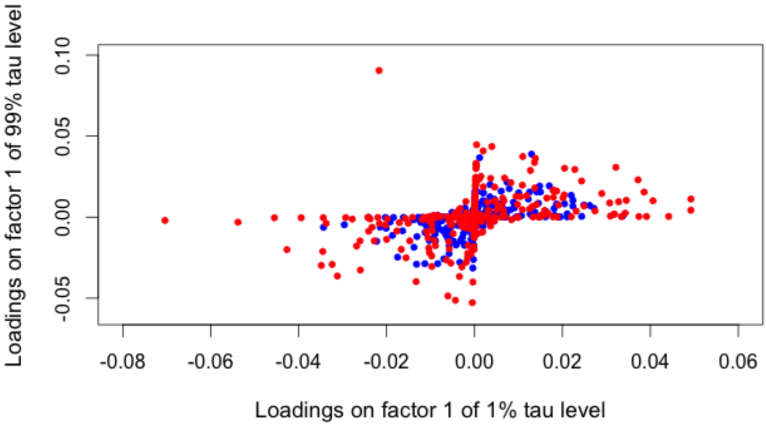


Figure 17: The first factor loadings for #1 and #19 individuals

Factor Loadings

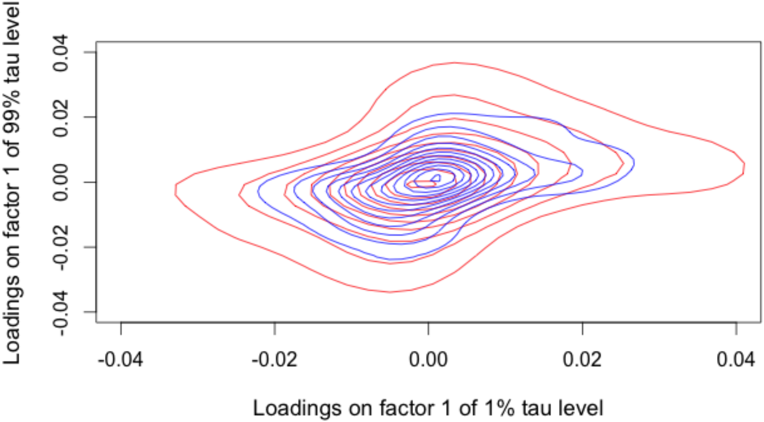


Figure 18: Iso-contour lines of ψ_1 for #1 and #19 individuals

Risk attitude - Stimulus Response

- Dispersion of the factor loadings under two τ levels

$$\text{dis}\left(\psi_1\right)_i^{\tau_1,\tau_2} = \frac{1}{256}\sum_{q=1}^{256}\sqrt{\left\{\psi_1\left(\tau_1\right)_{i,q} - \bar{\psi}_1\left(\tau_1\right)_i\right\}^2 + \left\{\psi_1\left(\tau_2\right)_{i,q} - \bar{\psi}_1\left(\tau_2\right)_i\right\}^2}$$

$$\beta_i = \alpha_0 + \alpha_1 \cdot \log \text{dis}\left(\psi_1\right)_{i,ainsL}^{\tau_1,\tau_2} + \alpha_2 \cdot \log \text{dis}\left(\psi_1\right)_{i,ainsR}^{\tau_1,\tau_2} + \alpha_3 \cdot \log \text{dis}\left(\psi_1\right)_{i,DMPFC}^{\tau_1,\tau_2} + \varepsilon_i \quad (12)$$

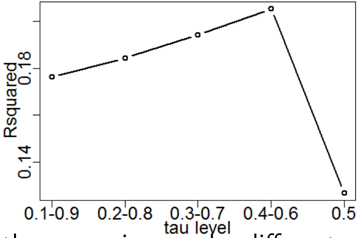


Figure 19: R^2 in the regressions under different τ levels (by pairs)

Risk attitude - Stimulus Response

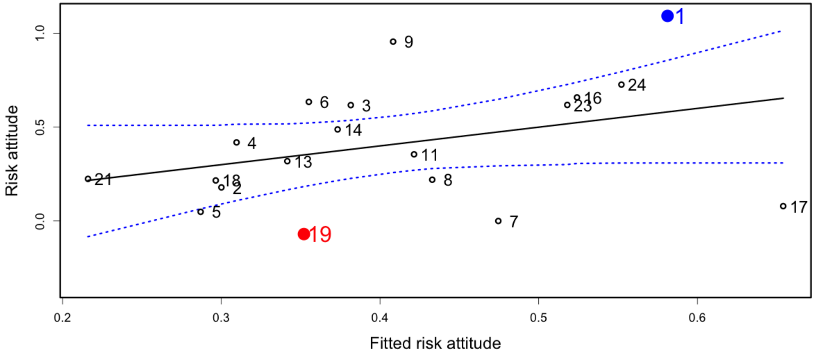


Figure 20: Fitted risk attitude by model given in (12) with $\tau_1 = 0.1, \tau_2 = 0.9$.

Risk attitude - Stimulus Response

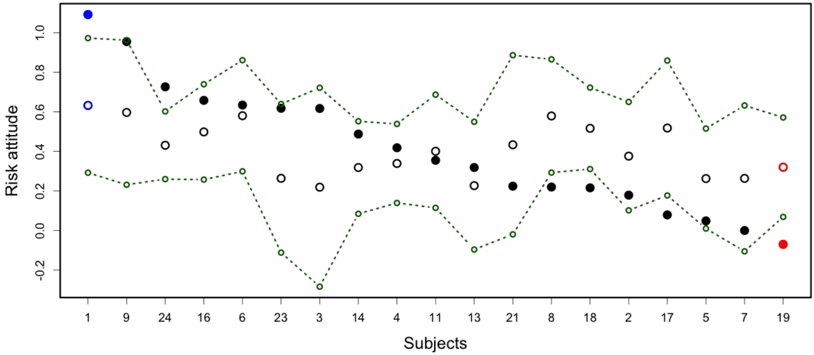


Figure 21: Fitted risk attitude (hollow points) by model given in (12) with $\tau_1 = 0.1, \tau_2 = 0.9$.

Factor Loadings

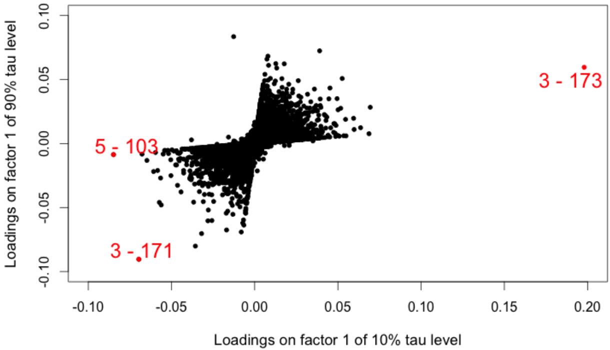


Figure 22: The first factor loadings for all curves $j = 1, \ldots, 4864$, where "3-171" denotes #3 individual's #171 question and so on.

Temperature Data

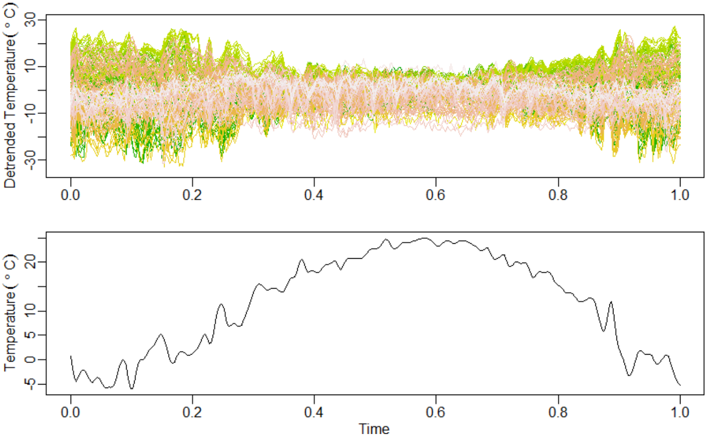


Figure 23: Top figure: detrended temperature series; bottom figure: trend

FASTECChinaTemper2008

Temperature Data - Factors

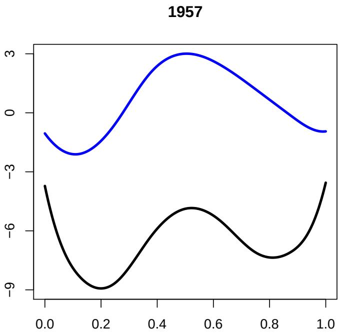
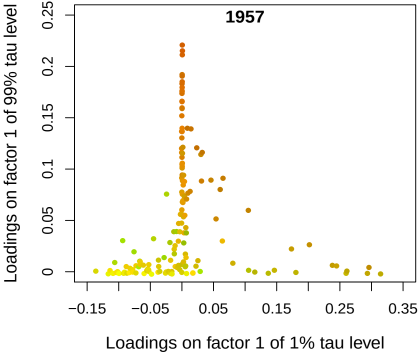


Figure 24: The first factor under 1% and 99% tail levels.

FASTEC_with_Expectiles

Temperature Data - Factor Loadings



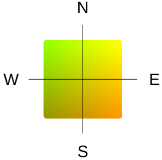


Figure 25: The first factor loadings for each station.

FASTEC_with_Expectiles

Temperature Data - Chinese Map

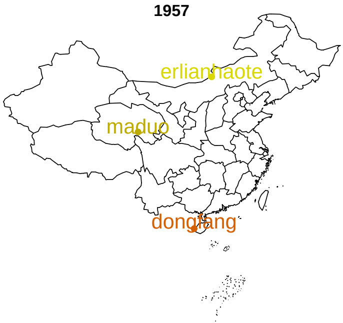


Figure 26: Chinese map marked with three selected weather stations.

FASTEC_with_Expectiles

Wind Data - Factors

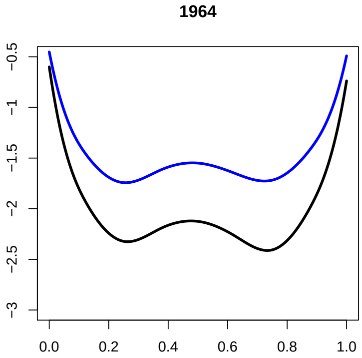


Figure 27: The first factor under 1% and 99% tail levels.

Wind Data - Factor Loadings

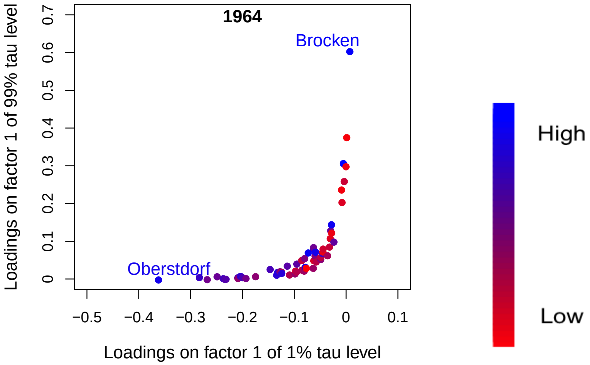


Figure 28: The first factor loadings for each station.

F

Wind Data - German Map

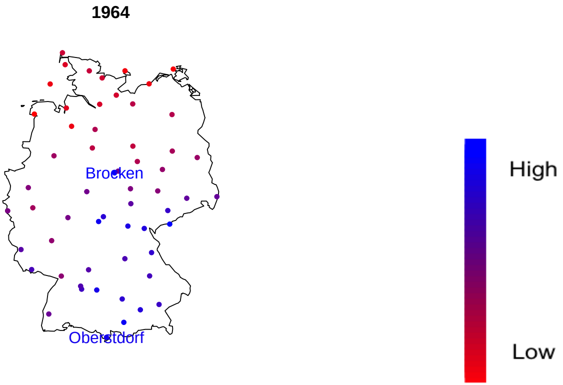


Figure 29: German map marked with two most extreme weather stations.

Mortality Data - Curves

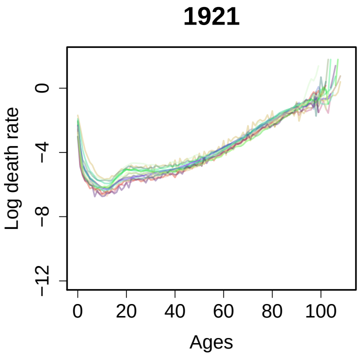


Figure 30: Log death rate curves from 1921 to 2011.

- In each year, a bundle of m curves over ages $0, 1, \ldots, 110$
- Estimate conditional expectile curves applying functional data analysis

Mortality Data - Factors

Figure 31: The first factor under 70% tail level.

- The common trend concerning quinquagenarian group
- Use factor loadings to detect the outliers
  - Good ones: Japan, Switzerland
  - Bad ones: Latvia, Russia

Mortality Data - Factor Loadings

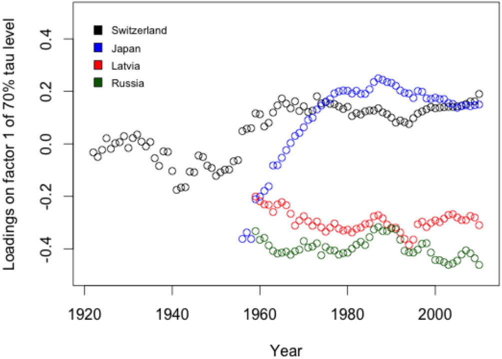


Figure 32: The first factor loadings for four representative countries in all years under 70% tail level: Switzerland, Japan, Latvia, Russia

Conclusions

- Principal factors capture the common patterns among curves
- TEC study discovers the extreme behaviors
- Consistency and convergence rate of the estimator are demonstrated by theorems
- Risk attitude implied by individual's choices and his brain reactions can be linked by statistical model

Factorizable Sparse Tail Event Curves with Expectiles

Shih-Kang Chao
Wolfgang K. Härdle
Chen Huang

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics and Economics
Humboldt–Universität zu Berlin
Department of Statistics, Purdue University
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
http://irtg1792.hu-berlin.de
http://www.stat.purdue.edu

References

Beck, A. and Teboulle, M. (2009)
A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems
SIAM Journal on Imaging Sciences, 2(1), 183-202

Breckling, J. and Chambers, R. (1988)
M-Quantiles
Biometrica, 75(4), 761-771

References

Cai, J.-F., Candès, E.J. and Shen, Z. (2010)
A Singular Value Thresholding Algorithm for Matrix Completion
SIAM Journal on Optimization, 20(4), 1956-1982

Chao, S.-K., Härdle, W.K. and Yuan, M. (2015)
Factorisable Sparse Tail Event Curves
SFB 649 Discussion Paper 2015-034, 2015

References

Majer, P., Mohr, P., Heekeren, H. and Härdle, W.K. (2015)
Portfolio Decisions and Brain Reactions via the CEAD method
Psychometrika, doi:10.1007/s11336-015-9441-5

Negahban, S.N., Ravikumar, P., Wainwright, M.J. and Yu, B. (2012)
A Unified Framework for High-Dimensional Analysis of M-Estimators with Decomposable Regularizers
Statistical Science, 27(4), 538-557

Value at Risk and Expected Shortfall

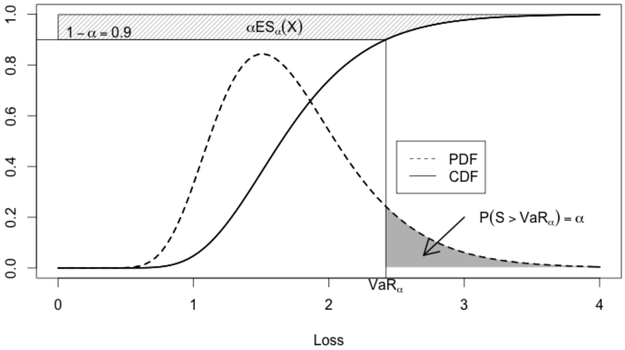


Figure 33: The relationship between VaR and ES in one plot. MSMVaRandES

Tail Event Curve

History of Expectiles

- Gravile - A. Goldberger
  - motivated by the interpretation of expectation as a center of gravity
- Projectile - G. Chamberlain
  - motivated by the fact that it solves a least squares problem
- Other alternative terminologies: Heftile, Loadile

▸ Tail Event Curve

Iterative Algorithm

- Initialize: $\boldsymbol{\Gamma}_0 = 0$, $\boldsymbol{\Sigma}_1 = 0$, step size $\delta_1 = 1$
- For $t = 1, 2, \ldots, T$
  - $\boldsymbol{\Gamma}_t = \arg\min_{\boldsymbol{\Gamma}} \left\{ \frac{g(\boldsymbol{\Gamma})}{L_{\nabla g}} + \frac{1}{2} \left\| \boldsymbol{\Gamma} - \left\{ \boldsymbol{\Omega}_t - \frac{1}{L_{\nabla g}} \nabla g\left(\boldsymbol{\Omega}_t\right) \right\} \right\|^2 \right\}$
  - when penalizing nuclear norm, $\boldsymbol{\Gamma}_t = \mathbf{P}\left(\mathbf{R} - \frac{\lambda}{L_{\nabla g}} \mathbf{I}_{p\times m}\right)_+ \mathbf{Q}^\top$, see Cai et al. (2010), where $\boldsymbol{\Omega}_t - \frac{1}{L_{\nabla g}} \nabla g\left(\boldsymbol{\Omega}_t\right) = \mathbf{PRQ}^\top$, by SVD
  - $\delta_{t+1} = \frac{1+\sqrt{1+4\delta_t^2}}{2}$
  - $\boldsymbol{\Omega}_{t+1} = \boldsymbol{\Gamma}_t + \frac{\delta_{t-1}}{\delta_{t+1}} \left(\boldsymbol{\Gamma}_t - \boldsymbol{\Gamma}_{t-1}\right)$
- $\widehat{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}_T$

▸ Model Estimation

Factorize $\widehat{\Gamma}_\lambda(\tau)$

- the number of nonzero singular values is the number of sparse factors: r
- dimension reduced from p to r
- singular value decomposition: $\widehat{\Gamma}_\lambda(\tau) = \mathbf{SVD}^\top$
- $f_k^\tau(\boldsymbol{X}_i) = \varphi_k^\top(\tau) = \sigma_k \mathbf{S}_{\cdot k}^\top \boldsymbol{X}_i$, $\psi = \mathbf{D}_{\cdot j}$, which is orthonormal

Model Estimation

RSC of Expectile Loss

- RSC holds for $g\left(\boldsymbol{\Gamma}\right)=(mn)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\rho_\tau\left(Y_{ij}-\boldsymbol{X}_i^\top\boldsymbol{\Gamma}_{\cdot j}\right)$ with curvature $\kappa>0$ and tolerance function $\xi(\cdot)$ if

$$g\left(\boldsymbol{\Gamma}^*+\boldsymbol{\Delta}\right)-g\left(\boldsymbol{\Gamma}^*\right)-\langle\langle\nabla g\left(\boldsymbol{\Gamma}^*\right),\boldsymbol{\Delta}\rangle\rangle\geq\kappa\|\boldsymbol{\Delta}\|_F^2-\xi^2\left(\boldsymbol{\Gamma}^*\right),\forall\boldsymbol{\Delta}\in\mathbb{C} \tag{13}$$

- $\rho_\tau\left(u\right)=\left|\tau-\mathbf{I}\left\{u<0\right\}\right|u^2$ satisfies

$$\rho_\tau\left(u+\delta\right)-\rho_\tau\left(u\right)-\rho_\tau'\left(u\right)\delta\geq\min\left(\tau,1-\tau\right)\delta^2 \tag{14}$$

RSC of Expectile Loss - ctd

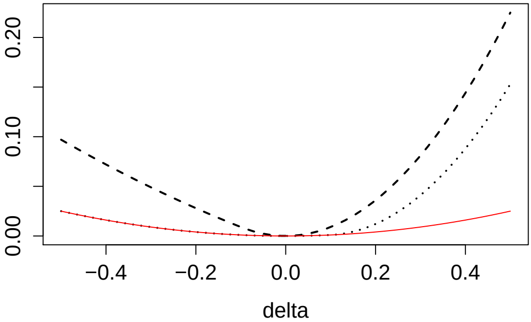


Figure 34: An example when $\tau = 0.9$, $\delta \in [-0.5, 0.5]$, where the two dash lines are the LHS of (14) w.r.t. δ (for $u = \pm 0.1$ respectively), and the red line is the lower bound

RSC of Expectile Loss - ctd

- $g\left(\boldsymbol{\Gamma}\right)$ is RSC with $\kappa = (m)^{-1} \min\left(\tau, 1-\tau\right) \sigma_{\min}\left(\frac{\boldsymbol{X}^\top \boldsymbol{X}}{n}\right)$ and $\xi(\cdot) = 0$
- if $\{\boldsymbol{X}_i\}_{i=1}^{n} \in \mathbb{R}^p$ are i.i.d. samples from $\mathrm{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$, for $n \geq 2\min\left(m, p\right)$, $\kappa = \frac{1}{9}(m)^{-1} \min\left(\tau, 1-\tau\right) \sigma_{\min}\left(\boldsymbol{\Sigma}\right)$ with probability greater than $1 - 4\exp\left(-n/2\right)$

Return

Decomposable Regularizers

- For $\mathcal{M} \subseteq \overline{\mathcal{M}}$ of $\mathbb{R}^{p\times m}$, a norm-based regularizer $\mathcal{R}$ is decomposable with respect to $\left(\mathcal{M}, \overline{\mathcal{M}}^{\perp}\right)$, if

$$\mathcal{R}(\mathbf{\Gamma}+\mathbf{\Delta}) = \mathcal{R}(\mathbf{\Gamma}) + \mathcal{R}(\mathbf{\Delta}), \forall \mathbf{\Gamma} \in \mathcal{M}, \mathbf{\Delta} \in \overline{\mathcal{M}}^{\perp} \tag{15}$$

- Nuclear norm is decomposable with respect to

$$\mathcal{M}(U,V) = \left\{\mathbf{\Gamma} \in \mathbb{R}^{p\times m} | \operatorname{row}(\mathbf{\Gamma}) \subseteq U, \operatorname{col}(\mathbf{\Gamma}) \subseteq V\right\}$$
$$\overline{\mathcal{M}}^{\perp}(U,V) = \left\{\mathbf{\Gamma} \in \mathbb{R}^{p\times m} | \operatorname{row}(\mathbf{\Gamma}) \subseteq U^{\perp}, \operatorname{col}(\mathbf{\Gamma}) \subseteq V^{\perp}\right\}$$

Return

More Assumptions

- $\{(\boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i=1}^n \in \mathbb{R}^{p+m}$ are i.i.d., $\{\boldsymbol{X}_i\}_{i=1}^n \in \mathbb{R}^p \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$
- Conditional on $\boldsymbol{X}_i$, $u_{ij} = \left\{ Y_{ij} - \boldsymbol{X}_i^\top \boldsymbol{\Gamma}_{\cdot j} \right\}_{j=1}^m$ are cross-sectional independent over j
- $u_{i1}, \ldots, u_{im}$ are sub-gaussian: $\exists C > 0$ such that $\mathrm{P}(|u_{ij}| > s) \leq \exp\left(1 - s^2/C^2\right)$, $j \in \{1, \ldots, m\}$
- $K_u \stackrel{\text{def}}{=} \max_{1 \leq j \leq m} \|u_{ij}\|_{\psi_2} = \max_{1 \leq j \leq m} \sup_{p \geq 1} p^{-1/2} \left(\mathrm{E}\,|u_{ij}|^p\right)^{1/p}$

Best choice of λ

Under the assumptions on sample setting,

$$\mathsf{P}\left(\|\nabla g\left(\boldsymbol{\Gamma}\right)\| \leq m^{-1} S \max\left(\tau, 1-\tau\right) \sqrt{K_u^2 \|\boldsymbol{\Sigma}\|} \sqrt{\frac{p+m}{n}} \right)$$
$$\geq 1 - 3 \times 8^{-(p+m)} - 4\exp\left(-n/2\right),$$

where S is an absolute constant.

Corallary

Experiment

- Incentive to be **rational**
  - Draw 1 ID task and multiply subject's choice by 100 EUR
    $9\% \times 100 = 9$ EUR
- Gaussian returns:
  - $\mu = 5\%, 7\%, 9\%, 11\%$
  - $\sigma = 2\%, 4\%, 6\%, 8\%$

ID Experiment

Single Investment

fMRI Experiment

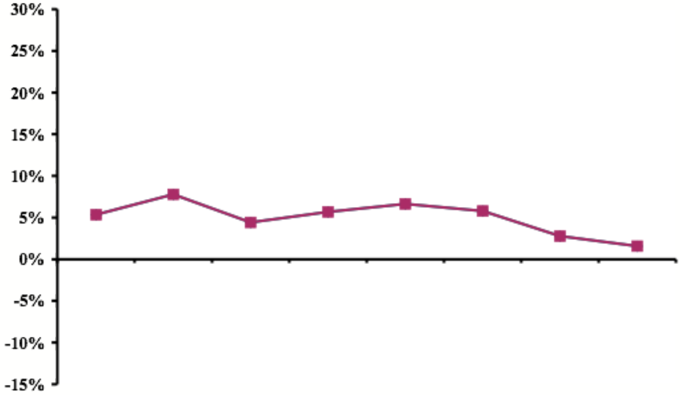


Figure 35: An example of return stream from single investment displayed to the subject during the experiment for 7 sec.; returns $r_i \sim N(\mu, \sigma^2)$, here $\mu = 5\%, \sigma = 2\%$

Correlated Portfolio

fMRI Experiment

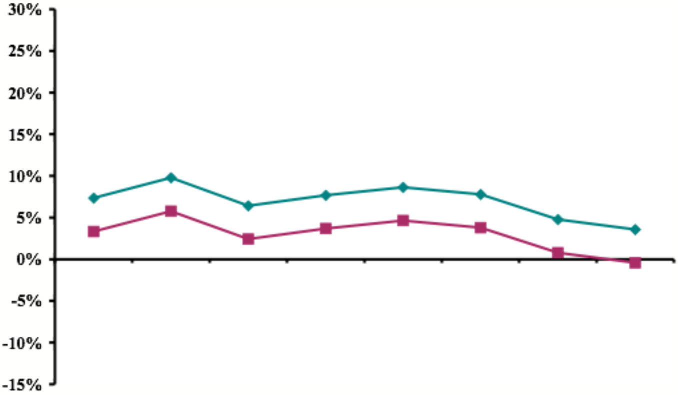


Figure 36: An example of return streams from correlated portfolio displayed to the subject during the experiment for 7 sec.; returns $r_i \sim \mathrm{N}(\mu, \sigma^2)$, here $\mu_1 = 5\%$,$\mu_2 = 9\%$ and $\sigma = 2\%$

Uncorrelated Portfolio

▸ fMRI Experiment

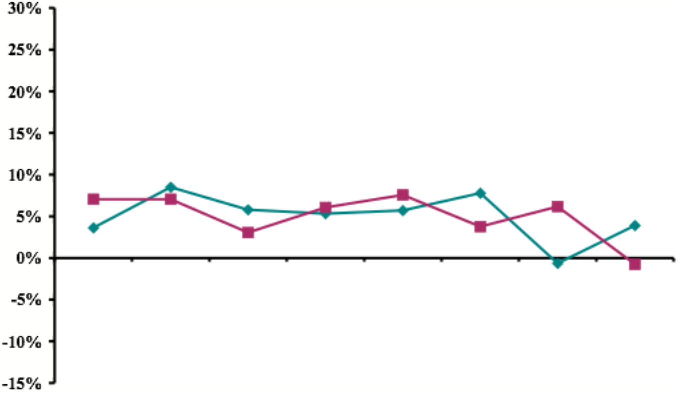


Figure 37: An example of return streams from uncorrelated portfolio displayed to the subject during the experiment for 7 sec.; returns $r_i \sim \mathrm{N}(\mu, \sigma^2)$, here $\mu = 7\%, \sigma = 2\%$

HRF

▸ fMRI dynamics

- Hemodynamic response function e.g. Double Gamma function

$$h(t) = \left(\frac{t}{5.4}\right)^6 \exp\left(-\frac{t-5.4}{0.9}\right) - 0.35\left(\frac{t}{10.8}\right)^{12} \exp\left(-\frac{t-10.8}{0.9}\right),\ t \geq 0\text{-time [sec]}$$

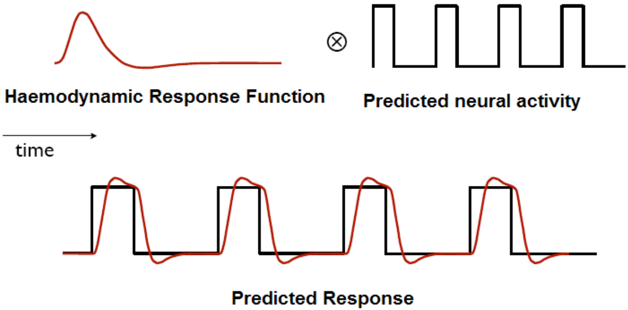


Figure 38: Predicted response as a convolution of a stimulus signal and a HRF. Figure modified from FEAT - FMRI.

Risk Attitude Parameter

- Risk-return choice model

$$V_r^i = \bar{x}_r - \beta_i S_r, \qquad 1 \leq i \leq 19, 1 \leq r \leq 256$$

  - x_r portfolio return stream, $\bar{x}_r$ average return (μ)
  - S_r standard deviation of x_r (risk)
  - V_r^i subjective value (unobserved), 5% risk free return
  - β_i risk attitude parameter

Risk Attitude Parameter

- Estimation of individual risk attitude by logistic regression

$$P\{\text{risky choice}|x\} = \frac{1}{1+\exp\{\bar{x} - \beta S(x) - 5\}}$$

$$P\{\text{sure choice}|x\} = 1 - \frac{1}{1+\exp\{\bar{x} - \beta S(x) - 5\}}$$

  risky choice - unknown return, sure choice - fixed, 5% return

- β estimated by maximum likelihood

Risk Attitude

Cluster Activation: aINS

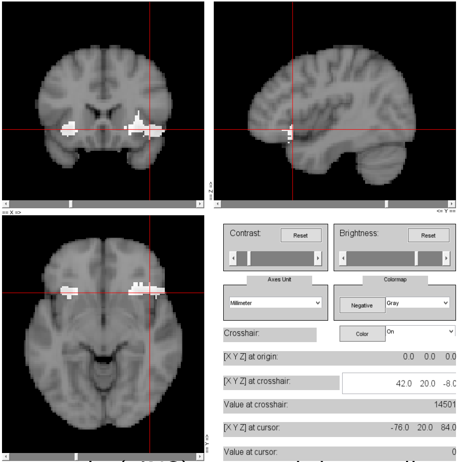


Figure 39: Anterior insula (aINS) activated during all type of investment decisions in the group-level analysis.

Cluster Activation: DMPFC

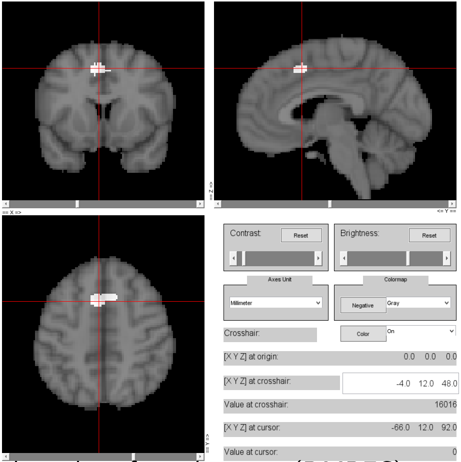


Figure 40: Dorsolateral prefrontal cortex (DMPFC) activated during all type of investment decisions in the group-level analysis. ▸ fMRI Data

B-Spline Basis for Cubic Splines

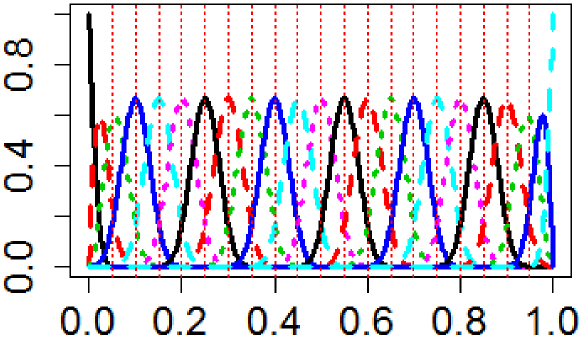


Figure 41: B-Spline basis for Cubic splines with 23 basis functions

Factor Analysis