FASTEC- FActorizable Sparse Tail Event Curves

Shih-Kang Chao
Joint work with
Wolfgang Karl Härdle, Ming Yuan

Department of Statistics, Purdue University
Ladislaus von Bortkiewicz Chair of Statistics, Humboldt-Universität zu Berlin
Department of Statistics, University of Wisconsin-Madison
http://www.stat.purdue.edu/~skchao74
http://lvb.wiwi.hu-berlin.de
http://www.stat.wisc.edu

Holding on the two ends...

Motivation

- Many data come as curves or bunch of time series
- Common structure analysis:
  - Tail (one end): τ is close to 0 or 1. Tail event curves (TEC)
  - Spread (both ends): between $\{q(\tau), q(1-\tau)\}$ τ-range; changes of τ-range: expanding, shrinking, shifting, shifting with expanding/shrinking, $0 < \tau < 1/2$
- Sparsity: common structure is reduced to a few factors
- FASTEC: FActorisable Sparse Tail Event Curve

FASTEC construction

- Data: $\{(\boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i=1}^n$ in $\mathbb{R}^{p+m}$ i.i.d.
- Linear model for τ-quantile curve of Y_j, $j = 1, ..., m$, $0 < \tau < 1$:

$$q_j(\tau|\boldsymbol{X}_i) = \boldsymbol{X}_i^\top \boldsymbol{\Gamma}_{*j}(\tau), \tag{1}$$

where coefficients for j response: $\boldsymbol{\Gamma}_{*j}(\tau) \in \mathbb{R}^p$

- Sparse factorisation: $f_k^\tau(\boldsymbol{X}_i) = \boldsymbol{\varphi}_k^\top(\tau)\boldsymbol{X}_i$ factors

$$q_j(\tau|\boldsymbol{X}_i) = \sum_{k=1}^{r} \psi_{j,k}(\tau) f_k^\tau(\boldsymbol{X}_i), \tag{2}$$

where r : number of factors;
$\boldsymbol{\Gamma}_{*j}(\tau) = (\sum_{k=1}^r \psi_{j,k}(\tau)\varphi_{k,1}(\tau), ..., \sum_{k=1}^r \psi_{j,k}(\tau)\varphi_{k,p}(\tau))$

FASTEC examples

- CAViaR: Y_{ij} log returns at i day and institution j; $\boldsymbol{X}_i$: $\cup_{j=1}^{m}(|Y_{i-1,j}|, Y_{i-1,j}^{-})$ is of $p = 2m$ dimension, $j = 1, ..., m$, $i = 1, ..., n$;
- Temperature data: Y_{ij}: temperature at i day and j weather station; $\boldsymbol{X}_i = (b_1(t_i), ..., b_p(t_i))$, where b_1,...,b_p: B-spline basis, $t = i/n$, $i = 1, ..., n$;
- Further application: image analysis, joint analysis of many images and find the common patterns

Temperature Data

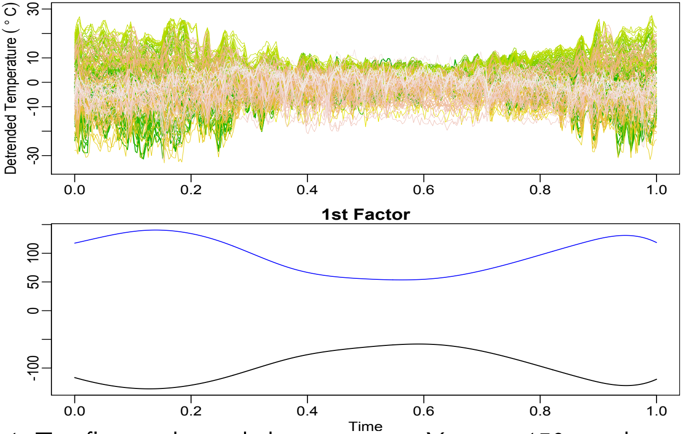


Figure 1: Top figure: detrended temperature Y_{ij}, $m = 159$ weather station, $t = i/n$ time point in year 2008, $n = 365$; bottom figure: quantile factors 1 $f_1^{0.01}(\boldsymbol{X}_i)$ and $f_1^{0.99}(\boldsymbol{X}_i)$; $p = n^{0.4} \approx 11$. FASTECChinaTemper2008

Financial Data

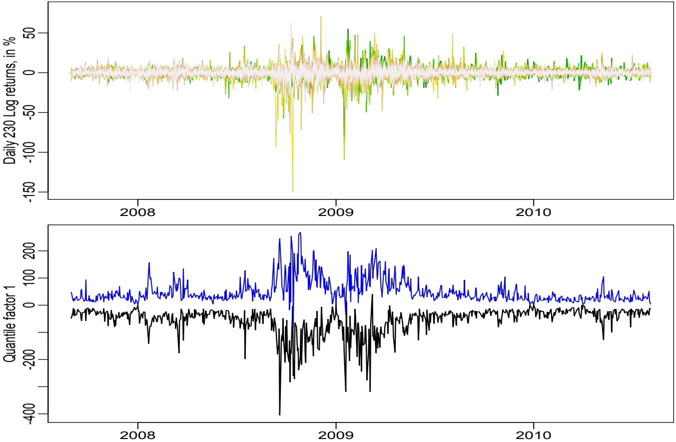


Figure 2: Top figure: log returns Y_{ij}, $n = 765$ ranging from Aug. 2007-Aug. 2010. $m = 230$ firm index. $p = 460$ covariate dimension; bottom figure: quantile factors 1 $f_1^{0.01}(\boldsymbol{X}_i)$, $f_1^{0.99}(\boldsymbol{X}_i)$. FASTECSAMCVaR

Spread gestalt

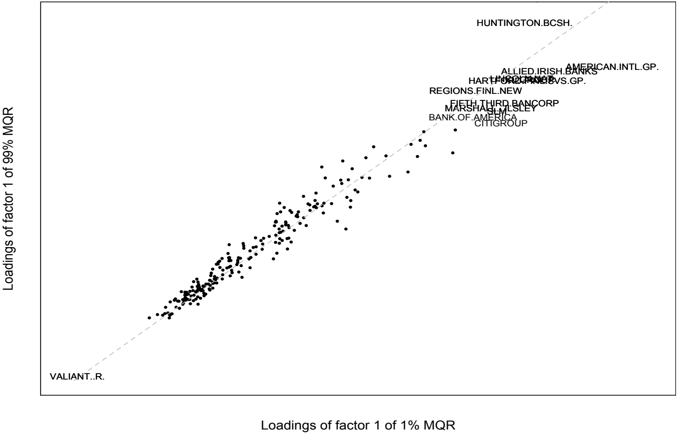


Figure 3: Loadings $(\psi_{j,1}(0.01), \psi_{j,1}(0.99))$ on factors 1 for 230 firms. Close distance indicates similar τ-range pattern. FASTECSAMCVaR

Tail behavior

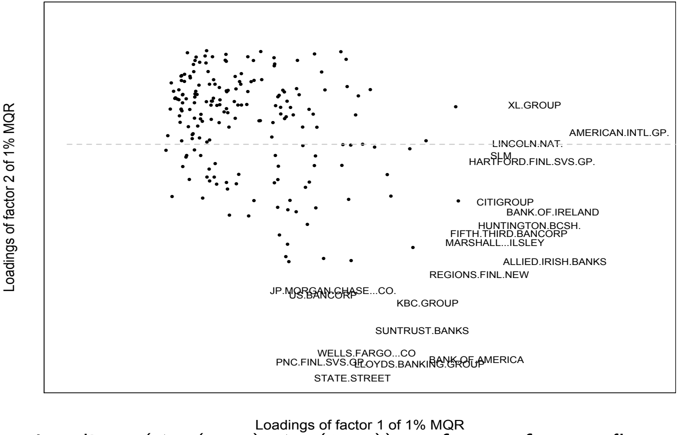


Figure 4: Loadings $(\psi_{j,1}(0.01), \psi_{j,2}(0.01))$ on factors for 230 firms. Close distance implies similar τ-quantile behavior. FASTECSAMCVaR

Challenges

- Implementation of FASTEC needs regularized multivariate quantile regression (MQR)
  - Estimation
  - Proper model tuning
  - Non-asymptotic error bounds
- Applications: High-dimensional joint Value-at-Risk analysis, common temperature risk
- Dimension reduction

Outline

1. Motivation ✓
2. High-dimensional multivariate quantile regression (MQR)
3. Oracle inequalities
4. Numerical analysis
5. Application: Sparse Asymmetric Multivariate Conditional Value-at-Risk (SAMCVaR) Model

Implementing FASTEC: MQR formulation

Recall $q_j(\tau|\boldsymbol{X}_i) = \boldsymbol{X}_i^\top \boldsymbol{\Gamma}_{*j}(\tau)$, for $\boldsymbol{\Gamma} = [\boldsymbol{\Gamma}_{*1}, ..., \boldsymbol{\Gamma}_{*m}]$,

$$L(\boldsymbol{\Gamma}) \stackrel{\text{def}}{=} \underbrace{(nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho_\tau(Y_{ij} - \boldsymbol{X}_i^\top \boldsymbol{\Gamma}_{*j})}_{(A)} + \lambda \underbrace{\|\boldsymbol{\Gamma}\|_*}_{(B)}, \qquad (3)$$

$$\widehat{\boldsymbol{\Gamma}}_{\lambda,\tau} \stackrel{\text{def}}{=} \arg \min_{\boldsymbol{\Gamma} \in \mathbb{R}^{p \times m}} L(\boldsymbol{\Gamma}) \qquad (4)$$

$\rho_\tau(u) = |\mathbf{I}(u \leq 0) - \tau||u|$. $\boldsymbol{\Gamma}_{*j}$: jth column of $\boldsymbol{\Gamma}$. ▸ Shape ρ_τ

- (A): quantile regression fitting quality. Ferguson (1967), Koenker and Bassett (1978), Koenker and Portnoy (1990)

Recall $q_j(\tau|\boldsymbol{X}_i) = \boldsymbol{X}_i^\top \boldsymbol{\Gamma}_{*j}(\tau)$, for $\boldsymbol{\Gamma} \in \mathbb{R}^{p \times m}$,

$$L(\boldsymbol{\Gamma}) \stackrel{\text{def}}{=} \underbrace{(nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho_\tau(Y_{ij} - \boldsymbol{X}_i^\top \boldsymbol{\Gamma}_{*j})}_{(A)} + \lambda \underbrace{\|\boldsymbol{\Gamma}\|_*}_{(B)}, \tag{5}$$

$$\widehat{\boldsymbol{\Gamma}}_{\lambda,\tau} \stackrel{\text{def}}{=} \arg \min_{\boldsymbol{\Gamma} \in \mathbb{R}^{p \times m}} L(\boldsymbol{\Gamma}) \tag{6}$$

- (B): nuclear norm $\|\boldsymbol{\Gamma}\|_* = \sum_{k=1}^{\text{rank}(\boldsymbol{\Gamma})} \sigma_k(\boldsymbol{\Gamma})$ prompts 0 for singular values, $\text{rank}(\boldsymbol{\Gamma})$= number of nonzero singular values
- $\lambda = \lambda_{n,p,m,\mathbf{X},\tau} > 0$ converges to 0 as $n \to \infty$
- $\widehat{\boldsymbol{\Gamma}}$ is estimated by Smooth Fast Iterative Shrinkage-Thresholding Algorithm (SFISTA), convergence rate is analyzed

Theorem (Convergence analysis of SFISTA)

Let $\{\mathbf{\Gamma}_t\}_{t=0}^{T}$ *be the SFISTA sequence, and* $\widehat{\mathbf{\Gamma}}_{\tau,\lambda}$ *minimize* (5) *for* $0 < \tau < 1$ *and* $\lambda > 0$. *Then for any* t *and* $\epsilon > 0$,

$$\left| L(\mathbf{\Gamma}_t) - L(\widehat{\mathbf{\Gamma}}_{\tau,\lambda}) \right| \leq \underbrace{\frac{\epsilon\{\tau \vee (1-\tau)\}^2}{2}}_{\text{Loss from smoothing}} + \underbrace{\frac{4\|\mathbf{\Gamma}_0 - \widehat{\mathbf{\Gamma}}_{\tau,\lambda}\|_{\mathrm{F}}^2 \|\mathbf{X}\|^2}{(t+1)^2 \epsilon m n}}_{\text{convergence of FISTA}}. \quad (7)$$

Requiring $L(\mathbf{\Gamma}_t) - L(\widehat{\mathbf{\Gamma}}_{\tau,\lambda}) \leq \epsilon$ *(e.g.* $\epsilon = 10^{-6}$*) yields*

$$t \geq 2 \frac{\|\widehat{\mathbf{\Gamma}}_{\tau,\lambda} - \mathbf{\Gamma}_0\|_{\mathrm{F}} \|\mathbf{X}\|}{\epsilon \sqrt{mn} \sqrt{1 - \frac{\{\tau \vee (1-\tau)\}^2}{2}}}. \quad (8)$$

$\|\mathbf{X}\|$*: spectral norm (largest singular value) of design matrix* $\mathbf{X}$.

Proof

FASTEC estimation

- High-dimensional setting: $p, m \to \infty$ with n, $m = \dim(\boldsymbol{Y}_i)$; $p = \dim(\boldsymbol{X}_i)$
- Sparsity in factor: $\mathrm{rank}(\boldsymbol{\Gamma})$ is finite and fixed
- Quality measures:
  - Prediction error: $\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|^2_{L_2(\Pi)} \stackrel{\mathrm{def}}{=} m^{-1}\mathsf{E}\|(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})^\top \boldsymbol{X}\|_2^2$, where Π is the distribution for $\boldsymbol{X}$
  - Frobenius error: $\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|^2_{\mathrm{F}} \stackrel{\mathrm{def}}{=} \mathrm{tr}\{(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})^\top\}$
  - Nuclear error: $\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_* = \sum_{k=1}^{\mathrm{rank}(\boldsymbol{\Gamma})} \sigma_k(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})$

Estimation noise

Tuning parameter λ depends on:

$$\Delta_\tau \stackrel{\text{def}}{=} \left\|(mn)^{-1}\mathbf{X}^\top \mathbf{W}_\tau\right\|$$

$(\mathbf{W}_\tau)_{ij} = \mathbf{I}\{Y_{ij} - \boldsymbol{X}_i^\top \boldsymbol{\Gamma}_{*j} \leq 0\} - \tau \sim \mathit{Bernoulli}(\tau)$, $\|\mathbf{X}\|$: matrix spectral norm

Lemma

Under Assumptions 1 and 2,

▸ Assumption

$$n^{-1}\|\mathbf{X}^\top \mathbf{W}_\tau\| \leq C^* \sqrt{\sigma_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{X}})\{\tau \vee (1-\tau)\}}\sqrt{\frac{p+m}{n}}, \qquad (9)$$

with probability greater than $1 - 3e^{-(p+m)\log 8} - \gamma_n$, *where* $\Sigma_{\boldsymbol{X}}$ *is the covariance matrix for* $\boldsymbol{X}_i$, $\gamma_n \to 0, C^* = 4\sqrt{2\frac{c_2}{C'}\log 8}$, C' *and* c_2 *are absolute constants.*

▸ γ_n

Nonasymptotic Risk Bounds

Theorem

Under regularity conditions and ▸ Assumption

$$\lambda = 2C^* \sqrt{\sigma_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{X}})\{\tau \vee (1-\tau)\}}\sqrt{\frac{p+m}{n}},$$

where C^* *and* $\Sigma_{\boldsymbol{X}}$ *are defined in previous page. Then*

$$\|\widehat{\boldsymbol{\Gamma}}_\tau - \boldsymbol{\Gamma}_\tau\|_{L_2(\Pi)} \leq \frac{C_0}{\underline{f}\sqrt{m}}\sqrt{\frac{\sigma_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{X}})}{\sigma_{\min}(\boldsymbol{\Sigma}_{\boldsymbol{X}})}}\sqrt{\tau \vee (1-\tau)}\sqrt{r}\sqrt{\frac{p+m}{n}}, \quad (10)$$

with probability greater than $1 - \gamma_n - 9(p+m)^{-2} - 3e^{-(p+m)\log 8}$ *and* $p+m > 3$, *where* $C_0 = 16\sqrt{2}\left\{\left(\sqrt{\frac{2}{C'}}+4\right)\sqrt{c_2} \vee 4\sqrt{2\frac{c_2}{C'}\log 8}\right\}$.

- Dimensionality:
  - When p, m fixed: the estimator converges in rate $n^{-1/2}$
  - Oracle property: performance depends on unknown number of parameters $r(p+m)$
- Design: condition number $\sigma_{\max}(\mathbf{\Sigma}_{\boldsymbol{X}})/\sigma_{\min}(\mathbf{\Sigma}_{\boldsymbol{X}})$, where $\mathbf{\Sigma}_{\boldsymbol{X}}$ is the covariance for $\boldsymbol{X}$
- Conditional densities:
  - $\underline{f} = \inf_{j \le m} \inf_{\boldsymbol{x}} f_{Y_{ij}|\boldsymbol{X}_i}(\boldsymbol{x}^\top \mathbf{\Gamma}_{*j}|\boldsymbol{x})$
  - Difficult to estimate at τ close to 0 or 1

Frobenius norm and nuclear norm bounds differ to the prediction error bound by a factor $\sqrt{m}$ and a constant

Tuning

- Δ_τ has the same distribution as

$$\Lambda_\tau = (nm)^{-1} \| \mathbf{X}^\top \widetilde{\mathbf{W}}_\tau \|, \quad (11)$$

  where $\widetilde{W}_{ij,\tau} = \mathbf{I}(U_{ij} \le 0) - \tau$, $\{U_{ij}\}$ for $i = 1, ..., n$ and $j = 1, ..., m$ are i.i.d. $\mathrm{U}(0, 1)$
- Λ_τ is pivotal (independent of unknown $\mathbf{\Gamma}$) conditioning on $\boldsymbol{X}$

Tuning

- Bound estimation noise with α quantile of Λ, for small $0 < \alpha < 1$:

$$\lambda_\tau = 2 \cdot q_{\Lambda_\tau}(1 - \alpha|\mathbf{X}), \tag{12}$$

where $q_{\Lambda_\tau}(1 - \alpha|\mathbf{X}) \stackrel{\text{def}}{=} (1 - \alpha)$-quantile of Λ_τ conditional on $\mathbf{X}$ is computed via simulation

- By symmetry, $\lambda_\tau = \lambda_{1-\tau}$
- Pivotal principle: QR-Lasso Belloni and Chernozuhkov (2011) and $\sqrt{\text{Lasso}}$ Belloni, Chernozuhkov and Wang (2011)

Simulation: symmetric situation

- $m = p = 500$, $n = 500$. Iteration=500.
- $\boldsymbol{X}_i$ i.i.d. $N(0, \boldsymbol{\Sigma})$ with $\Sigma_{ij} = 0.5^{|i-j|}$.

$$\boldsymbol{Y}_i = \boldsymbol{\Gamma}^\top \boldsymbol{X}_i + \varepsilon_i, \; \varepsilon_i \sim N(0, \mathbf{I}_m) \text{ i.i.d. } \varepsilon_i \perp \boldsymbol{X}$$

- $\boldsymbol{\Gamma}$ generation: sampling entries from i.i.d. $N(0, 1)$
  1. Model LS (less sparse): The last 375 singular values of $\boldsymbol{\Gamma}$ are 0, $r = \operatorname{rank}(\boldsymbol{\Gamma}) = 125$
  2. Model MS (moderately sparse): Set the first 10 singular values to 30 and the rest 0, $r = 10$
  3. Model ES (extremely sparse): Set the first singular value to 20 and the rest 0, $r = 1$

Simulation: asymmetric situation

- Simulate Y_{ij} with asymmetric conditional quantiles
- $m = p = 500$, $n = 500$. Iteration=500.
- Generating $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ with $\mathrm{rank}(\mathbf{\Gamma}_1) = 2$ and $\mathrm{rank}(\mathbf{\Gamma}_2) = r_2$:
- Model AES (Asymmetric Extremely Sparse): $r_2 = 2$
- Model AMS (Asymmetric Moderately Sparse): $r_2 = 10$

Generating $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$

Simulation: asymmetric situation

- $\boldsymbol{X}_i = \Phi(\tilde{\boldsymbol{X}}_i)$ where $\tilde{\boldsymbol{X}}_i$ i.i.d. $N(0, \boldsymbol{\Sigma})$ with $\Sigma_{ij} = 0.5^{|i-j|}$. $\Phi(\cdot)$: cdf of $N(0, 1)$
- $\{U_{ij}\}$ i.i.d. $U[0, 1]$, $i = 1, ..., n$, $j = 1, ..., m$

$$Y_{ij} = \Phi^{-1}(U_{ij})\boldsymbol{X}_i^\top \left\{\boldsymbol{\Gamma}_{1,*j}\mathbf{I}(U_{ij} < 0.5) + \boldsymbol{\Gamma}_{2,*j}\mathbf{I}(U_{ij} \geq 0.5)\right\}$$

- Quantiles of Y_{ij} given $\boldsymbol{X}$:

$$q_j(\tau|\boldsymbol{X}) = \Phi^{-1}(\tau)\boldsymbol{X}_i^\top\boldsymbol{\Gamma}_{1,*j}, \quad \tau < 0.5;$$
$$q_j(\tau|\boldsymbol{X}) = \Phi^{-1}(\tau)\boldsymbol{X}_i^\top\boldsymbol{\Gamma}_{2,*j}, \quad \tau \geq 0.5.$$

- Given $\boldsymbol{X}_i$, Y_{ij} independent in j

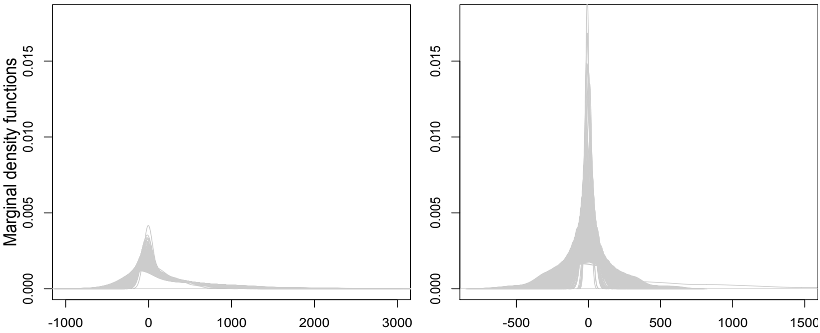


Figure 5: 500 marginal densities (kernel estimators) of $\boldsymbol{Y}_i$ in asymmetric situation. Left figure: AMS shows more asymmetry as the right tail corresponds to higher rank $\boldsymbol{\Gamma}_2$; right: AES shows less asymmetry as the rank of $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are equal.

Performance

- Prediction error: $\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|^2_{L_2(\Pi)}$
  1. V shape: tail quantiles have larger error
  2. For more sparse model: larger λ
- Frobenius error and nuclear norm error show similar patterns as prediction error
- Estimated number of nonzero singular values

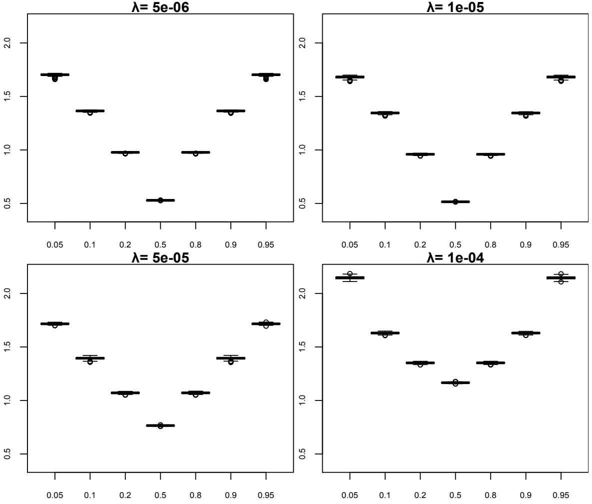


Figure 6: Model LS Prediction Error box plots. Symmetric "V" shape is observed for different choices of λ. Model MS and ES perform similarly.

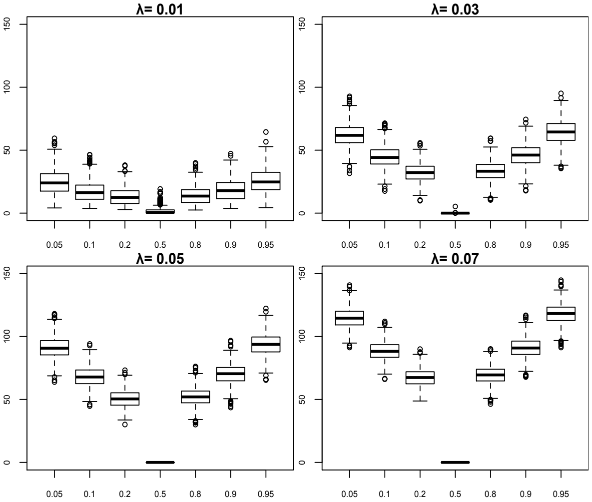


Figure 7: Model AES Prediction Error box plots. Prediction errors present symmetric "V" shape since $\mathrm{rank}(\mathbf{\Gamma}_1) = \mathrm{rank}(\mathbf{\Gamma}_2)$.

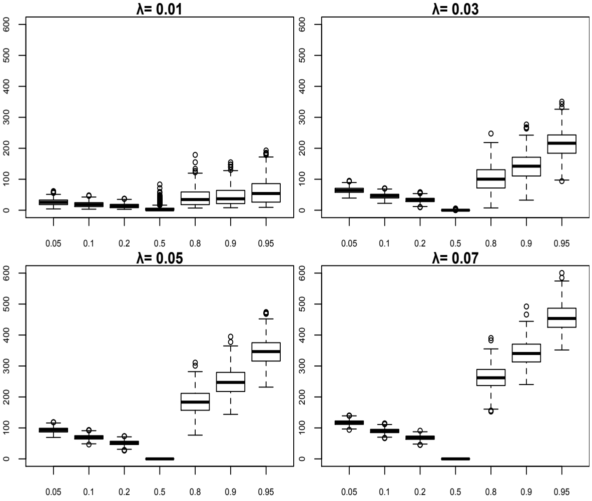


Figure 8: Model AMS Prediction Error box plots. For $\tau > 0.5$ the prediction errors are higher as $\mathrm{rank}(\mathbf{\Gamma}_1) < \mathrm{rank}(\mathbf{\Gamma}_2)$.

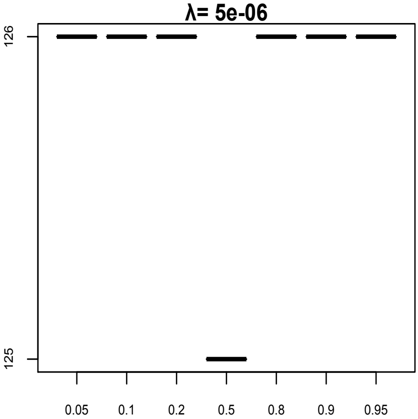


Figure 9: Model LS Estimated number of nonzero singular values box plot. True number of singular value is 125. The result is the same for other choice of λ with certain threshold.

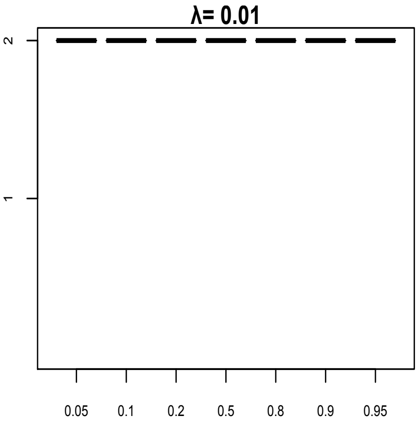


Figure 10: Model AES Estimated number of nonzero singular values box plot. The true number is 2. The result is the same for other choice of λ with certain threshold.

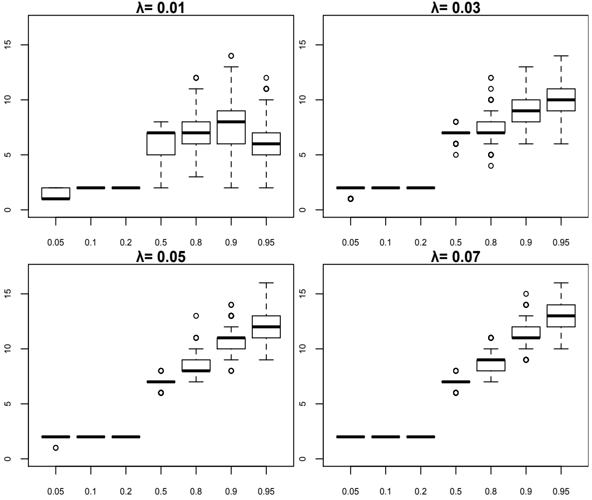


Figure 11: Model AMS Estimated number of nonzero singular values box plots. The true number is 2 for $\tau < 0.5$ and 10 for $\tau \geq 0.5$.

Sparse Asymmetric Multivariate Conditional Value-at-Risk (SAMCVaR)

$$q_{t,j}(\tau|\mathcal{F}_{t-1}) = \boldsymbol{X}_{t-1}^\top \boldsymbol{\Gamma}_{*j},$$
$$\boldsymbol{X}_{t-1} = (|Y_{t-1,1}|, \dots, |Y_{t-1,m}|, Y_{t-1,1}^-, \dots, Y_{t-1,m}^-)^\top \in \mathbb{R}^{2m},$$

where $Y^- = \max\{-Y, 0\}$

- Engle and Manganelli (2004): Conditional Autoregressive Value-at-Risk (CAViaR)
- White et al. (2008): Univariate Multi-Quantile CAViaR (MQ-CAViaR)
- White et al. (2015) "VAR for VaR": estimate bivariate VAR due to computational burden

Factorization

- Factorisation: $r = \mathrm{rank}(\boldsymbol{\Gamma})$,

$$q_{t,j}(\tau|\mathcal{F}_{t-1}) = \sum_{k=1}^{r} \psi_{j,k}(\tau) f_k^{\tau}(\boldsymbol{X}_t) \tag{13}$$

$$f_k^{\tau}(\boldsymbol{X}_t) = \sum_{l=1}^{m} \varphi_{1,k,l}(\tau)|Y_{t-1,l}| + \sum_{l=1}^{m} \varphi_{2,k,l}(\tau) Y_{t-1,l}^{-}, \tag{14}$$

- Flow from component j to f_k^{τ} :

$$\frac{\partial f_k^{\tau}}{\partial(|Y_j|, Y_j^{-})} = \{\varphi_{|\cdot|,k,j}(\tau), \varphi_{-,k,j}(\tau)\}.$$

Sensitivity of j quantile to $f_k(\tau)$: $\dfrac{\partial q_j(\tau|\boldsymbol{X})}{\partial f_k^{\tau}} = \psi_{j,k}(\tau).$

Goals

- Leverage effect: $Y^{-}_{t-1,j} > 0$ implies the increase in $\sigma_{t,j}$. Black (1976) and Engle and Ng (1993)
  - Is leverage effect symmetric? i.e., $|\varphi_{-,k,j}(\tau)| = |\varphi_{-,k,j}(1-\tau)|$?
- Risk sensitivity analysis with τ-range: plot of $\{\psi_{j,1}(\tau), \psi_{j,1}(1-\tau)\}$

Data

- Data period: August 31, 2007 to August 5, 2010. 766 daily closing price for each stock in the sample.

	Banks	Financial Services	Insurances	Total
EU	47	22	27	96
North America	25	17	28	70
Asia	47	14	3	64
Total	119	53	58	$m = 230$

Table 1: Financial firms summary.

- $p = 2m = 460$ (2 transformations of lag return $|Y_{t-1,j}|$, $Y^{-}_{t-1,j}$)
- Downloaded from Simone Manganelli's website
- Using tuning method introduced previously: $\lambda = 0.0247$ for both $\tau = 1\%$ and 99%

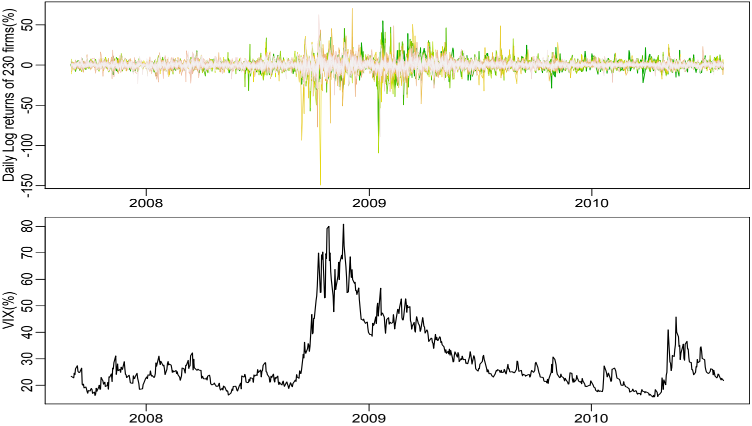


Figure 12: Time series plots of log returns Y_{ij}, i ranging from Aug. 31, 2007-Aug. 5, 2010. $n = 765$. $j = 1, \ldots, 230$ firm. The lower figure shows the time series of VIX. FASTECSAMCVaR

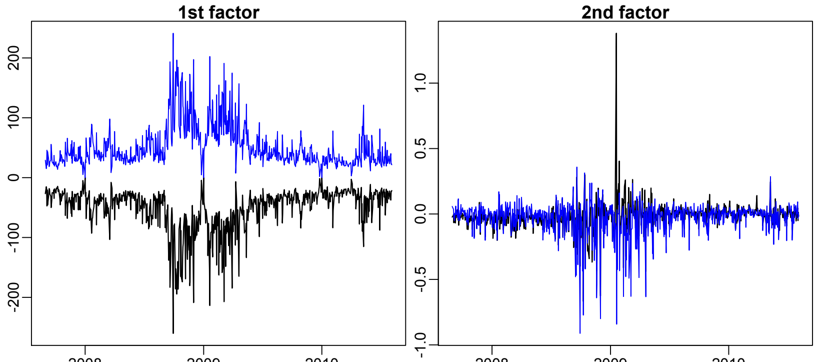


Figure 13: Time series of first factors $f_1^{0.01}$, $f_1^{0.99}$ (left) and second factors $f_2^{0.01}$, $f_2^{0.99}$ (right). Large deviation periods of first factors correspond to that of VIX. The magnitude of factor 2 is much smaller than that of factor 1. FASTECSAMCVaR

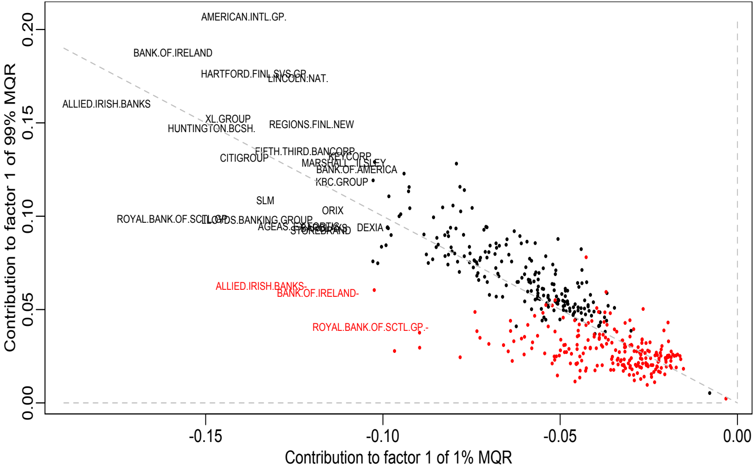


Figure 14: Scatter plot of $(\varphi_{|\cdot|,1,j}(0.01), \varphi_{|\cdot|,1,j}(0.99))$ and $(\varphi_{-,1,j}(0.01), \varphi_{-,1,j}(0.99))$ for $j = 1, ..., 230$. $Y^{-}_{t-1,j}$ relates more to left dispersion, while $|Y_{t-1,j}|$ contributes symmetrically. FASTECSAMCVaR

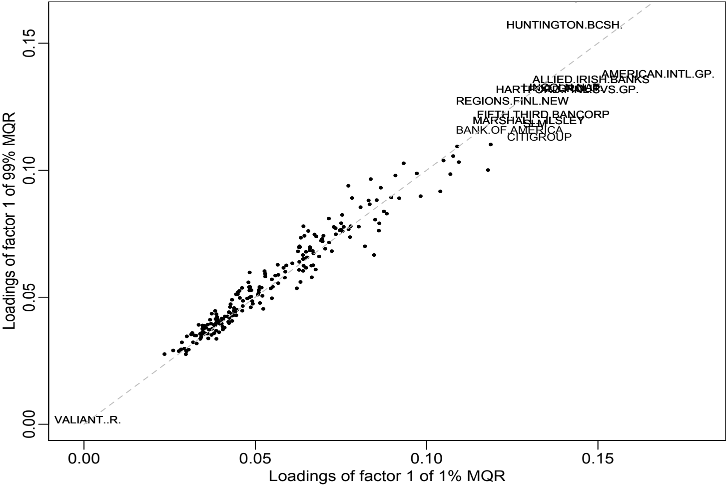


Figure 15: Scatter plot of loadings $(\psi_{j,1}(0.01), \psi_{j,1}(0.99))$ for $j = 1, ..., 230$ firms. Firms on the northeast corner are more associated to the extreme event of the market. FASTECSAMCVaR

FASTEC: SAMCVaR

- Figure 5-7
  - leverage effect: $Y^{-}_{t-1,j}0$ leads to left τ-range expansion
  - $|Y_{t-1,j}|$ contributes symmetrically to τ-range
- Figure 5-8 Large loadings $(\psi_{j,1}(0.01), \psi_{j,1}(0.99))$, large τ-range

Chinese Temperature Data

- Temperature data from $m = 159$ weather stations in China in year 2008, downloaded from Research Data Center of CRC 649

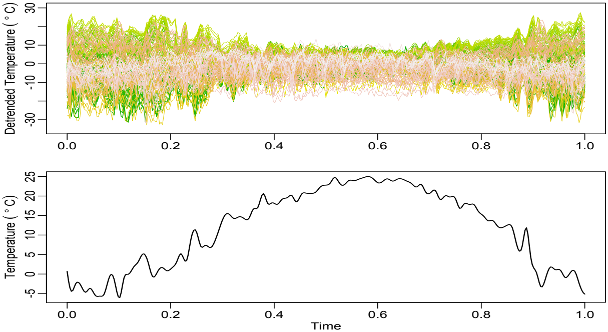


Figure 16: Upper: detrended temperature $Y_j(t)$ and yearly trend by smoothing spline. j: weather station, $t \in [0, 1]$ time point in year 2008. Lower: trend. Detrending FASTECChinaTemper2008

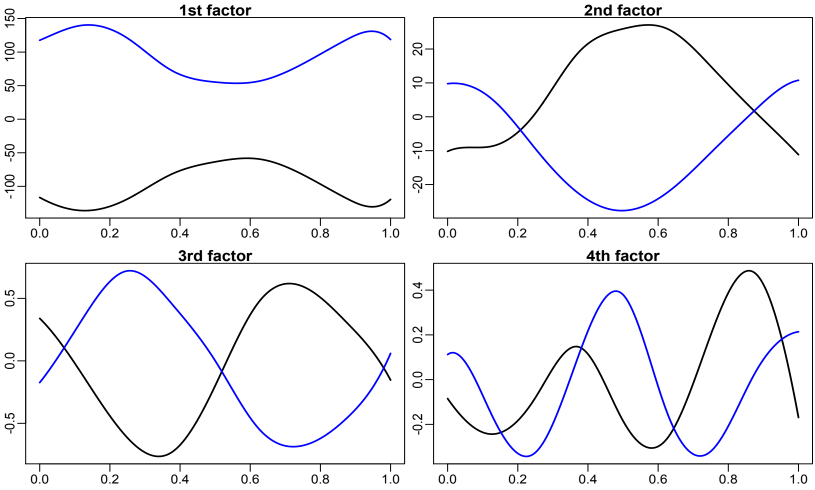


Figure 17: The first 4 factor curves. $\tau = 90\%$. $\tau = 10\%$.
FASTECChinaTemper2008

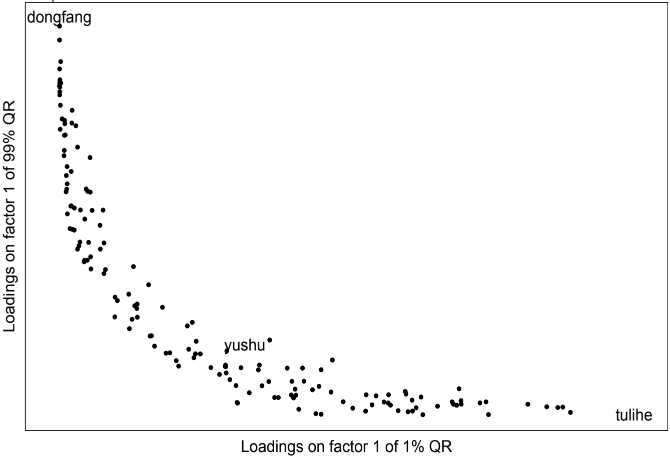
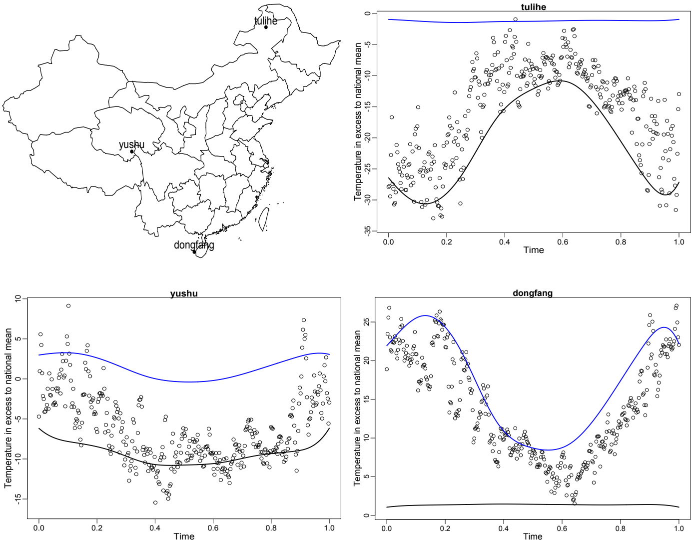


Figure 18: Scatter plot of factor loadings of weather station j, $j = 1, ..., 159$, demonstrates a "L"-like shape: stations associated with factor 1 of 1% have almost no association with that of 99%.

FASTECChinaTemper2008


tulihe
yushu
dongfang
tulihe
Temperature in excess to national mean
0
-5
-10
-15
-20
-25
-30
-35
0.0
0.2
0.4
0.6
0.8
1.0
Time
yushu
Temperature in excess to national mean
10
5
0
-5
-10
-15
0.0
0.2
0.4
0.6
0.8
1.0
Time
dongfang
Temperature in excess to national mean
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
Time

Temperature analysis

- The algorithm classifies the northern and southern temperature patterns
- "L" like shape in Figure 18: stations associated with factor 1 of 1% have almost no association with that of 99%
- Stations in the middle cannot be explained by either northern or southern temperature pattern
- Yuchu: region avg. 4000 meters high above sea level, highland climate with reverse seasonality

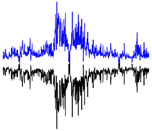
Summary and Extensions

- Conditional quantiles are useful for studying tail events and spread of dispersion
- Nuclear norm regularized multivariate quantile regression
- Algorithm and oracle properties are derived

Further research directions:

- Expectile regression, support vector machine (non-smooth loss)
- Confidence intervals for singular values
- Nonconvex penalty. e.g. nonconvex adaptive nuclear norm $\|\mathbf{\Gamma}\|_* = \sum_{i=1}^{p \wedge m} w_i d_i(\mathbf{\Gamma})$ by Chen, Dong and Chan (2013, Biometrika)

FASTEC- FActorizable Sparse Tail Event Curves



Shih-Kang Chao
Joint work with
Wolfgang Karl Härdle, Ming Yuan

Department of Statistics, Purdue University
Ladislaus von Bortkiewicz Chair of Statistics, Humboldt-Universität zu Berlin
Department of Statistics, University of Wisconsin-Madison
http://www.stat.purdue.edu/~skchao74
http://lvb.wiwi.hu-berlin.de
http://www.stat.wisc.edu

Check function

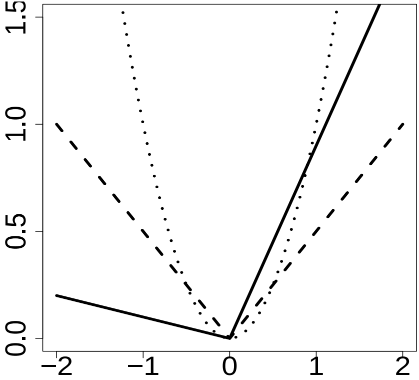


Figure 19: Solid line: $\tau = 0.9$. Dashed line: $\tau = 0.5$. Dotted line: square loss u^2 (OLS regression). LQRcheck

Loss function

Nonsmooth loss function: $\frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}\rho_\tau\left(Y_{ij} - \boldsymbol{X}_i^\top \boldsymbol{\Gamma}_{*j}\right) + \lambda\|\boldsymbol{\Gamma}\|_*$

Introduce dual variables

$$\max_{\Theta_{ij}\in[\tau-1,\tau]} \ell(\boldsymbol{\Gamma},\boldsymbol{\Theta})$$

$$\ell(\boldsymbol{\Gamma},\boldsymbol{\Theta}) = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}\Theta_{ij}(Y_{ij} - \boldsymbol{X}_i^\top \boldsymbol{\Gamma}_{*j})$$

Smoothing by Nesterov (2005)

$$f_\kappa(\boldsymbol{\Gamma}) = \max_{\Theta_{ij}\in[\tau-1,\tau]}\left\{\ell(\boldsymbol{\Gamma},\boldsymbol{\Theta}) - \frac{\kappa}{2}\|\boldsymbol{\Theta}\|_F^2\right\}$$

$$\nabla_{\boldsymbol{\Gamma}} f_\kappa(\boldsymbol{\Gamma}) = -(mn)^{-1}\mathbf{X}^\top[[(\kappa mn)^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\Gamma})]]_\tau$$

Lipschitz constant $M = (\kappa m^2 n^2)^{-1}\|\mathbf{X}\|^2$

$\kappa = \epsilon/2mn$

Project on low rank space

$$S_\lambda(\boldsymbol{\Gamma}) \stackrel{\text{def}}{=} \mathbf{U}(\mathbf{D} - \lambda\mathbf{I})_+\mathbf{V}^\top$$

Algorithm

$[[\cdot]]_\tau$ and Theorem of Nesterov

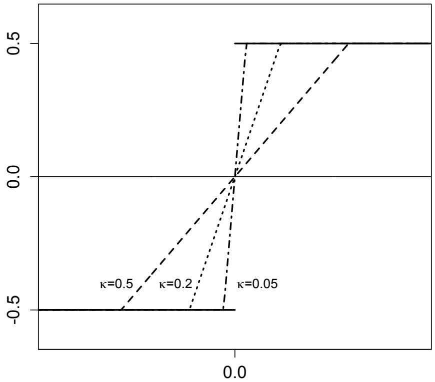


Figure 20: $\mathbf{X} = 1$, $m = p = n = 1$. Solid line: $\psi_\tau(u) = \tau - \mathbf{I}(u \leq 0)$ with $\tau = 0.5$; Dashed, dotted, dot-dash line: smoothing gradient $[[\kappa^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\Gamma})]]_\tau$, $\kappa = 0.5, 0.2, 0.05$. ▸ Algorithm

Nonsmooth loss function: $\frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}\rho_\tau\left(Y_{ij}-\boldsymbol{X}_i^\top\boldsymbol{\Gamma}_{*j}\right)+\lambda\|\boldsymbol{\Gamma}\|_*$

Introduce dual variables

$$\max_{\Theta_{ij}\in[\tau-1,\tau]}\ell(\boldsymbol{\Gamma},\boldsymbol{\Theta})$$

$$\ell(\boldsymbol{\Gamma},\boldsymbol{\Theta})=\frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}\Theta_{ij}(Y_{ij}-\boldsymbol{X}_i^\top\boldsymbol{\Gamma}_{*j})$$

Smoothing by Nesterov (2005)

$$f_\kappa(\boldsymbol{\Gamma})=\max_{\Theta_{ij}\in[\tau-1,\tau]}\left\{\ell(\boldsymbol{\Gamma},\boldsymbol{\Theta})-\frac{\kappa}{2}\|\boldsymbol{\Theta}\|_F^2\right\}$$

$$\nabla_{\boldsymbol{\Gamma}}f_\kappa(\boldsymbol{\Gamma})=-(mn)^{-1}\mathbf{X}^\top[[(\kappa mn)^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\Gamma})]]_\tau$$

Lipschitz constant $M=(\kappa m^2n^2)^{-1}\|\mathbf{X}\|^2$

$\kappa=\epsilon/2mn$

Project on low rank space

$$S_\lambda(\boldsymbol{\Gamma})\stackrel{\text{def}}{=}\mathbf{U}(\mathbf{D}-\lambda\mathbf{I})_+\mathbf{V}^\top$$

▸ Algorithm

▸ Proximity operator

Detrending of Chinese temperature data

- Chapter 4 of Ramsay and Silverman (2005): smooth discretized data with smoothing spline
- Estimation of mean function and smoothing are done jointly by minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\left[Y_{ij}-\hat{\mu}(t_i)\right]^2+\eta\int[D^2\hat{\mu}(s)]^2ds \tag{15}$$

where $\eta > 0$ is a smoothing parameter selected by cross-validation and $\hat{\mu}$ is fitted by cubic spline basis

Introduction-temperature data Application-temperature data

$$[[a_{ij}]]_\tau = \begin{cases} \tau, & \text{if } a_{ij} \geq \tau; \\ a_{ij}, & \text{if } \tau - 1 < a_{ij} < \tau; \\ \tau - 1, & \text{if } a_{ij} \leq \tau - 1. \end{cases}$$

Theorem

For any $\kappa > 0$, $f_\kappa(\boldsymbol{\Gamma})$ is well-defined, convex and continuously-differentiable function in $\boldsymbol{\Gamma}$ with the gradient $\nabla f_\kappa(\boldsymbol{\Gamma}) = -(mn)^{-1}\mathbf{X}^\top \boldsymbol{\Theta}^*(\boldsymbol{\Gamma}) \in \mathbb{R}^{p\times m}$, where $\boldsymbol{\Theta}^*(\boldsymbol{\Gamma})$ is the optimal solution to $\max_{\Theta_{ij}\in[\tau-1,\tau]} \left\{(mn)^{-1}\ell(\boldsymbol{\Gamma},\boldsymbol{\Theta}) - \frac{\kappa}{2}\|\boldsymbol{\Theta}\|_{\mathrm{F}}^2\right\}$, namely

$$\boldsymbol{\Theta}^*(\boldsymbol{\Gamma}) = [[(\kappa mn)^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\Gamma})]]_\tau. \tag{16}$$

Moreover, the gradient $\nabla f_\kappa(\boldsymbol{\Gamma})$ is Lipschitz continuous with the Lipschitz constant $M = (\kappa m^2 n^2)^{-1}\|X\|^2$.

▸ Smoothing the loss

Definition (Proximity Operator)

Let $\mathcal{X} = \mathbb{R}^{p \times n}$ with inner product $\langle \mathbf{A}, \mathbf{B} \rangle = \operatorname{tr}(\mathbf{A}^\top \mathbf{B})$ and $\|\cdot\|$ be the induced norm. $f : \mathcal{X} \to \mathbb{R}$ a lower semicontinuous convex function. The proximity operator of f, $S_f : \mathcal{X} \to \mathcal{X}$:

$$S_f(\mathbf{Y}) \stackrel{\text{def}}{=} \arg \min_{\mathbf{X} \in \mathcal{X}} \left\{ f(\mathbf{X}) + \frac{1}{2} \|\mathbf{X} - \mathbf{Y}\|^2 \right\}, \forall \mathbf{Y} \in \mathcal{X}.$$

Theorem (Cai et al. (2010))

SVD: $\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$. *The proximity operator* $S_\lambda(\cdot)$ *of* $\lambda\|\cdot\|_*$ *is*

$$S_\lambda(\mathbf{Y}) \stackrel{\text{def}}{=} \mathbf{U}(\mathbf{D} - \lambda \mathbf{I})_+ \mathbf{V}^\top, \tag{17}$$

▸ Estimating Γ ▸ Smoothing the loss

Proof.

- $\ell(\Gamma) = (nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho_\tau(Y_{ij} - X_i^\top \Gamma_{*j})$
- $L(\Gamma) \stackrel{\text{def}}{=} \ell(\Gamma) + \lambda \|\Gamma\|_*$
- $\widetilde{L}(\Gamma) = f_\kappa(\Gamma) + \lambda \|\Gamma\|_*$
- $f_\kappa(\Gamma) = \min_{\Theta \in [\tau-1,\tau]^{n\times m}} \ell(\Gamma, \Theta) - \frac{\kappa}{2}\|\Theta\|_F^2$

$$L(\Gamma_t) - L(\widehat{\Gamma}) = L(\Gamma_t) - \widetilde{L}(\Gamma_t) + \widetilde{L}(\Gamma_t) - \widetilde{L}(\widehat{\Gamma}) + L(\widehat{\Gamma}) - \widetilde{L}(\widehat{\Gamma}).$$

1. $\widetilde{L}(\Gamma) \le L(\Gamma) \le \widetilde{L}(\Gamma) + \kappa \max_{\Theta \in [\tau-1,\tau]^{n\times m}} \frac{\|\Theta\|_F^2}{2} \le \widetilde{L}(\Gamma) + \kappa \mu(\tau)^2 \frac{nm}{2}$
2. BT(2009): $\left|\widetilde{L}(\Gamma_t) - \widetilde{L}(\widehat{\Gamma})\right| \le \frac{2M\|\Gamma_0 - \widehat{\Gamma}\|_F^2}{(t+1)^2}$, $M = (\kappa m^2 n^2)^{-1}\|X\|^2$: Lipschitz constant of $\nabla f_\kappa(\Gamma)$,

□

▸ Convergence of SFISTA ▸ Proof of oracle property

Nonasymptotic risk bounds

Generalization of support using projections:

- SVD: $\mathbf{A} = \sum_{j=1}^{r} \sigma(\mathbf{A})\mathbf{u}_j\mathbf{v}_j^\top$ for matrix $\mathbf{A}$
- $\mathbf{U}_r = [\mathbf{u}_1, ..., \mathbf{u}_r]$, $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_r]$.
  $P_{\mathbf{A},1} = \mathbf{U}_r\mathbf{U}_r^\top$; $P_{\mathbf{A},2} = \mathbf{V}_r\mathbf{V}_r^\top$ are orthogonal projections
- $\mathcal{P}_{\mathbf{A}}(\mathbf{S}) \stackrel{\text{def}}{=} \mathbf{S} - P_{\mathbf{A},1}^{\perp}\mathbf{S}P_{\mathbf{A},2}^{\perp}$; $\mathcal{P}_{\mathbf{A}}^{\perp}(\mathbf{S}) \stackrel{\text{def}}{=} P_{\mathbf{A},1}^{\perp}\mathbf{S}P_{\mathbf{A},2}^{\perp}$

The cone

$$\mathcal{K}(\mathbf{\Gamma}, c_0) \stackrel{\text{def}}{=} \left\{\mathbf{S} \in \mathbb{R}^{p\times m} : \|\mathcal{P}_{\mathbf{\Gamma}}^{\perp}(\mathbf{S})\|_* \le c_0\|\mathcal{P}_{\mathbf{\Gamma}}(\mathbf{S})\|_*\right\}. \tag{18}$$

The norm: $\|\mathbf{S}\|_{L_2(\Pi)}^2 \stackrel{\text{def}}{=} m^{-1}\mathsf{E}_{\Pi}\|\mathbf{S}^\top \boldsymbol{X}_i\|_2^2$ Nonasymptotic Risk Bounds

Assumption (Sampling setting)

Samples $(\boldsymbol{X}_1, \boldsymbol{Y}_1), ..., (\boldsymbol{X}_n, \boldsymbol{Y}_n)$ are i.i.d. copies of $(\boldsymbol{X}, \boldsymbol{Y})$ random vectors in $\mathbb{R}^{p+m}$. $F^{-1}_{Y_{ij}|\boldsymbol{X}_i}(\tau|\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\Gamma}_{*j}(\tau)$. Conditioning on $\boldsymbol{X}_i$, Y_{ij} is independent in j.

Assumption (Covariance matrix condition)

Let the covariance matrix of $\boldsymbol{X}$ be $\boldsymbol{\Sigma}_{\boldsymbol{X}}$, assume that $0 < \sigma_{\min}(\boldsymbol{\Sigma}_{\boldsymbol{X}}) < \sigma_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{X}}) < \infty$. Moreover, assume the sample covariance matrix of covariates $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{X}} = \frac{1}{n}\mathbf{X}^\top\mathbf{X}$ satisfies

$$\mathrm{P}\left[\sigma_{\min}(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}) \geq c_1 \sigma_{\min}(\boldsymbol{\Sigma}_{\boldsymbol{X}}), \sigma_{\max}(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}) \leq c_2 \sigma_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{X}})\right] \geq 1 - \gamma_n. \quad (19)$$

Covariates come from a joint p-Gaussian distribution $N(0, \boldsymbol{\Sigma}_{\boldsymbol{X}})$: $c_1 = 1/9$, $c_2 = 9$ and $\gamma_n = 4\exp(-n/2)$ from Wainwright (2009)

▸ Nonasymptotic Risk Bounds ▸ Estimation noise

Assumption (Conditional density condition)

There exist $\underline{f} > 0$ and $\bar{f}' < \infty$ such that $|\frac{\partial}{\partial y_j} f_{Y_{ij}|\boldsymbol{X}_i}(y_i|\boldsymbol{x})| \leq \bar{f}'$ and $\inf_{j\leq m} \inf_{\boldsymbol{x}} f_{Y_{ij}|\boldsymbol{X}_i}(\boldsymbol{x}^\top \boldsymbol{\Gamma}_{*j}|\boldsymbol{x}) \geq \underline{f}$, where $f_{Y_{ij}|\boldsymbol{X}_i}$ is the conditional density function of Y_{ij} on $\boldsymbol{X}_i$.

Assumption (Restricted eigenvalue)

For a given probability distribution Π for $\boldsymbol{X}$,

$$\beta_{\boldsymbol{\Gamma},3} \stackrel{\text{def}}{=} \inf \left\{ \beta > 0 : \beta \|\mathcal{P}_{\boldsymbol{\Gamma}}(\boldsymbol{\Delta})\|_{\mathrm{F}} \leq \|\boldsymbol{\Delta}\|_{L_2(\Pi)},\ \forall \boldsymbol{\Delta} \in \mathcal{K}(\boldsymbol{\Gamma}, 3) \right\} > 0. \tag{20}$$

A rough lower bound: $\beta_{\boldsymbol{\Gamma},3} \geq m^{-1/2} \sqrt{\sigma_{\min}(\boldsymbol{\Sigma}_{\boldsymbol{X}})}$ ▸ Nonasymptotic Risk Bounds

Assumption (Restricted nonlinearity)

$$\nu \stackrel{\text{def}}{=} \frac{3}{8}\frac{\underline{f}}{\bar{f}'} \inf_{\substack{\boldsymbol{\Delta}\in\mathcal{K}(\boldsymbol{\Gamma},3)\\ \boldsymbol{\Delta}\neq\mathbf{0}}} \frac{\|\boldsymbol{\Delta}\|^3_{L_2(\Pi)}}{m^{-1}\sum_{j=1}^m \mathsf{E}[|\boldsymbol{X}_i^\top \boldsymbol{\Delta}_{*j}|^3]}, \tag{21}$$

$$\nu > \frac{C'_\tau}{\underline{f}\sqrt{m}}\sqrt{\frac{\sigma_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{X}})}{\sigma_{\min}(\boldsymbol{\Sigma}_{\boldsymbol{X}})}}\sqrt{\tau\vee(1-\tau)}\left(\sqrt{\frac{\log(p+m)}{n}}+\sqrt{\frac{p+m}{n}}\right)\sqrt{r}. \tag{22}$$

Section 2.5 of Belloni and Chernozuhkov (2011) calculate ν for various data generating processes

Nonasymptotic Risk Bounds

Asymmetric situation: Γ generation

1. Basis vectors $\{v_1, v_2\}$ and $\{u_1, ..., u_{r_2}\}$ in $\mathbb{R}^p$. Components in v_j and u_k follow $U[0,1]$
2. $\mathbf{\Gamma}_{1,*j} = a_{1,j} v_1 + a_{2,j} v_2$, $a_{1,j}, a_{2,j} \sim U[0,1]$ i.i.d.;
   $\mathbf{\Gamma}_{2,*j} = b_{1,j} u_1 + ... + b_{r_2,j} u_{r_2}$, $b_{1,j},...,b_{r_2,j} \sim U[0,1]$ i.i.d.

Asymmetric Models

References

Beck, A. and Teboulle, M.
A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems
SIAM J. Imaging Sciences (2009) Vol. 2, No. 1, 183-202

Belloni, A. and Chernozhukov, V.
ℓ_1-penalized quantile regression in high-dimensional sparse models,
The Annals of Statistics (2011) Vol. 39, No. 1, 82-130.

Belloni, A., Chernozhukov, V. and Wang, L.
Square-root lasso: pivotal recovery of sparse signals via conic programming
Biometrika(2011) Vol. 98, No. 4, 791-806.

Black, F.
Studies of stock market volatility changes
Proceedings of the American Statistical Association (1976), 177-181

References

Bunea, F., She, Y. and Wegkamp, M. H.
Optimal selection of reduced rank estimators of high-dimensional matrices
The Annals of Statistics (2011) Vol. 39, No. 2, 1282-1309

Engle, R. and Manganelli, S.
CAViaR: Conditional autoregressive value at risk by regression quantiles
Journal of Business & Economic Statistics (2004) Vol. 22, 367-381

Engle, R. F. and Ng, V.
Measuring and testing the impact of news on volatility
Journal of Finance (1993) Vol. 48, 1749-1778.

Fazel, M.
Matrix rank minimization with applications
Ph.D. thesis (2002), Standford University

References

Koenker, R. and Portnoy, S.
M estimation of multivariate regressions
Journal of American Statistical Association(1990) Vol. 85(412), 1060-1068.

Koltchinskii, V., Lounici, K. and Tsybakov A. B.
Nuclear-Norm Penalization and Optimal Rates for Noisy Low-Rank Matrix Completion
The Annals of Stat. (2011), Vol. 39, No. 5, 2302-2329

Ramsay J. O. and Silverman B. W.
Functional Data Analysis
Springer (2005), New York

Reinsel G. C. and Velu R. P.
Multivariate Reduced-Rank Regression
Springer (1998), New York

References

Wainwright, M. J.
Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso),
IEEE Transactions on Information Theory(2009) 55: 2183-2202.

White, H., Kim, T.-H. and Manganelli, S.
Modeling autoregressive conditional skewness and kurtosis with multi-quantile CAViaR
Volatility and Time Series Econometrics: A Festschrift in Honor of Robert F. Engle.

White, H., Kim, T.-H. and Manganelli, S.
VAR for VaR: measuring systemic risk using multivariate regression quantiles
MPRA Paper No. 35372

References

Yuan, M., Ekici, A., Lu, Z. and Monteiro, R.
Dimension reduction and coefficient estimation in multivariate linear regression
J. R. Stat. Soc. Ser. B Stat. Methodol. (2007) Part 3, 329-346.