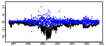
Quantile Regression with high dimensional Single-Index Models

Wolfgang Karl Härdle
Weining Wang, Lixing Zhu
Yan Fan, Lining Yu

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

Financial risk

Figure 1: Financial risk

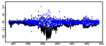


Objective

- Risk patterns depend on covariates X
- Dimensionality issues, $X \in \mathbb{R}^p, p \to \infty$
- Variable selection for Quantile Regression (QR)
- CoVaR, single index model (SIM)

Challenges

- Model of tails of conditional distribution
- Dimension reduction
- SIM estimation combined with variable selection
- Alternatives to MAVE (minimum average variance estimation)



Single Index Model

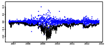
- Observations $\{X_i, Y_i\}_{i=1}^n$ with

$$Y_i = g(\beta^{*\top} X_i) + \varepsilon_i, \tag{1}$$

  where $g(\cdot)$ is the link function, and $\beta^* \in \mathbb{R}^p$. $\{\varepsilon_i\}_{i=1}^n$ are independent.
- p is possible large: $p \to \infty$.
- $\mathsf{E}_{Y|X=x}(\varepsilon) = 0$ for mean regression.
- $F^{-1}_{\varepsilon|X=x}(\tau) = 0$ for quantile regression.

What is known?

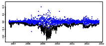
- MAVE method, Xia et al.(2002)
- Application in banking, environmental statistics
- First order "free lunch" $\sqrt{n}$ rate
- A one dimensional problem for estimating $g(\cdot)$



High dimensional SIM

- How to estimate nonzero β_j ?
- Which rates can we allow for p ?
- What are the consequences for estimating $g(\cdot)$?
- Sparsistency ?

Go to details



Outline

1. Motivation ✓
2. Single index model
3. Simulations
4. Applications
5. Further work

A quasi-likelihood approach

Recall (1):

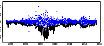
$$\min_{\beta} \mathsf{E}_{(Y|X=x)} \rho_\tau\{Y - g(\beta^\top x)\}, \tag{2}$$

Quantile regression:

$$\rho_\tau(u) = \tau u \mathbf{1}\{u \in (0,\infty)\} - (1-\tau) u \mathbf{1}\{u \in (-\infty, 0)\}. \tag{3}$$

Expectile regression:

$$\rho_\tau(u) = \tau u^2 \mathbf{1}\{u \in (0,\infty)\} + (1-\tau) u^2 \mathbf{1}\{u \in (-\infty, 0)\}. \tag{4}$$



Likelihood approximations

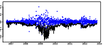
$$g(\beta^\top X_i) \approx g(\beta^\top x) + g'(\beta^\top x)\beta^\top (X_i - x). \tag{5}$$

Approximations

$$L_x(\beta) \stackrel{\text{def}}{=} \mathsf{E}\,\rho_\tau\{Y - g(\beta^\top x) - g'(\beta^\top x)\beta^\top (X - x)\} K_h\{\beta^\top (X - x)\} \tag{6}$$

$$L_{n,x}(\beta) \stackrel{\text{def}}{=} n^{-1}\sum_{i=1}^{n} \rho_\tau\{Y_i - g(\beta^\top x) - g'(\beta^\top x)\beta^\top (X_i - x)\} K_h\{\beta^\top (X_i - x)\} \tag{7}$$

where $K_h(.) = K(./h)/h$ with $K(.)$ a kernel function and h a bandwidth.



A simple trick

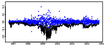
Minimize average contrast (w.r.t. β):

$$L_n(\beta) \stackrel{\text{def}}{=} \sum_{j=1}^{n} L_{n,X_j}(\beta)$$

$$= \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\rho_\tau\left\{Y_i - g(\beta^\top X_j) - g'(\beta^\top X_j)\beta^\top(X_i - X_j)\right\} K_h\{\beta^\top(X_i - X_j)\}. \quad (8)$$

Therefore (in first approach):

$$\widehat{\beta} \approx \arg\min_{\beta} L_n(\beta).$$

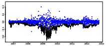


One more trick

Let $a_j = g(\beta^\top X_j)$, $b_j = g'(\beta^\top X_j)$, estimate β by:

$$\min_{(a_j,b_j)'\mathrm{s},\beta} \sum_{j=1}^{n} \sum_{i=1}^{n} \rho_\tau \big(Y_i - a_j - b_j X_{ij}^\top \beta\big) \omega_{ij}(\beta), \tag{9}$$

where $X_{ij} \stackrel{\mathrm{def}}{=} X_i - X_j$, $\omega_{ij}(\beta) \stackrel{\mathrm{def}}{=} K_h(X_{ij}^\top \beta) / \sum_{l=1}^{n} K_h(X_{lj}^\top \beta)$.



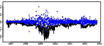
The final trick

Penalize the dimension p and estimate β by:

$$\min_{(a_j,b_j)'\mathrm{s},\beta} \sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau\left(Y_i - a_j - b_j X_{ij}^\top \beta\right)\omega_{ij}(\beta) + \sum_{l=1}^{p} \gamma_\lambda(|\beta_l^{(0)}|)|\beta_l|, \quad (10)$$

where $\gamma_\lambda(t)$ is some non-negative function.

▸ Go to details



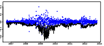
How does this work?

- $\widehat{\beta}^{(0)}$ initial estimator of β^* (linear QR with variable selection).
- For $t = 0, 1, 2, \cdots$, given $\widehat{\beta}^{(t)}$, standardize $\widehat{\beta}^{(t)}$, $\|\widehat{\beta}^{(t)}\| = 1$, $\widehat{\beta}_1^{(t)} = 1$, $d_l \stackrel{\text{def}}{=} \gamma_\lambda(|\widehat{\beta}_l^{(t)}|)$. Then compute

$$(\widehat{a}_j^{(t)}, \widehat{b}_j^{(t)}) \stackrel{\text{def}}{=} \operatorname*{arg\,min}_{(a_j, b_j)'\text{s}} \sum_{i=1}^{n} \rho_\tau\big(Y_i - a_j - b_j X_{ij}^\top \widehat{\beta}^{(t)}\big)\omega_{ij}(\widehat{\beta}^{(t)})$$

- Given $(\widehat{a}_j^{(t)}, \widehat{b}_j^{(t)})$, solve

$$\widehat{\beta}^{(t+1)} = \operatorname*{arg\,min}_{\beta} \sum_{j=1}^{n} \sum_{i=1}^{n} \rho_\tau\big(Y_i - \widehat{a}_j^{(t)} - \widehat{b}_j^{(t)} X_{ij}^\top \beta\big)\omega_{ij}(\widehat{\beta}^{(t)}) + \textstyle\sum_{l=1}^{p} d_l |\beta_l|.$$

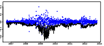


Some definitions

Let $\beta^* = (\beta^{*\top}_{(1)}, \beta^{*\top}_{(0)})^\top$ with $\beta^*_{(1)} \stackrel{\text{def}}{=} (\beta_1, \ldots, \beta_q)^\top \neq 0$ and $\beta^*_{(0)} = (\beta_{q+1}, \ldots, \beta_p)^\top = 0$. $X_{i(1)} \stackrel{\text{def}}{=}$ sub vector of X_i corresponding to $\beta^*_{(1)}$, $X_{i(0)}$ corresponding to $\beta^*_{(0)}$.

$$\widehat{\beta}^0 \stackrel{\text{def}}{=} \arg\min_{\beta_{(1)}} \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} \rho_\tau \left\{ Y_i - a_{j(1)} - b_{j(1)} X^\top_{ij(1)} \beta_{(1)} \right\} K_h\{\beta^\top_{(1)}(X_{i(1)} - X_{j(1)})\}.$$

where $a_j = g(\beta^\top_{(1)} X_{j(1)})$, $b_j = g'(\beta^\top_{(1)} X_{j(1)})$, $X_{ij(1)} = X_{i(1)} - X_{j(1)}$, $Z_i = X_i^\top \beta^*$.



An amuse gueule of theory

Denote $\widehat{\beta}$ as the final estimate of β^*.

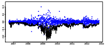
Theorem

Under A 1-5, the estimators $\widehat{\beta}^0$ and $\widehat{\beta}$ exist and $\mathrm{P}(\widehat{\beta}^0 = \widehat{\beta}) \to 1$. Moreover,

$$\mathrm{P}(\widehat{\beta}^0 = \widehat{\beta}) \geq 1 - (p - q)\exp(-C' n^{\alpha}) \tag{11}$$

for $0 < \alpha < 1/2$.

Go to details



Antipasti Theory

Theorem

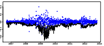
Let $0 < \alpha < \frac{\alpha_2}{2} < \frac{1}{2}$, $\frac{\alpha_2}{2} < \alpha_1$, $D_n = o(n^{\alpha_1 - \alpha_2/2})$, $q = \mathcal{O}(n^{\alpha_2})$. $\| \sum_i \sum_j X_{ij(1)} \omega_{ij(1)} X_{ij(0)} \|_{2,\infty} = \mathcal{O}(n^{1-\alpha_1})$, $\lambda = \mathcal{O}\left(\sqrt{q/n}\right)$. *Then*

$$\|\widehat{\beta}_{(1)} - \beta^*_{(1)}\| = \mathcal{O}_p\{(\lambda D_n + n^{-1/2})\sqrt{q}\} \tag{12}$$

For any unit vector $\boldsymbol{b}$ *in* $\mathbb{R}^q$, *we have*

$$\boldsymbol{b}^\top C_{0(1)}^{-1} \sqrt{n} (\widehat{\beta}_{(1)} - \beta^*_{(1)}) \xrightarrow{\mathcal{L}} N(0,\, \sigma_\tau^2) \tag{13}$$

Go to details



Antipasti Theory

where $\widehat{\beta}_{(1)} \stackrel{\text{def}}{=} \big(\widehat{\beta}_j\big)_{j\in\mathcal{M}^*}$, $\mathcal{M}_* = \{j : \beta_j^* \neq 0\}$ be the true model.

$C_{0(1)} \stackrel{\text{def}}{=} \mathsf{E}\{[g'(Z_i)]^2[\mathsf{E}(X_{(1)}|Z_i) - X_{i(1)}][\mathsf{E}(X_{(1)}|Z_i) - X_{i(1)}]^\top\}$,

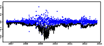
$\psi_\tau(\varepsilon)$ is a selection of the subgradient of $\rho_\tau(\varepsilon)$ and

$\sigma_\tau^2 = \mathsf{E}[\psi_\tau(\varepsilon_i)]^2/[\partial^2 \mathsf{E}\,\rho_\tau(\varepsilon_i)]^2$,

where

$$\partial^2 \mathsf{E}\,\rho_\tau(\cdot) = \frac{\partial^2 \mathsf{E}\,\rho_\tau(\varepsilon_i - v)^2}{\partial v^2}\bigg|_{v=0} \tag{14}$$

Go to details



Main Course

Theorem

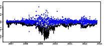
Under A 1-5, the event $\mathcal{B}_n \stackrel{\text{def}}{=} \{\widehat{\beta} = \beta^*\}$*.* $\mathrm{P}(\mathcal{B}_n) \to 1$ *and* $\mathrm{P}(\mathcal{B}_n^c) \to 0$*. Let* $\mu_j \stackrel{\text{def}}{=} \int u^j K(u)du$ *and* $\nu_j \stackrel{\text{def}}{=} \int u^j K^2(u)du$*,* $j = 0, 1, \ldots$*. If* $nh^3 \to \infty$ *and* $h \to 0$*, we have*

$$\sqrt{nh}\left\{\widehat{g}(x^\top\widehat{\beta}) - g(x^\top\beta^*) - \frac{1}{2}h^{5/2}g''(x^\top\beta^*)\mu_2[\partial\,\mathsf{E}\,\psi_\tau(\varepsilon)]\right\}$$

$$\xrightarrow{\mathcal{L}} N\left(0,\ \nu_0\sigma_\tau^2/f_Z(z)\right).$$

and

$$\sqrt{nh^3}\left\{\widehat{g}'(x^\top\widehat{\beta}) - g'(x^\top\beta^*)\right\} \xrightarrow{\mathcal{L}} N\left(0,\ \nu_2\sigma_\tau^2/[f_Z(z)\mu_2^2]\right).$$



Link functions

Consider the link functions $g(\cdot)$:

- Model 1

$$Y_i = 5\cos(D \cdot u_i) + \exp(-D \cdot u_i^2) + \varepsilon_i, \tag{15}$$

where u_i is the index: $u_i = X_i^\top \beta$, $D = 0.8$ is a scaling constant and ε_i is the error term.

- Model 2

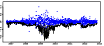
$$Y_i = \sin\{\pi(a \cdot u_i - b)\} + \varepsilon_i, \tag{16}$$

with the parameters $a = 0.1$, $b = 0.4$.

- Model 3

$$Y_i = 10\sin(D \cdot u_i) + \sqrt{|\sin(u_i) + \varepsilon_i|} \tag{17}$$

with $D = 0.1$.



Criteria

1. Standardized $L2$ norm:

$$Dev \stackrel{\mathrm{def}}{=} \sum_{i=1}^{p} \frac{\|\beta_i - \widehat{\beta}_i\|_2}{\|\beta\|_2}$$

2. Sign consistency:

$$Acc \stackrel{\mathrm{def}}{=} \sum_{i=1}^{p} |sign(\beta_i) - sign(\widehat{\beta}_i)|;$$

Criteria

3. Least angle:

$$\textit{Angle} \stackrel{\text{def}}{=} \frac{< \beta, \widehat{\beta} >}{\|\beta\|_2 \cdot \|\widehat{\beta}\|_2}$$

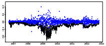
4. Relative error:

$$\textit{Error} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} \left| \frac{g(X_i^\top \beta^*) - \widehat{g}(X_i^\top \beta^*)}{g(X_i^\top \beta^*)} \right|$$

Criteria - quantile regression

$g(\cdot)$	τ	*Dev*	*Acc*	*Angle*	*Error*
Model 1	0.9	2.3 (0.3)	0.66 (0.40)	0.999 (0.1)	0.96 (0.1)
	0.5	3.7 (4.9)	0.04 (0.03)	0.998 (0.2)	0.15 (0.1)
Model 2	0.9	2.8 (5.6)	0.13 (0.81)	0.997 (0.1)	8.16 (1.1)
	0.5	8.2 (6.4)	0.02 (0.13)	0.995 (0.1)	7.51 (4.7)
Model 3	0.9	3.2 (5.9)	0.20 (0.92)	0.997 (0.2)	11.50 (7.9)
	0.5	1.1 (0.8)	0.07 (0.26)	0.986 (0.1)	5.34 (1.6)

Table 1: Criteria evaluated under different models and quantiles. The error ε follows a $\mathrm{N}(0, 0.1)$ distribution. In 100 simulations we set $n = 100, p = 10, q = 2$. *Dev* and *Error* are reported in 10^{-2}. Standard deviations are given in brackets.



Criteria - quantile regression

$g(\cdot)$	*Dev*	*Acc*	*Angle*	*Error*
Model 1	6.5 (8.0)	3.5 (0.2)	0.934 (0.1)	1.6 (0.5)
Model 2	5.2 (11.1)	2.8 (0.6)	0.933 (0.1)	2.1 (5.6)
Model 3	4.1 (5.9)	0.6 (0.8)	0.992 (0.2)	2.0 (9.1)

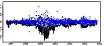
Table 2: Criteria evaluated under different models. The error ε follows a $\mathrm{N}(0, 0.1)$ distribution. In 100 simulations we set $n = 100, p = 120, q = 9, \tau = 0.9$. *Dev* and *Error* are reported in 10^{-2}. Standard deviations are given in brackets.

Value at Risk

- Value-at-Risk (VaR) is the most known measure for quantifying and controlling the risk of a portfolio.
- The VaR of a financial institution i at $\tau \in (0,1)$:

$$\mathrm{P}(X_{i,t} \leq VaR^{\tau}_{i,t}) \stackrel{def}{=} \tau,$$

where $X_{i,t}$ represents the asset return of financial institution i at time t.

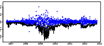


CoVaR

- Adrian and Brunnermeier (AB) (2011) proposed CoVaR.
- The CoVaR of a risk factor j given X_i at level $\tau \in (0,1)$:

$$\mathrm{P}\{X_{j,t} \le CoVaR^{\tau}_{j|i,t} | X_{i,t} = VaR^{\tau}(X_{i,t}), M_{t-1}\} \stackrel{def}{=} \tau,$$

here M_{t-1} is a vector of macro prudential variables.



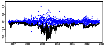
Quantile regression

- CoVaR technique (AB).
- Two linear quantile regressions:

$$
\begin{aligned}
X_{i,t} &= \alpha_i + \gamma_i^\top M_{t-1} + \varepsilon_{i,t}, && (18)\\
X_{j,t} &= \alpha_{j|i} + \beta_{j|i} X_{i,t} + \gamma_{j|i}^\top M_{t-1} + \varepsilon_{j,t}. && (19)
\end{aligned}
$$

- $F^{-1}_{\varepsilon_{i,t}}(\tau|M_{t-1}) = 0$ and $F^{-1}_{\varepsilon_{j,t}}(\tau|M_{t-1}, X_{i,t}) = 0$, then:

$$
\begin{aligned}
\widehat{VaR}^{\tau}_{i,t} &= \widehat{\alpha}_i + \widehat{\gamma}_i^\top M_{t-1}, && (20)\\
\widehat{CoVaR}^{\tau}_{j|i,t} &= \widehat{\alpha}_{j|i} + \widehat{\beta}_{j|i} \widehat{VaR}^{\tau}_{i,t} + \widehat{\gamma}_{j|i}^\top M_{t-1}. && (21)
\end{aligned}
$$



Quantile regression and SIM

- Generalize (19):

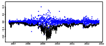
$$X_{j,t} = g(S^\top \beta_{j|S}) + \varepsilon_{j,t}, \tag{22}$$

where $S \stackrel{def}{=} [M_{t-1}, R]$, R is a vector of log returns. $\beta_{j|S}$ is a $p \times 1$ vector, p large.

- $F^{-1}_{\varepsilon_{j,t}}(\tau|S) = 0$, then:

$$\widehat{CoVaR}^{\tau}_{j|\widehat{S}} = \widehat{g}(\widehat{S}^\top \widehat{\beta}_{j|S}), \tag{23}$$

where $\widehat{S} \stackrel{def}{=} [M_{t-1}, \widehat{V}]$, where $\widehat{V}$ is the estimated VaR in (20).



Model classification

- A general model:

$$Y = f(X) + \varepsilon.$$

- An additive structure:

$$f(X) = \sum_{l=1}^{L} g_l(\beta_l^\top x_l).$$

- $L = 1$ is SIM model:

$$f(X) = g(\beta^\top X), \quad \text{where} \quad X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

and $x_1 \in \mathbb{R}^q$, $x_2 \in \mathbb{R}^{p-q}$.

Model classification

- $L = 2$ is partial linear model (PLM)

$$\begin{aligned} f(X) &= g_1(\beta_1^\top x_1) + g_2(\beta_2^\top x_2) \\ &= \beta_1^\top x_1 + g_2(x_2), \end{aligned}$$

for $g_1 = id$, $\beta_1 \in \mathbb{R}^{p-1}$, $\beta_2 = 1$, where

$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

and $x_1 \in \mathbb{R}^{p-1}$, $x_2 \in \mathbb{R}^1$.

Dataset

- City national corp (CYN) (ranked as a small firm).
- Choose 20 firms ranked higher than CYN and 7 macro prudential variables.
- Time period is from January 5, 2006 to September 6, 2012, $T = 1670$.

Descriptive statistics of CYN

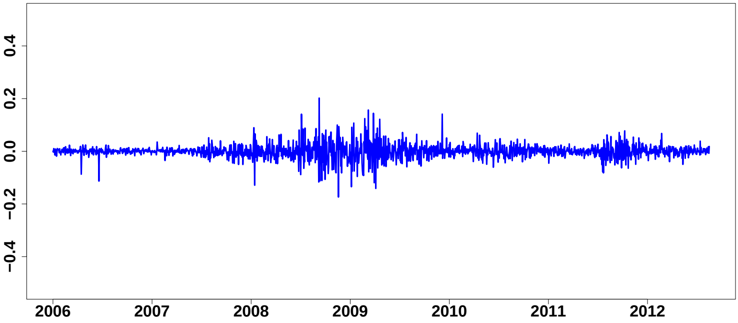
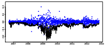


Figure 2: Log returns of CYN

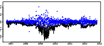


Descriptive statistics of CYN

	Mean	SD	Skewness	Kurtosis
Before crisis	-0.0004	0.0209	0.2408	12.1977
In crisis	-9.247×10^{-5}	0.0312	0.1326	8.9544

Table 3: Descriptive statistics

- Jarque Bera Test is performed: log returns of CYN are not normally distributed.
- Unit root test is conducted: log returns of CYN are stationary.



Estimation of VaR

- Window size: $n = 100$.
- 7 Macro prudential variables are applied.
- Method: quantile regression.
- $\tau = 0.05$.
- $T = 1570$ estimated VaR by moving window estimation.

Estimation of VaR

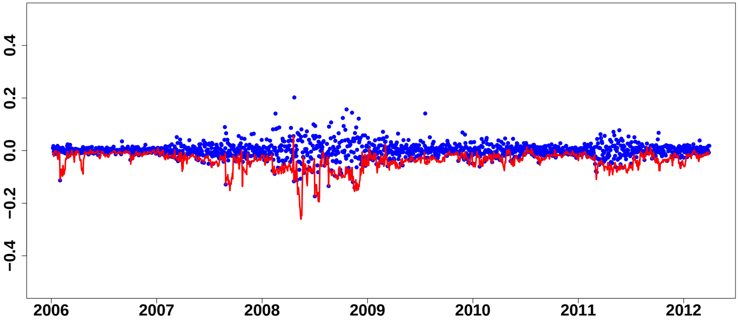


Figure 3: Log returns of CYN (blue) and VaR of log returns of CYN (red), $\tau = 0.05$, $T = 1570$, window size $n = 100$, refer to (20).

Estimation of VaR

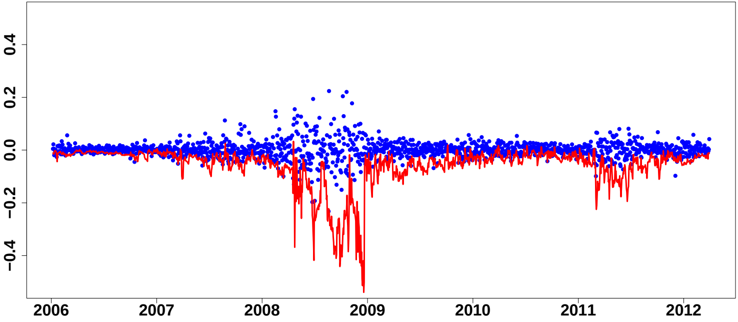
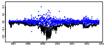


Figure 4: Log returns of JPM (blue) and VaR of log returns of JPM (red), $\tau = 0.05$, $T = 1570$, window size $n = 100$, refer to (20).

Estimation of CoVaR

- Window size: $n = 126$.
- Original variables: $p = 27$.
- Method: $L1$-norm quantile regression.
- $\tau = 0.05$.
- Bandwidth: $h_\tau = h_{mean}\left[\tau(1-\tau)\varphi\{\Phi^{-1}(\tau)\}^{-2}\right]^{0.2}$.
- Where h_{mean}: use direct plug-in methodology of a local linear regression described by Ruppert, Sheather and Wand (1995).
- Survived variables: Different $\widehat{q}$ in each window.
- $T = 1544$ estimated CoVaR by moving window estimation.



Estimation of CoVaR

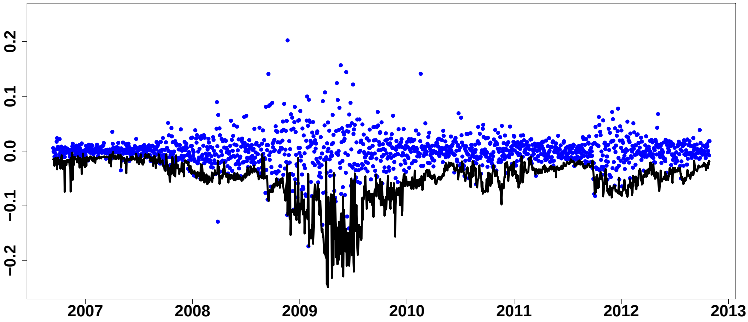


Figure 5: Log returns of CYN (blue) and the estimated CoVaR (black), $\tau = 0.05$, $T = 1544$, window size $n = 126$, refer to (23).

The link function

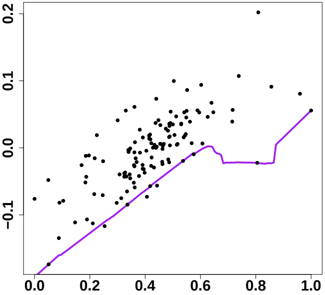
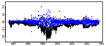


Figure 6: The link functions



The link function

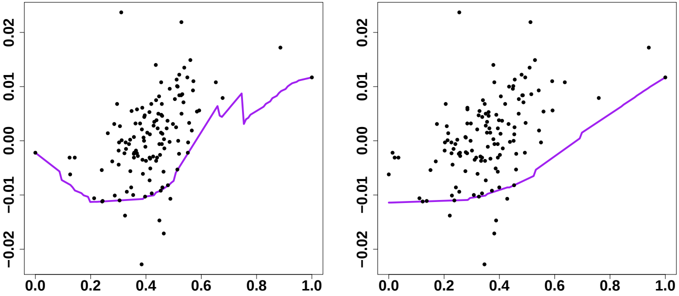


Figure 7: The estimated link function, window size $n = 126$, starting date: 20061227, $\tau = 0.05$, $h(\mathit{left}) = 0.0176$, $h(\mathit{right}) = 0.02$, $p = 27$, $\widehat{q} = 9$: JPM, WFC, BK, FITB, CMA, liquidity, 3MT, yield and S&P.

The link function

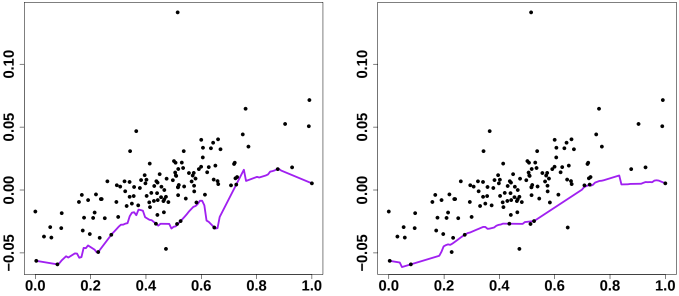


Figure 8: The estimated link function, window size $n = 126$, starting date: 20100126, $\tau = 0.05$, $h(left) = 0.0118$, $h(right) = 0.02$, $p = 27$, $\widehat{q} = 5$: C, STT, RF, FITB, ZION.

The influential variables

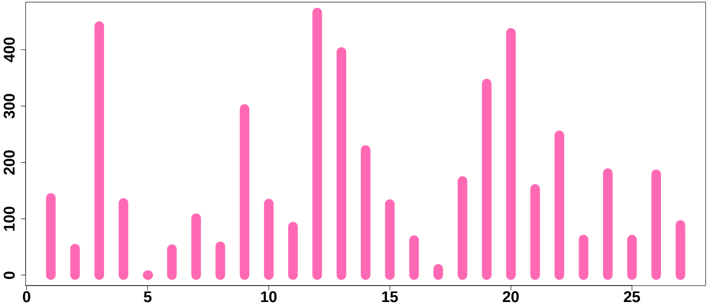
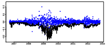


Figure 9: The frequency of the firms and macro prudential variables. The X-axis represents these 27 variables, and the Y-axis stands for the frequency with which the variables survived in the moving window estimation.

Go to details



Further work

- CoVaR estimation in Expectile situation.
- CoVaR estimation in Composite Quantile regression situation.
- Backtesting VaR and CoVaR.

The penalty term

- Lasso, Tibshirani (1996): $\gamma_\lambda(x) = \lambda$
- SCAD, Fan and Li (2001):

$$\gamma_\lambda(x) = \lambda\{\mathbf{1}(x \leq \lambda) + \frac{(a\lambda - x)_+}{(a-1)\lambda}\},$$

- The adaptive Lasso, Zou (2006): $\gamma_\lambda(x) = \lambda|x|^{-a}$ for some $a > 0$.

Return

Assumptions

A 1 . The kernel function $K(\cdot)$ is a continuous symmetric probability density function having at least four-order finite moment. And the link function $g(\cdot)$ has continuous second derivative.

A 2 . Assume $\rho_k(x)$ are all strictly convex. And suppose $\psi_k(x)$, the derivative (or a subgradient of) of $\rho_k(x)$, satisfy (1) it is Lipschitz continuous; (2) $\mathsf{E}\,\psi_k(\varepsilon_i) = 0$ and $\inf_{|v|\leq c} \partial\, \mathsf{E}\,\psi_k(\varepsilon_i - v) = C_1$ where $\partial\, \mathsf{E}\,\psi_k(\varepsilon_i - v)$ is the partial derivative with respect to v, and C_1 is a constant.

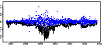
Assumptions

A 3 . The error term ε_i is independent of X_i. Let $X_{i(1)}$ denote the sub-vector of X_i consisting of its first q elements. Let $Z_i = X_i^\top \beta^*$ and $Z_{ij} = Z_i - Z_j$. Define $\mathsf{E}\{g'(Z_i)^2(\mathsf{E}(X_{i(1)}|Z_i) - X_{i(1)})(\mathsf{E}(X_{i(1)}|Z_i - X_{i(1)})\}^\top \stackrel{\text{def}}{=} C_{0(1)}$, and the matrix C_0 satisfies $L_1 \leq \lambda_{\min}(C_0) \leq \lambda_{\max}(C_0) \leq L_2$ for positive constants L_1 and L_2. There exists a constant $c_0 > 0$ such that $\sum_{i=1}^n \{\|X_{i(1)}\|/\sqrt{n}\}^{2+c_0} \to 0$.

A 4 . Assume $\sqrt{n}\gamma_\lambda(|\tilde{\beta}_l|) \to 0$ for $\beta_l^* \neq 0$ and $\sqrt{n}\gamma_\lambda(|\tilde{\beta}_l|) \to \infty$ for $\beta_l^* = 0$. Furthermore assume $qh \to 0$ as n goes to infinity.

A 5 . The error term ε_i satisfies $\mathsf{E}\,\varepsilon_i = 0$ and $\mathrm{Var}(\varepsilon_i) < \infty$. Assume that $\mathsf{E}\left|\psi^m(\varepsilon_i)/m!\right| \leq s_0 K^m$ where s_0 and K are constants.

▸ Return



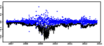
Subgradient

If $f : U \to \mathbb{R}$ is a real-valued convex function defined on a convex open set in the Euclidean space $\mathbb{R}^n$, a vector v in that space is called a subgradient at a point x_0 in U if for any x in U one has

$$f(x) - f(x_0) \geq v \cdot (x - x_0)$$

where the dot denotes the dot product.

Return

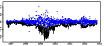


Matrix norm

Assume A is a $m \times n$ matrix

$$\|A\|_{\alpha,\beta} = \max_{x \neq 0} \frac{\|Ax\|_\beta}{\|x\|_\alpha}$$

Return



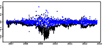
Sparsistency

Let $\widehat{\beta}$ be an estimator of β. $\widehat{\beta}$ is sparsistent if

$$\lim_{n\to\infty} \mathrm{P}\{sign(D\widehat{\beta}) = sign(D\beta)\} = 1$$

where D is the incidence matrix. Consider a undirected graph G defined by a set of vertices V and undirected edges E which are unordered pairs of vertices. We construct an orientation of G by defining a head $e^+ \in e$ and tail $e^- \in e$. The incidence matrix $D \in \mathbb{R}^{E\times V}$ for the oriented graph is the matrix whose $D_{e,v}$ entry is 1 if $v = e^+$, -1 if $v = e^-$ and 0 otherwise.

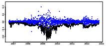
Return



The financial firms and macro prudential variables

The financial frims:	15. Northern Trust Corp (NTRS)
1. Bank of America Corp (BAC)	16. M & T Bank Corp (MTB)
2. JP Morgan Chase & Co (JPM)	17. Hudson City Bancorp Inc (HCBK)
3. Citigroup Inc (C)	18. Comerica Inc (CMA)
4. Wells Fargo & Co (WFC)	19. Huntington Bancshares Inc (HBAN)
5. U.S. Bancorp (USB)	20. Zions Bancorp (ZION)
6. PNC Financial Services Group (PNC)	The macro prudential variables:
7. Bank of New York Mellon (BK)	21. VIX
8. Capital One Financial Corp (COF)	22. Short term liquidity spread (liquidity)
9. Suntrust Banks Inc (STI)	23. Daily change in the 3-month Treasury maturities (3MT)
10. State Street Corp (STT)	24. Change in the slope of the yield curve (yield)
11. BB&T Corp (BBT)	25. Change in the credit spread (credit)
12. Financial Corp New (RF)	26. Daily Dow Jones U.S. Real Estate index returns (D_J)
13. Fifth Third Bancorp (FITB)	27. S&P500 returns (S&P)
14. KeyCorp (KEY)	

Return

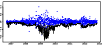


References

Adrian, T. and Brunnermeier, M. K.
CoVaR
Staff Reports 348, Federal Reserve Bank of New York, 2011

Bradic, J., Fan, J. and Wang, W.
Penalized composite quasi-likelihood for ultrahigh dimensional variable selection
J. R. Statist. Soc. B 73(3): 325-349. 2011

Chao, S. K., Härdle, W. K. and Wang, W.
Quantile regression in Risk Calibration
Discussion Paper 2012-006, CRC 649, Humboldt-Universität zu Berlin, 2012

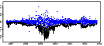


References

Engle, R. F. and Manganelli, S.
CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles
Journal of Business and Economic Statistics 22: 367-381, 2004

Fan, J. and Li, R.
Variable selection via nonconcave penalized likelihood and its oracle properties
J. Amer. Statist. Assoc. 96: 1348-1360. 2001

Franke, J., Härdle, W. and Hafner, C.
Statistics of Financial Markets
Springer, 2004

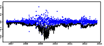


References

Hao, L. and Naiman, Q. D.
Quantile Regression
Sage Publications, 2007

Hautsch, N., Schaumburg, J. and Schienle, M.
Financial Network System Risk Contributions
Discussion Paper 2011-072, CRC 649, Humboldt-Universität zu Berlin, 2011

Huber, P. J.
Robust estimation of a location parameter
The Annals of Mathematical Statistics 35(1): 73-101. 1964

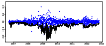


References

John C. Hull
Risk Management and Financial Institutions – 2nd ed.
Prentice Hall, 2010

John Y. Campbell, Andrew W. Lo and A. Craig MacKinlay
The Econometrics of Financial Markets
Princeton University Press, 1996

Julien Leider
A Quantile Regression Study of Climate Change in Chicago
available on www.siam.org, 2012

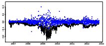


References

Koenker, R. and Bassett, G. W.
Robust tests for heteroscedasticity based on regression quantiles
Econometrica 50: 43-61. 1982

Koenker, R. and Hallock, K. F.
Quantile regression
Journal of Econometric Perspectives 15(4): 143-156. 2001

Leng, C., Xia, Y. and Xu, J.
An adaptive estimation method for semiparametric models and dimension reduction
In preparation, 2010

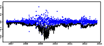


References

Li, Y. and Zhu, J.
L1- norm quantile regression
Journal of Computational and Graphical Statistics 17: 163-185. 2008

Newey, W. and Powell, J.
Asymmetric least squares estimation and testing
Econometrica 55: 819-847. 1987

Roger Koenker
Quantile Regression
Cambridge Univ. Press, 2000

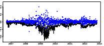


References

Schnabel, S. and Eilers, P.
Optimal expectile smoothing
Computational Statistics and Data Analysis 53(12): 4168-4177. 2009

Serfling, R. J.
Approximation theorems of mathematical statistics
Wiley, New York. 1980

Sharpnack, J. Rinaldo, A. and Singh, A.
Sparsistency of the Edge Lasso over Graphs
Appearing in Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS) 2012

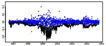


References

Tibshirani, R.
Regression shrinkage and selection via the lasso
J. R. Statist. Soc. B 58(1): 267-288. 1996

Xia, Y., Tong, H., Li, W. and Zhu, L.
An adaptive estimation of dimension reduction space
J. R. Statist. Soc. B 64, Part 3, pp. 363-410, 2002

Yu, K. and Jones, M. C.
Local linear quantile regression
J. Amer. Statist. Assoc. 98: 228-237, 1998



References

Yuan, M. and Lin, Y.
Model selection and estimation in regression with grouped variables
J. R. Statist. Soc. B 68(1): 49-67. 2006

Zou, H.
The adaptive lasso and its oracle properties
J. Amer. Statist. Assoc. 101: 476. 2006