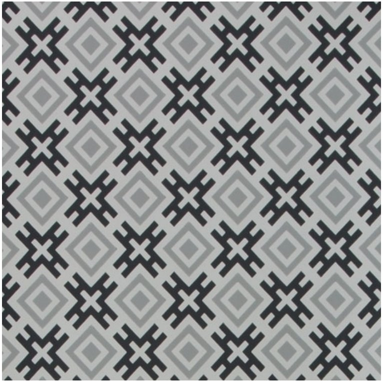
Nuclear Norm Penalized Large Multiple Quantile Regression

Shih-Kang Chao
Wolfgang Karl Härdle
Ming Yuan

Department of Statistics
University of Wisconsin-Madison
Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
https://www.stat.wisc.edu

Patterns adaptation?



(a) Y: Modern English design

(b) X: Ancient Islamic design

Association between large sets of variables

We consider the multiple regression model

$$\boldsymbol{y}_i = \Gamma^\top \boldsymbol{x}_i + E_i, \quad i = 1, ..., n, \tag{1}$$

where $\boldsymbol{y}_i \in \mathbb{R}^m$, $\boldsymbol{x}_i \in \mathbb{R}^p$, $\Gamma \in \mathbb{R}^{p\times m}$. E_i is a vector of i.i.d. random variables. $Y = [\boldsymbol{y}_1, ..., \boldsymbol{y}_n]^\top$ and $X = [\boldsymbol{x}_1, ..., \boldsymbol{x}_n]^\top$

- Classical applications:
  - David and Tso (1982): engine test ratings to composition of hydrocarbon fuel
  - Gudmundsson (1977): macroeconomic variables of UK
  - more examples in Reinsel and Velu (1998)
- Large VAR model: Yuan et al. (2007), Han and Liu (2013).

$$\boldsymbol{y}_t = \Gamma^\top \boldsymbol{y}_{t-1} + E_t, \quad t = 1, ..., T,$$

where $\boldsymbol{y}_t$: (a subset of) S&P500 stock returns.

Challenges

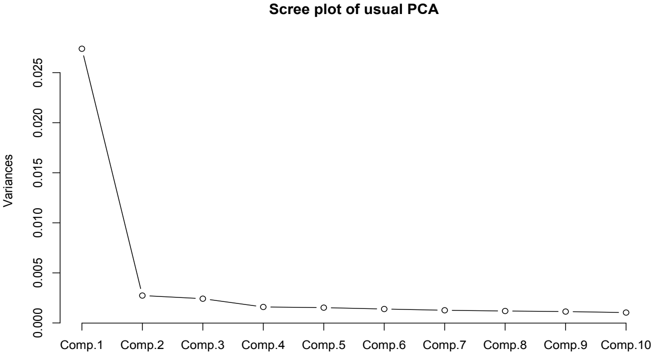
(C1) Direct estimation (eqn.-by-eqn. estimation) neglects the correlation in $\boldsymbol{y}$: $\widehat{\Gamma} = \arg \min_{\Gamma \in \mathbb{R}^{p \times m}} \|Y - X\Gamma\|_{\mathrm{F}}$, where $\|\cdot\|_{\mathrm{F}}$: matrix Frobenius norm

(C2) Using canonical correlation or principal component analysis: depends sample covariance matrix, which is inaccurate when $p/n \to c \in (0, 1)$.
$X \in \mathbb{R}^{n \times p}$: random matrix with i.i.d. standard normal entries.
- Marcenko-Pastur law (1967): the empirical distribution of the eigenvalues of $X^\top X/n$ is supported on interval $\big((1-\sqrt{c})^2, (1+\sqrt{c})^2\big)$
- Geman (1980): $\lambda_1 \to (1+\sqrt{c})^2$ a.s.

(C3) Behavior of the marginal conditional distributions of $\boldsymbol{y}$?

Table 1: Principal components of return of constituents of S&P500. $m = 495$ firms. Data period: Oct. 26, 2012-Sep. 24, 2014. $n = 499$.

PC	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6
Prop. of var.	.279	.028	.025	.016	.016	.014
Cum. prop.	.279	.307	.332	.348	.364	.378



Reduced-rank model

(C1) and (C2) are addressed. Recall model (1)

$$\boldsymbol{y}_i = \Gamma^\top \boldsymbol{x}_i + E_i, \quad i = 1, \ldots, n.$$

- Reinsel and Velu (1998): Allowing Γ matrix degenerate in rank: $\mathrm{rank}(\Gamma) = r \leq \min\{p, m\}$. r needs to be selected.
  - Introducing $\min\{p, m\} - r$ linear restrictions (unknown a priori)
  - Reducing number of unknowns: requiring less samples
  - Convenience for interpretation: factor analysis
- Rank is data driven: Penalized method.
  - Bunea et al. (2010, AoS):
    $\widehat{\Gamma} = \arg \min_{\Gamma \in \mathbb{R}^{p \times m}} \|Y - X\Gamma\|_{\mathrm{F}} + \lambda \, \mathrm{rank}(\Gamma)$
  - Yuan et al. (2007, JRSSB) and Bunea et al. (2010, AoS):
    $\widehat{\Gamma} = \arg \min_{\Gamma \in \mathbb{R}^{p \times m}} \|Y - X\Gamma\|_{\mathrm{F}} + \lambda \|\Gamma\|_*$, where $\|\cdot\|_*$: nuclear norm

Reduced-rank model v.s. factor analysis

- Suppose $\mathrm{rank}(\Gamma) = r$.
- *Singular value decomposition* of Γ yields $\Gamma = UDV^\top$. Where $D \in \mathbb{R}^{p\times m}$ is rectangular diagonal matrix with only r nonzero entries.

We can write the model (1) as a factor regression model:

$$\boldsymbol{y}_i = \Gamma^\top \boldsymbol{x}_i + E_i = VDU^\top \boldsymbol{x}_i \stackrel{\mathrm{def}}{=} VF_i + E_i. \tag{2}$$

where $F_i = DU^\top \boldsymbol{x}_i$ are factors and V are factor loadings

Factors for quantiles other than the mean

Approaching (C3) using quantile regression concept based on (2). Modeling the $\tau-$quantile of each y_j in $\boldsymbol{y}$ by

$$\boldsymbol{y}_i = Q_\tau F_{\tau,i} + E_i, \tag{3}$$

where F_τ of dimension r_τ is the factor at level τ and Q_τ the factor loadings at level τ. E_i is errors with τ quantile 0

- Reduced rank concept enters by r_τ
- In general, r_τ and F_τ vary with τ
- F_τ explain the joint variation of marginal τ-quantiles
- By looking at different τ, we gain more insight into the marginal conditional distribution of $\boldsymbol{y}$ by using joint factors

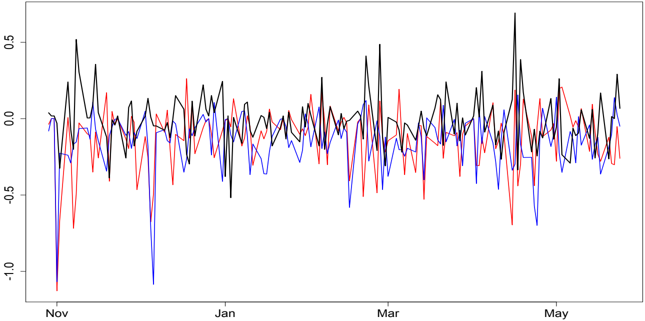


Figure 1: Score functions of first principal component of year 2012-2013. $\boldsymbol{y}_t = \Gamma^\top \boldsymbol{y}_{t-1} + E_t, \quad t = 1, ..., 150$. PCA; 5%; 95%.

	5%	95%	PCA
5%	1	0.25	0.16
95%		1	-0.03
PCA			1

Outline

1. Motivation ✓
2. Algorithm and convergence analysis
3. Oracle inequalities
4. Simulation
5. Application to S&P500 data

Multiple quantile regression

- Factors unobservable and big p or m relative to n?
  Solution: using reduced-rank (quantile) regression
- Rank is unknown?
  Solution: penalization method
- Rank penalization is not convex, NP-hard?
  Solution: matrix nuclear norm (sum of singular values), denoted by $\|\cdot\|_*$, convex envelope of rank (Fazel, 2002)
- Notations: $Y = [\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_n]^\top$, $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n]^\top$ and $E = [E_1, ..., E_n]^\top$

Loss function

$$L(\Gamma) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho(Y_{ij} - X_i^\top \Gamma_j) + \lambda \|\Gamma\|_*, \tag{4}$$

where $\rho(u) = \rho_\tau(u) = |\mathbf{I}(u \leq 0) - \tau||u|$.

- The loss function is akin to those considered by Yuan et al. (2007) and Bunea et al. (2010) in mean regression framework
- $\widehat{\Gamma}_{\lambda,\tau} \stackrel{\text{def}}{=} \arg \min_{\Gamma \in \mathbb{R}^{p \times m}} L(\Gamma)$

Smooth the loss function

Nonsmooth loss function: $\sum_{i=1}^{n}\sum_{j=1}^{m}\rho\left(Y_{ij}-X_i^{\top}\Gamma_j\right)+\lambda\|\Gamma\|_*$

Given weights $\{w_{ij}\}$, $w_{ij}\in\{\tau,1-\tau\}$

$$\sum_{i=1}^{n}\sum_{j=1}^{m}w_{ij}\left|Y_{ij}-X_i^{\top}\Gamma_j\right|$$

Introduce dual variables

$$\max_{\xi_{ij}\in[-1,1]}\sum_{i=1}^{n}\sum_{j=1}^{m}w_{ij}\xi_{ij}\left(Y_{ij}-X_i^{\top}\Gamma_j\right)$$

Smooth

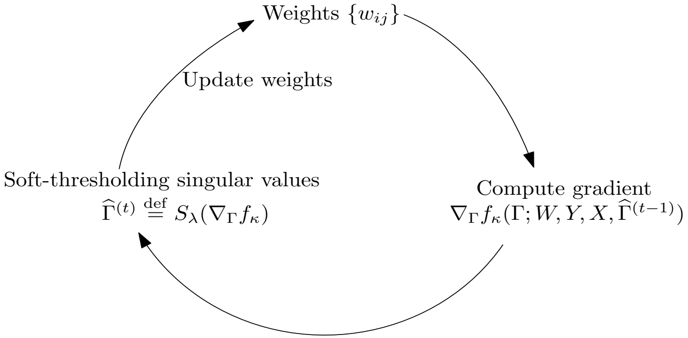
$$f_\kappa(\Gamma;W)=\max_{\xi_{ij}\in[-1,1]}\left\{\ell(\Gamma,\Xi;W)-\tfrac{\kappa}{2}\|\Xi\|_{\mathrm{F}}^2\right\}$$

Lipschitz gradient $\nabla_\Gamma f_\kappa(\Gamma;W)\overset{\text{def}}{=}-\kappa^{-1}X^{\top}(W\odot(Y-X\Gamma)\odot W)$
Lipschitz constant $\leq\kappa^{-1}\|X\|_2^2$

Project on low rank space

Proximity operator: $S_\lambda(Y)\overset{\text{def}}{=}U(\Sigma-\lambda I)_+V^{\top}$

Reweighted Smoothing Proximal Gradient



The process is accelerated by FISTA (Beck and Teboulle, 2009)

Reweighted smoothing proximal gradient algorithm

Input: Y, X, λ, $\kappa = \epsilon/mn$, $L = \frac{mn}{\epsilon}\|X\|^2$;

Initialization: $\Gamma_0 = 0$, $\Omega_1 = 0$, step size $\delta_1 = 1$, $w_{ij}^{(0)} = |\mathbf{I}(Y_{ij} \leq 0) - \tau|$;

for $t = 1, 2, ..., T$ **do**

$\quad \Gamma_t = S_{\lambda/L}\left(\Omega_t - \frac{1}{L} G_\Gamma(\Omega_t; W^{(t-1)})\right)$;

$\quad \delta_{t+1} = \frac{1+\sqrt{1+4\delta_t^2}}{2}$;

$\quad \Omega_{t+1} = \Gamma_t + \frac{\delta_t - 1}{\delta_{t+1}}(\Gamma_t - \Gamma_{t-1})$;

$\quad w_{ij}^{(t)} = |\mathbf{I}(Y_{ij} - X_i^\top \Gamma_j \leq 0) - \tau|$;

end

Output $\widehat{\Gamma} = \Gamma_T$

Convergence analysis

Theorem

Let Γ^* be the optimal solution for minimizing (4) and Γ_t by the approximate solution at t iteration. Set $\kappa = \epsilon / mn$. Then for any $t \geq 1$,

$$|L(\Gamma_t) - L(\Gamma^*)| \leq \frac{\epsilon}{2} + \frac{2mn\|\Gamma_0 - \Gamma^*\|_{\mathrm{F}}^2 \|X\|_2^2}{(t+1)^2 \epsilon}. \tag{5}$$

If we require $L(\Gamma_T) - L(\Gamma^*) \leq \epsilon$. Then the number of iteration T is bounded by

$$2\frac{\sqrt{mn}\|\Gamma^* - \Gamma_0\|_{\mathrm{F}}\|X\|_2}{\epsilon}. \tag{6}$$

Oracle Inequalities for Nuclear Norm Penalized Problems

- Koltchinskii et al. (2011), AoS: trace regression model, matrix completion, minimax lower bound
- Bunea et al. (2011), AoS: Multiple regression model, Rank Selection Criterion (RSC)
- Rohde & Tsybakov (2011), AoS: trace regression model, Schatten-p norm penalty
- Koltchinskii (2013): trace regression model, convex, differentiable empirical loss

Notations

For any $S \in \mathbb{R}^{p\times m}$, $u \in \mathbb{R}$

- Quantile regression: $\rho(u) = \left|\mathbf{I}(u \le 0) - \tau\right||u|$
- Subgradient: $\psi(u) = \mathbf{I}(u \le 0) - \tau$
- Empirical loss: $P_n(\rho \bullet S) \stackrel{\text{def}}{=} (nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho\left(Y_{ij} - \boldsymbol{x}_i^\top S_j\right)$
- Expected loss: $P(\rho \bullet S) \stackrel{\text{def}}{=} m^{-1}\mathrm{E}\left[\sum_{j=1}^{m} \rho\left(Y_{ij} - \boldsymbol{x}_i^\top S_j\right)\right]$
- $\widehat{\Gamma} \stackrel{\text{def}}{=} \min\limits_{S\in\mathbb{D}} \{P_n(\rho \bullet S) + \lambda\|S\|_*\}$
- A natural way to measure the error: excess risk
  $\mathcal{E}(\widehat{\Gamma}) \stackrel{\text{def}}{=} P(\rho \bullet \widehat{\Gamma}) - \inf_{S\in\mathbb{R}^{p\times m}} P(\rho \bullet S)$

Assumptions

1. $Q \geq P_n(\rho \bullet 0)$ for some $Q > 0$
2. $|\boldsymbol{x}^\top S_j| \leq a_j$ a.s. for all $S \in \mathbb{D}$
3. $\nu(a) = \min_{j \leq m} \inf_{|z| < 2a_j} f_{\varepsilon_j}(z|X) > 0$

We define the sparsity concept as below.

- SVD: $A = \sum_{j=1}^{r} \sigma(A) u_j v_j^\top$ with orthogonal vectors $u_1, ..., u_r \in \mathbb{R}^p$ and $v_1, ..., v_r \in \mathbb{R}^m$.
- Support of A: (S_1, S_2) in which $S_1 = \text{span}\{u_1, ..., u_r\}$ and $S_2 = \text{span}\{v_1, ..., v_r\}$.
- Projections P_{S_1} (resp. P_{S_2}): the projection on S_1 (resp. S_2); orthogonal projections: $P_{S_1^\perp}$ (resp. $P_{S_2^\perp}$)
- $\mathcal{P}_A(S) \stackrel{\text{def}}{=} S - P_{S_1^\perp} S P_{S_2^\perp}$; $\mathcal{P}_A^\perp(S) \stackrel{\text{def}}{=} P_{S_1^\perp} S P_{S_2^\perp}$

Assumptions: Restricted Eigenvalue Property

Koltchinskii et al. (2011): Let $\mathbb{D} \subset \mathbb{R}^{p\times m}$: convex closed; $A \in \mathbb{R}^{p\times m}$. $b > 0$. Define the cone

$$\mathcal{K}(\mathbb{D}; A; b) \stackrel{\text{def}}{=} \left\{ S \in \mathbb{D} : \|\mathcal{P}_A^{\perp}(S)\|_* \le b\|\mathcal{P}_A(S)\|_* \right\}. \tag{7}$$

Given a probability distribution Π on $\mathbb{R}^p$ for X, define

$$\beta^{(b)}(\mathbb{D}; A; \Pi) \stackrel{\text{def}}{=} \inf\left\{\beta > 0 : \|\mathcal{P}_A(S)\|_{\mathrm{F}} \le \beta \|S\|_{L_2(\Pi)},\ S \in \mathcal{K}(\mathbb{D}; A; b)\right\}, \tag{8}$$

where $\|S\|_{L_2(\Pi)} \stackrel{\text{def}}{=} m^{-1}\mathsf{E}\|S^\top \boldsymbol{x}\|_2^2$. We focus on $b = 5$.

RE assumption : $\beta(A) \stackrel{\text{def}}{=} \beta^{(5)}(\mathbb{D}; A; \Pi) < \infty$ for some A

Mimicking the Best

Theorem

Let $U_X > 0$ *such that* $\|\boldsymbol{x}/p\|_2 \le U_X$ *a.s.,* $D(p, m, n) = \log(p + m) + 2m \log n$*,* $\sigma_X = (\mathrm{E}\boldsymbol{x}^\top \boldsymbol{x}/p)^{1/2}$ *and*

$$\lambda \ge 16\sqrt{2}\mu(\tau)\left(\frac{\sigma_X\sqrt{p}}{\sqrt{mn}}\sqrt{D(p, m, n)}\bigvee\frac{U_X\sqrt{p}}{\sqrt{mn}}D(p, m, n)\right).$$

Then with probability at least $1 - e^{-t}$*,*

$$\mathcal{E}(\widehat{\Gamma}) \le \inf_{S\in\mathbb{D}}\left[\mathcal{E}(S) + \left\{\frac{4}{\nu(a)}\beta(S)^2 rank(S)\lambda^2 \bigwedge 2\lambda\|S\|_*\right\} + C(a, \tau)\frac{t(S;\lambda)}{n}\right], \tag{9}$$

where $\mu(\tau) = \max\{\tau, 1 - \tau\}$*,* $C(a, \tau) \stackrel{\text{def}}{=} C\left(\frac{\mu(\tau)^2}{\nu(a)} + (\max_j a_j)\mu(\tau)\right)$*,*
$t(S;\lambda) = t + 3\log\left(B\log_2\left(\max_j a_j^{-1} \vee Q \vee \lambda \vee n \vee \|S\|_* \vee 2\right)\right)$*.*

An Oracle Inequality for MQR

Theorem

With probability at least $1 - 1/n$,

$$P(\rho \bullet \widehat{\Gamma}) - P(\rho \bullet \Gamma)$$

$$\leq C_* \underbrace{\frac{\mu(\tau)^2}{\nu(a)}}_{\text{Error}} \underbrace{\{U_X^2 \vee (\mathrm{E}\boldsymbol{x}^\top \boldsymbol{x}/p)\}}_{\text{Design}} \frac{\overbrace{rank(\Gamma)m \vee p}^{\text{oracle number of parameters}}}{n} + c_* \frac{\log n}{n}, \quad (10)$$

where c_* *is a constant and* C_* *depends on* $\beta(\Gamma)$. $\mu(\tau) = \tau \vee (1 - \tau)$, $U_X \geq \|\boldsymbol{x}/p\|_2$ *a.s.*

- The performance depends on the oracle number of parameters rather than the actual number of unknowns pm
- $pm > n$ is fine. However, $\max\{p, m\} < n$ is inevitable

Simulation

Parameters: $p = m = 500$. $n = 500, 800$. No. of iteration $I = 300$
Data generation:

1. $\Gamma \in \mathbb{R}^{p\times m}$: each entry follows i.i.d. normal distribution; truncate the last 375 singular values (i.e. keep only the first 125 singular values)
2. $\boldsymbol{x}_i \in \mathbb{R}^p$: i.i.d. generated from multi. Gaussian distribution with mean 0 and correlation $0.5^{|i-j|}$
3. $E \in \mathbb{R}^{n\times m}$: i.i.d. standard normal distribution. Response variable is then generated from

$$Y = X\Gamma + E. \tag{11}$$

Table 2: Average Frobenius loss $\|\widehat{\Gamma} - \Gamma\|_{\mathrm{F}}$. The "()" indicate the standard deviation.

λ	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$	$\tau = 0.9$	$\tau = 0.95$
50	284.35	153.78	122.97	108.85	122.97	153.81	284.30
	(3.33)	(1.90)	(1.27)	(0.88)	(1.25)	(1.9)	(3.29)
100	285.48	194.08	151.23	123.52	151.21	194.08	285.44
	(3.29)	(4.73)	(1.76)	(1.37)	(1.71)	(4.70)	(3.25)
150	286.42	273.88	175.82	148.76	175.78	273.84	286.39
	(3.26)	(3.60)	(1.95)	(1.62)	(1.93)	(3.58)	(3.23)
200	287.64	277.54	194.36	187.53	194.38	277.50	287.60
	(3.24)	(3.48)	(2.34)	(1.92)	(2.21)	(3.47)	(3.19)
50	82.31	53.15	41.96	37.30	41.97	53.13	82.33
	(1.42)	(0.88)	(0.26)	(0.15)	(0.27)	(0.89)	(1.44)
100	180.64	72.42	44.83	37.32	44.79	72.33	181.70
	(45.57)	(1.49)	(0.83)	(0.15)	(0.9)	(1.53)	(45.21)
150	212.57	89.32	52.57	37.32	52.54	89.27	212.56
	(5.23)	(2.25)	(1.2)	(0.15)	(1.25)	(2.25)	(5.23)
200	214.95	102.15	62.87	46.09	62.76	102.23	214.98
	(5.17)	(3.09)	(1.72)	(2.79)	(1.76)	(3.05)	(5.19)

Table 3: Average maximal loss $\|\widehat{\Gamma} - \Gamma\|_{max}$. The "()" indicate the standard deviation.

λ	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$	$\tau = 0.9$	$\tau = 0.95$
50	2.76	1.52	1.23	1.08	1.23	1.53	2.77
	(0.16)	(0.09)	(0.07)	(0.07)	(0.07)	(0.09)	(0.16)
100	2.77	1.90	1.50	1.23	1.51	1.91	2.78
	(0.16)	(0.11)	(0.09)	(0.07)	(0.09)	(0.13)	(0.16)
150	2.78	2.66	1.72	1.48	1.74	2.66	2.78
	(0.16)	(0.15)	(0.1)	(0.09)	(0.11)	(0.15)	(0.16)
200	2.79	2.69	1.90	1.83	1.92	2.70	2.79
	(0.16)	(0.15)	(0.11)	(0.1)	(0.12)	(0.15)	(0.16)
50	0.82	0.51	0.40	0.36	0.40	0.52	0.82
	(0.05)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.06)
100	1.77	0.72	0.43	0.36	0.43	0.73	1.78
	(0.45)	(0.05)	(0.02)	(0.02)	(0.03)	(0.05)	(0.45)
150	2.08	0.89	0.52	0.36	0.52	0.89	2.08
	(0.13)	(0.06)	(0.03)	(0.02)	(0.04)	(0.06)	(0.14)
200	2.10	1.02	0.63	0.45	0.63	1.02	2.10
	(0.13)	(0.07)	(0.04)	(0.04)	(0.05)	(0.07)	(0.14)

Table 4: Average number of factor selected. The "()" indicate the standard deviation.

λ	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$	$\tau = 0.9$	$\tau = 0.95$
50	1.00	126.00	125.00	125.00	125.00	126.00	1.00
	(0)	(0.06)	(0)	(0)	(0)	(0)	(0)
100	1.00	1.00	126.00	125.00	126.00	1.00	1.00
	(0)	(0)	(0.06)	(0)	(0)	(0)	(0)
150	1.00	1.00	88.08	125.00	83.08	1.00	1.00
	(0)	(0)	(57.46)	(0)	(59.35)	(0)	(0)
200	1.00	1.00	1.00	125.00	1.00	1.00	1.00
	(0)	(0)	(0)	(0)	(0)	(0)	(0)
50	125.08	125.00	125.00	125.00	125.00	125.00	125.08
	(0.27)	(0)	(0)	(0)	(0)	(0)	(0.27)
100	23.50	125.00	125.00	125.00	125.00	125.00	25.17
	(48.02)	(0.06)	(0)	(0)	(0)	(0.06)	(49.36)
150	1.00	125.99	125.00	125.00	125.00	125.99	1.00
	(0)	(0.1)	(0)	(0)	(0)	(0.08)	(0)
200	1.00	76.00	125.00	125.00	125.00	78.50	1.00
	(0)	(61.24)	(0)	(0)	(0)	(60.67)	(0)

Table 5: Average l_2 loss $\|sv(\widehat{\Gamma}) - sv(\Gamma)\|_2$. The "()" indicate the standard deviation.

λ	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$	$\tau = 0.9$	$\tau = 0.95$
50	183.46	61.48	49.34	56.02	49.35	61.50	183.43
	(3.76)	(1.28)	(0.32)	(0.78)	(0.32)	(1.3)	(3.76)
100	185.02	92.12	59.51	45.55	59.49	92.17	185.00
	(3.73)	(4.6)	(1.08)	(0.32)	(1.06)	(4.59)	(3.74)
150	186.39	172.22	75.39	52.18	75.38	172.20	186.38
	(3.73)	(4.08)	(1.55)	(0.85)	(1.55)	(4.09)	(3.76)
200	188.23	176.59	90.16	81.00	90.21	176.57	188.20
	(3.75)	(4.02)	(2.17)	(1.79)	(2.1)	(4.04)	(3.75)
50	40.78	32.56	31.30	27.81	31.31	32.57	40.80
	(0.68)	(0.27)	(0.21)	(0.12)	(0.22)	(0.27)	(0.69)
100	105.06	37.07	30.92	27.81	30.93	37.04	105.81
	(32.62)	(0.61)	(0.25)	(0.12)	(0.27)	(0.6)	(32.34)
150	127.96	44.30	30.98	27.81	30.98	44.27	127.94
	(4.47)	(0.97)	(0.25)	(0.12)	(0.25)	(0.97)	(4.44)
200	129.83	49.86	33.44	26.71	33.40	49.90	129.83
	(4.43)	(1.4)	(0.54)	(0.42)	(0.53)	(1.38)	(4.42)

Table 6: Average l_∞ loss $\|sv(\widehat{\Gamma})-sv(\Gamma)\|_\infty$. The "()" indicate the standard deviation.

λ	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$	$\tau = 0.9$	$\tau = 0.95$
50	72.67	16.02	13.69	8.66	13.70	16.01	72.68
	(1.68)	(0.83)	(0.63)	(0.25)	(0.64)	(0.85)	(1.74)
100	71.39	15.95	16.62	6.22	16.58	15.95	71.39
	(1.67)	(2.82)	(0.85)	(0.36)	(0.87)	(2.82)	(1.72)
150	69.97	64.28	17.29	8.67	17.29	64.29	69.95
	(1.63)	(1.72)	(0.39)	(0.36)	(0.38)	(1.76)	(1.66)
200	67.80	63.69	15.82	12.53	15.81	63.69	67.80
	(1.57)	(1.56)	(0.4)	(0.38)	(0.39)	(1.6)	(1.64)
50	17.03	14.44	10.52	4.28	10.54	14.45	17.02
	(0.79)	(0.63)	(0.49)	(0.12)	(0.52)	(0.63)	(0.77)
100	54.15	16.19	9.43	4.32	9.45	16.16	54.64
	(21.83)	(0.86)	(0.53)	(0.12)	(0.56)	(0.83)	(21.61)
150	67.50	20.10	10.57	4.33	10.56	20.07	67.50
	(3.24)	(0.82)	(0.65)	(0.12)	(0.63)	(0.79)	(3.19)
200	66.96	21.36	14.05	3.51	14.00	21.34	66.95
	(3.1)	(0.63)	(0.91)	(0.25)	(0.87)	(0.59)	(3.05)

Table 7: Average computational time. The "()" indicate the standard deviation.

λ	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$	$\tau = 0.9$	$\tau = 0.95$
50	66.34	461.84	436.71	432.81	435.68	459.48	67.58
	(0.65)	(1.72)	(1.15)	(1.54)	(0.74)	(1.1)	(0.00)
100	60.14	262.10	239.28	218.30	238.63	261.18	60.60
	(0.49)	(0.20)	(1.43)	(1.43)	(0.28)	(1.71)	(0.56)
150	55.81	63.71	166.31	131.20	166.34	63.29	55.68
	(0.12)	(0.09)	(0.96)	(0.31)	(0.73)	(0.55)	(0.21)
200	50.36	55.08	129.96	75.10	128.59	55.65	50.33
	(0.39)	(0.15)	(0.87)	(0.05)	(0.01)	(0.31)	(0.06)
50	892.32	682.96	1196.53	634.90	1198.71	686.44	894.87
	(8.77)	(13.39)	(21.75)	(8.05)	(19.08)	(14.56)	(15.52)
100	276.28	484.50	440.60	612.84	449.42	485.61	271.60
	(6.44)	(8.91)	(6.41)	(9.03)	(7.15)	(8.92)	(5.98)
150	111.58	398.30	318.33	603.94	317.57	397.17	111.09
	(1.63)	(6.70)	(5.62)	(8.86)	(5.64)	(7.33)	(1.83)
200	101.07	348.28	269.71	227.06	270.91	350.86	101.30
	(1.69)	(5.55)	(5.59)	(1.28)	(4.00)	(5.29)	(1.80)

S&P500 component distributional analysis

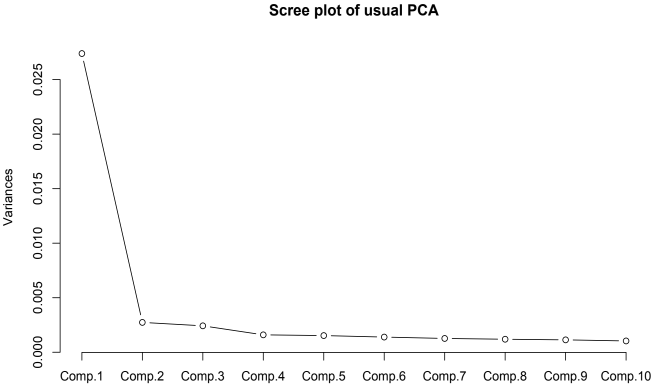
- Data: Daily returns for the components of S&P500 from Oct. 26, 2012-Sep. 24, 2014
- Data sourse: Datastream "constituent list"
- 495 firms are under consideration, due to missing data
- The following model is considered:

$$\boldsymbol{y}_t = \Gamma^\top \boldsymbol{y}_{t-1} + E_t, t = 1, ..., T = 499.$$

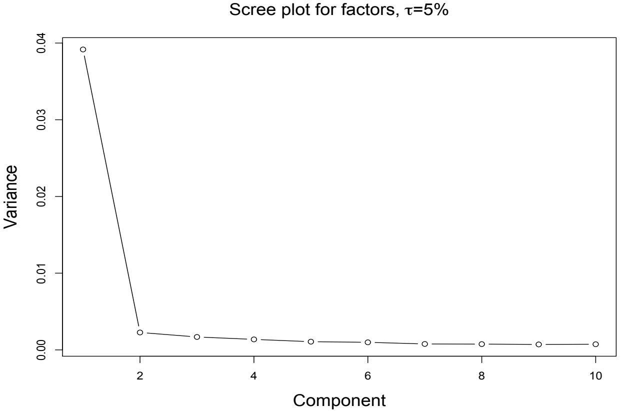
  $\boldsymbol{y}_t$: log return of the 495 firms. Γ is 495×495

- $\lambda = 0.31$, obtained by oracle result times 0.0005

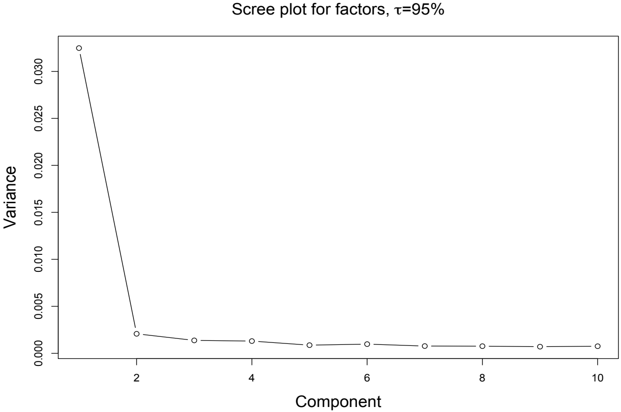
PC	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6
Prop. of var.	.279	.028	.025	.016	.016	.014
Cum. prop.	.279	.307	.332	.348	.364	.378



PC	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6
Prop. of var.	.543	.025	.014	.013	.012	.010
Cum. prop.	.543	.568	.582	.595	.607	.617



PC	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6
Prop. of var.	.584	.017	.014	.012	.010	.009
Cum. prop.	.584	.601	.615	.627	.637	.646



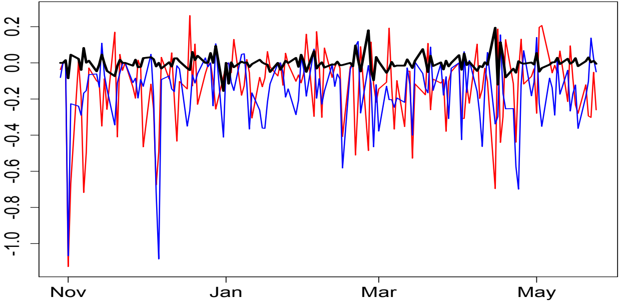


Figure 2: Score functions of first principal component of year 2012-2013. $\boldsymbol{y}_t = \Gamma^\top \boldsymbol{y}_{t-1} + E_t, \quad t = 1, ..., 150$. **VIX returns**; 5%; 95%.

	5%	95%	*VIX*
5%	1	0.25	-0.43
95%		1	0.35
VIX			1

Nuclear Norm Penalized Large Multiple Quantile Regression

Shih-Kang Chao
Wolfgang Karl Härdle
Ming Yuan

Department of Statistics
University of Wisconsin-Madison
Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
https://www.stat.wisc.edu

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

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

Han F. and Liu H.
A Direct Estimation of High Dimensional Stationary Vector Autoregressions
arXiv: 1307.0293v2 (2013)

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.