Wang, W., Fan, Y., Härdle, W.K. & Zhu, L. (2016). Single index based CoVaR with very high dimensional covariates. Journal of Business & Economic Statistics, doi: 10.1080/07350015.2016.1180990

Single index based CoVaR with very high dimensional covariates*†

Yan Fan ‡, Wolfgang Karl Härdle§, Weining Wang¶, Lixing Zhu‖

August 5, 2016

*We thank the Editor, the Associate Editor and the referees for valuable comments.

†The financial support from the Deutsche Forschungsgemeinschaft via CRC 649 "Ökonomisches Risiko", Humboldt-Universität zu Berlin, IRTG 1792, and the Research Grants Council of Hong Kong via G-HK012/10 is gratefully acknowledged. We also gratefully acknowledge the funding from DAAD ID 50746311.

‡School of Business Information, Shanghai, 201620, China

§Professor at Humboldt-Universität zu Berlin and Director of C.A.S.E. - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany. Singapore Management University, 50 Stamford Road, Singapore 178899

¶Corresponding Author. Lecturer(assistant professor) in statistics, King's College London. Hermann-Otto-Hirschfeld Junior Professor in Nonparametric Statistics and Dynamic Risk Management at the Ladislaus von Bortkiewicz Chair of Statistics of Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany. Email:wangwein@cms.hu-berlin.de

‖Chair Professor at Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

Abstract

Systemic risk analysis reveals the interdependencies of risk factors especially in tail event situations. In applications the focus of interest is on capturing joint tail behavior rather than a variation around the mean. Quantile and expectile regression are used here as tools of data analysis. When it comes to characterizing tail event curves one faces a dimensionality problem, which is important for CoVaR (Conditional Value at Risk) determination. A projection based single index model specification may come to the rescue but for ultra high dimensional regressors one faces yet another dimensionality problem and needs to balance precision vs. dimension. Such a balance is achieved by combining semi parametric ideas with variable selection techniques. In particular, we propose a projection based single index model specification for very high dimensional regressors. This model is used for practical CoVaR estimates with a systemically chosen indicator. In simulations we demonstrate the practical side of the semiparametric CoVaR method. The application to the US financial sector shows good backtesting results and indicate market coagulation before the crisis period.

Keyword:

Quantile Single-index Regression, Minimum Average Contrast Estimation, CoVaR, Composite Quasi-Maximum Likelihood Estimation, Lasso, Model Selection

JEL Classification: C00, C14, C50, C58

1 Introduction

It is known to be a challenging task to manage financial risk due to joint extreme events, reflecting the fact that in times of crisis losses tend to spread across a portfolio. The key interest is to understand and forecast the risk exposure of e.g. a financial institution in the market for firm leaders or to identify and select systemic risk relevant factors for government regulators. There is a large amount of literature on measuring systemic risk. We focus on the line of research adopting quantile methods to quantify the tail dependence among financial institutions. In particular, Adrian and Brunnermeier (2011) propose a systemic risk measure, called CoVaR, with balance sheet characteristics driven individual risk exposure. Furthermore, Hautsch et al. (2014) introduce an applicable measure of a firm's systemic relevance, explicitly accounting for the company's interconnectedness within the financial sector.

The underlying statistical setting involved is a two-stage linear quantile regression. Several elements of the existing CoVaR methodology are, however, based on questionable assumptions: First, a significant degree of nonlinearity occurs when modeling conditional tail curves. Second, the number of potential risk factors is large in comparison with the amount of available observations. Third, the selected factors are difficult to be interpreted, and need to be summarized to an index. Therefore, one calls for a data driven technique that combines dimension reduction, variable selection and generalized tail events e.g. expectiles. In this paper we address these points and provide a practical CoVaR estimate together with a systemically chosen indicator. The systemic indicator is chosen by the single index approach, which has a unique feature: the index that yields interpretability and low dimension simultaneously. However, in the case of ultra high dimensional regressors X the single index approach suffers from singularity problems. Efficient variable selection is the strategy to employ here. Specifically we consider composite regression with general weighted loss and possible ultra high dimensional covariates. Our setup is general, and includes quantile, expectile (and therefore mean as a special case) regression. We offer theoretical properties and demonstrate our method with applications to firm risk analysis in a CoVaR estimation context.

The basic element of our CoVaR estimation is quantile regression(QR). In many fields of applications such as quantitative finance, econometrics, marketing and also medical and biological sciences, QR is a fundamental element for data analysis, modeling and inference. An application in finance is the analysis of time varying Value-at-Risk (VaR) using the Conditional Autoregressive Value at Risk (CaViaR) model, see Engle and Manganelli (2004). The QR estimation may be seen as an estimation problem by assuming an asymmetric ALD (asymmetric Laplace distribution) pseudo likelihood, which not necessarily return an efficient estimator. Therefore, different flexible loss functions are considered in the literature to improve the estimation efficiency, such as, composite quantile regression, Zou et al. (2008), Kai et al. (2010) and Kai et al. (2011). Moreover, Bradic et al. (2011) propose a general loss function framework for linear models, with a weighted sum of different kinds of loss functions, and the weights are selected to be data driven. Another type of loss considered is in Newey and Powell (1987) corresponding to expectile regression (ER). This is similar in spirit to QR but contains mean regression as a special case. Nonparametric expectile smoothing work with applications to demography can be found in Schnabel and Eilers (2009). The ER curves are alternatives to the QR curves and give us an alternative regression picture.

The difficulty of characterizing an entire distribution partly arises from the high dimensionality of covariates. This asks to strike a balance between model flexibility and statistical precision. To crack this tough nut, dimension reduction techniques of semiparametric type, such as the single index model, came into the focus of statistical modeling. Wu et al. (2010) and Kong and Xia (2012) consider quantile regression via a single index model. However, to our knowledge there is no further literature on generalized QR for the single-index model.

In addition to the dimension reduction, there is also the problem (incurred in our CoVaR estimation procedure) of choosing the right variables for projection. This motivates our second goal of this research: variable selection. Kong and Xia (2007), Wang and Yin (2008) and Zeng et al. (2012) focus on variable selection in mean regression for the single index model. The set of ideas presented there, however, have never been applied to a quantile, composite quantile framework or to a even more general (composite) quasi-likelihood framework. The semiparametric single index approach that we consider herein will be a good tool for practitioners, as it combines flexibility in modeling with applicability for even very high dimensional data.

This article is organized as follows: In Section 2, we introduce the basic setup and the estimation algorithm. In Section 3, we build up asymptotic theorems for our model. In Section 4, simulations are carried out. In Section 5, we illustrate our methodology by estimating CoVaR. All the technical details can be found in the appendix.

2 MACE for Single Index Model

Let X and Y be p dimensional, continuous random variables respectively, (p can be very large, namely of the rate $\exp(n^{\delta})$, where (δ is a constant whose range will be defined in Condition 4 in Section 3. The single index model (SIM) is defined to be:

$$Y = g(X^{\top}\beta^{*}) + \varepsilon, \tag{2.1}$$

where $g(\cdot) : \mathbb{R}^1 \longmapsto \mathbb{R}^1$ is an *unknown* smooth link function, β^* is the vector of index parameters, ε is a continuous variable with mean zero. The interest here is to simultaneously estimate β^* and $g(\cdot)$. The assumptions on error structure can be seen in Condition 3.

2.1 Quasi-Likelihood for the Single Index Model

Several estimation techniques exist for (2.1), among which the average derivative estimator (ADE) method is one of the oldest ones, see Härdle and Stoker (1989). The semiparametric SIM (2.1) also permits a one-step projection pursuit interpretation, therefore estimation tools from this stream of literature might also be employed, see Huber (1985). The minimum average variance estimation (MAVE) technique aimed at simultaneous estimation of (β^*, $g(\cdot)$) was proposed by Xia et al. (2002). Here we will apply a Minimum Average Contrast Estimation Approach, called MACE.

Similar to MAVE, the MACE technique uses double integration but allow more general loss functions. Our estimation framework is new in three aspects. First, we consider a general class of contrast functions that allow us to identify and estimate conditional quantiles, expectiles and other tail specific objects. Second, we consider the situation where p might be very large and we add penalty terms that lead to an automatic model selection framework of e.g. the least absolute shrinkage and selection operator (Lasso) or Smoothly Clipped Absolute Deviation (SCAD) type. Third, we implement a composite estimation technique for efficiency improvement.

In our theoretical setup, we identify the parameter via a minimum contrast with ρ_{w} as the contrast function. It corresponds, as mentioned above, to a quasi maximum likelihood framework: the direction β^* (for known $g(\cdot)$) is the solution of

$$\min_{\beta} \mathsf{E}\, \rho_{\mathsf{w}}\{Y - g(X^\top \beta)\}, \tag{2.2}$$

with the general quasi-likelihood loss function $\rho_{\mathsf{w}}(\cdot) = \sum_{k=1}^{K} \mathsf{w}_k \rho_k(\cdot)$, where $\rho_1(\cdot)$, . . . , $\rho_K(\cdot)$ are convex loss functions and $\mathsf{w}_1, ..., \mathsf{w}_K$ are positive weights.

Equivalently, β is the solution to

$$\mathsf{E}(\psi_{\mathsf{w}}\{Y - g(X^\top \beta)\}|X) = 0 \quad a.s.$$

(where $\psi_{\mathsf{w}}(\cdot)$ is the derivative (a subgradient) of $\rho_{\mathsf{w}}(\cdot)$). This weighted loss function includes many situations such as ordinary least square, quantile regression(QR), expectile regression(ER), composite quantile regression(CQR) and so on. For model identification, we assume that the L_2-norm of β^*, $\|\beta^*\|_2 = 1$ and the first component of β^* is positive.

The standard situation of QR is with $K = 1$ and the conditional quantile function $F^{-1}_{\varepsilon|X}(\tau) = 0$. This means to take the loss function as:

$$\rho_{\mathsf{w}}(u) = \tau u \mathbf{1}(u \geq 0) - (1-\tau)u\mathbf{1}(u < 0), \tag{2.3}$$

where $\mathbf{1}(A)$ is equal to 1 if A is true and 0 otherwise. Moreover, for ER with $K = 1$, we have:

$$\rho_{\mathsf{w}}(u) = \tau u^2 \mathbf{1}(u \geq 0) + (1-\tau)u^2\mathbf{1}(u < 0). \tag{2.4}$$

The general form of $\rho_{\mathsf{w}}(\cdot)$ boils downs to CQR when one employs K different quantiles $\tau_1, \tau_2, \ldots, \tau_K$, with $w_k = 1/K, k = 1, \ldots, K$ and

$$\rho_k(u) = \tau(u - b_k)\mathbf{1}(u - b_k \geq 0) + (1-\tau)(u - b_k)\mathbf{1}(u - b_k < 0), \tag{2.5}$$

where b_k is the τ_k quantile of the error distribution, see Bradic et al. (2011).

Let us now launch the MACE. First, we approximate $g(X_i^\top \beta)$ for $x^\top \beta$ near $X_i^\top \beta$:

$$g(X_i^\top \beta) \approx g(x^\top \beta) + g'(x^\top \beta)(X_i - x)^\top \beta. \tag{2.6}$$

In the context of local linear smoothing, a first order proxy of β (given x) can therefore be constructed by minimizing:

$$L_x(\beta, g(\cdot)) \stackrel{\text{def}}{=} \mathsf{E}\, \rho_{\mathsf{w}}\{Y - g(x^\top \beta) - g'(x^\top \beta)(X_i - x)^\top \beta\}. \tag{2.7}$$

The empirical version of (2.7) requires minimizing, with respect to β and function

$g(\cdot)$:

$$L_{n,x}(\beta, g(\cdot)) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^{n} \rho_{\mathsf{w}}\{Y_i - g(x^\top \beta) - g'(x^\top \beta)(X_i - x)^\top \beta\} K_h\{(X_i - x)^\top \beta\} \quad (2.8)$$

where $K_h(\cdot)$ is a kernel function with $K_h(u) = h^{-1}K(u/h)$ and h is a bandwidth parameter. We adopt now the double integration idea of MAVE, i.e. we integrate with respect to the empirical distribution function of the covariates leading to the following loss function:

$$\begin{aligned} L_n(\beta, g(\cdot)) \quad &\stackrel{\text{def}}{=} \quad n^{-2} \sum_{j=1}^{n} \sum_{i=1}^{n} \rho_{\mathsf{w}} \left\{ Y_i - g(X_j^\top \beta) - g'(X_j^\top \beta)(X_i - X_j)^\top \beta \right\} \\ &\qquad K_h\{(X_i - X_j)^\top \beta\}. \end{aligned} \quad (2.9)$$

Minimizing (2.9) with respect to β and $g(\cdot)$ is the basic idea.

For simplicity, from now on we write $g(X_j^\top \beta)$ and $g'(X_j^\top \beta)$ as $a(X_j)$ and $b(X_j)$ or a_j and b_j respectively. The calculation of the above minimization problem can be decomposed into two subproblems, motivated by e.g. Leng et al. (2008),

a) Given β, the estimation of $a(\cdot)$ and $b(\cdot)$ are obtained through local linear minimization.

b) Given $a(\cdot)$ and $b(\cdot)$, the minimization with respect to β is carried out by the interior point method.

2.2 Variable Selection for Single Index Model

The dimension of covariates (p) is large, even one can allow $p = \mathcal{O}\{\exp(n^\delta)\}$, so selecting important covariates is a necessary step. Without loss of generality assume that the first q components of β^* minimizing (2.2) are non-zero. To point this out write $\beta^* = (\beta_{(1)}^{*\top}, \beta_{(0)}^{*\top})^\top$ with $\beta_{(1)}^* \stackrel{\text{def}}{=} (\beta_1, \ldots, \beta_q)^\top \neq 0$ and $\beta_{(0)}^* \stackrel{\text{def}}{=} (\beta_{q+1}, \ldots, \beta_p)^\top = 0$ element-wise. Accordingly we denote $\mathbf{X}_{(1)}$ and $\mathbf{X}_{(0)}$ as the first q and the last $p-q$ column of design matrix $\mathbf{X}$, corresponding to $\beta_{(1)}^{*\top}$ and $\beta_{(0)}^{*\top}$ respectively.

Suppose $\{(X_i, Y_i)\}_{i=1}^n$ are n independent and identically distributed (i.i.d.) copies of (X, Y). Consider first estimating the SIM coefficient β^* by solving the optimization problem

$$\min_{(a_j,b_j)'\text{s},\beta} n^{-1} \sum_{j=1}^n \sum_{i=1}^n \rho_{\mathsf{w}}\big(Y_i - a_j - b_j X_{ij}^\top \beta\big)\omega_{ij}(\beta) + \sum_{l=1}^p \gamma_\lambda(|\hat{\beta}_l^{(0)}|)|\beta_l|, \tag{2.10}$$

where $X_{ij} \stackrel{\text{def}}{=} X_i - X_j$, $\omega_{ij}(\beta) \stackrel{\text{def}}{=} K_h(X_{ij}^\top\beta)/\sum\limits_{i=1}^n K_h(X_{ij}^\top\beta)$. Here $\gamma_\lambda(t)$ is some non-negative function, and $\hat{\beta}^{(0)}$ is an initial estimator of β^* (eg. linear QR with variable selection). The penalty term in (2.10) is quite general and it covers the most popular variable selection criteria as special cases: the Lasso Tibshirani (1996) with $\gamma_\lambda(x) = \lambda$, the SCAD Fan and Li (2001) with

$$\gamma_\lambda(x) = \lambda\{\mathbf{1}(|x| \le \lambda) - \frac{(|x|^2 - 2c_1\lambda|x| + \lambda^2)_+}{|x|(c_1-1)2\lambda}\mathbf{1}(\lambda < |x| \le c_1\lambda) + \frac{(c_1+1)\lambda}{2|x|}\mathbf{1}(|x| > c_1\lambda)\},$$

with ($c_1 > 2$) and $\gamma_\lambda(x) = \lambda|x|^{-c_2}$ for some $c_2 > 0$ corresponding to the adaptive Lasso, Zou (2006).

We propose to estimate β^* in (2.10) with the MACE iterative procedure described below. Denote $\hat{\beta}_{\mathsf{w}}$ the final estimate of β^*. Specifically, for $t = 1, 2, \cdots$, iterate the following two steps. Denote $\hat{\beta}^{(t)}$ as the estimate at step t.

a) Given $\hat{\beta}^{(t)}$, standardize $\hat{\beta}^{(t)}$ so that $\hat{\beta}^{(t)}$ has length one and positive first component. Then compute

$$(\hat{a}_j^{(t)}, \hat{b}_j^{(t)}) \stackrel{\text{def}}{=} \underset{(a_j,b_j)'\text{s}}{\arg\min} \sum_{i=1}^{n} \rho_{\mathsf{w}}\big(Y_i - a_j - b_j X_{ij}^{\top}\hat{\beta}^{(t)}\big)\omega_{ij}(\hat{\beta}^{(t)}). \tag{2.11}$$

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

$$\hat{\beta}^{(t+1)} = \arg\min_{\beta} \sum_{j=1}^{n}\sum_{i=1}^{n} \rho_{\mathsf{w}}\big(Y_i - \hat{a}_j^{(t)} - \hat{b}_j^{(t)} X_{ij}^{\top}\beta\big)\omega_{ij}(\hat{\beta}^{(t)}) + n\sum_{l=1}^{p} \hat{d}_l^{(t)}|\beta_l|, \tag{2.12}$$

where $\hat{d}_l^{(t)} \stackrel{\text{def}}{=} \gamma_\lambda(|\hat{\beta}_l^{(t)}|)$. Please note here that the kernel weights $\omega_{ij}(\cdot)$ use the $\hat{\beta}^{(t)}$ from the step before.

When choosing the penalty parameter λ, we adopt a C_p-type criterion as in Yuan and Lin (2006) instead of the computationally involved cross validation method. We choose the optimal weights of the convex loss functions ρ_{w} by minimizing the asymptotic variance of the resulting estimator of β^*, and the bandwidth h by criteria proposed in Yu and Jones (1998) for $g(\cdot)$.

3 Main Theorems

Define $\hat\beta_{\mathsf{w}} \stackrel{\text{def}}{=} (\hat\beta_{\mathsf{w}(1)}^\top, \hat\beta_{\mathsf{w}(2)}^\top)^\top$ as the estimator for $\beta^* \stackrel{\text{def}}{=} (\beta_{(1)}^{*\top}, \beta_{(2)}^{*\top})^\top$ attained by the procedure in (2.11) and (2.12). Let $\hat\beta_{\mathsf{w}(1)}$ and $\hat\beta_{\mathsf{w}(2)}$ be the first q components and the remaining $p-q$ components of $\hat\beta_{\mathsf{w}}$ respectively. If in the iterations, we have the initial estimator $\hat\beta_{(1)}^{(0)}$ as a $\sqrt{n/q}$ consistent one for $\beta_{(1)}^*$ (2.12), we will obtain with a very high probability, an oracle estimator of the following type, say $\hat\beta_{\mathsf{w}} = (\hat\beta_{\mathsf{w}(1)}^\top, \mathbf{0}^\top)^\top$, since the oracle knows the true active set $\mathcal{M}_* \stackrel{\text{def}}{=} \{l : \beta_l^* \neq 0\}$. The following theorem shows that the penalized estimator enjoys the oracle property. Define $\hat\beta^0$ (note that it is different from the initial estimator $\hat\beta_{(1)}^{(0)}$) as the minimizer with the same loss in (2.10) but within subspace $\{\beta \in \mathbb{R}^p : \beta_{\mathcal{M}_*^c} = \mathbf{0}\}$.

We make the following assumptions for the proofs of the theorems in this paper. Let $Z_i \stackrel{\text{def}}{=} X_i^\top \beta^*$ and $Z_{ij} \stackrel{\text{def}}{=} Z_i - Z_j$.

Condition 1. The kernel $K(\cdot)$ is a continuous symmetric function. The link function $g(\cdot) \in C^2$, where C^2 is the function space consisting of functions with second order continuous derivatives.

Condition 2. Assume that for all $k = 1, \cdots, K$, $\rho_k(x)$ is convex and not continuous on finite number of points. Suppose $\psi_k(x)$, the derivative (or a subgradient of) of $\rho_k(x)$, satisfies $\mathsf{E}\{\psi_k(\varepsilon)|Z_i\}$ would only be a function related to k such that $\mathsf{E}\{\psi_{\mathsf{w}}(\varepsilon)|Z_i\} = 0$, a.s., $\mathsf{E}\{\psi_k^2(\varepsilon)|Z_i\} < \infty$. Let $H_i(c) \stackrel{\text{def}}{=} \inf_{|v|\le c} \partial\, \mathsf{E}\, \psi_{\mathsf{w}}(\varepsilon_i - v) = C_1$, where $\partial\, \mathsf{E}\, \psi_k(\varepsilon - v)$ is the partial derivative with respect to v, and c and C_1 are positive constants.

Condition 3. $\{(X_i, Y_i)\}_{i=1}^n$ be n i.i.d. copies of (X, Y). The density of $\beta^{*\top} X$ is bounded with bounded absolute continuous first-order derivatives on its support. Let $X_{i(1)}$ denote the sub-vector of X_i consisting of its first q elements.

Define

$$C_{1(1)} \stackrel{\text{def}}{=} \mathsf{E}\left[\mathsf{E}_{\varepsilon_i|Z_i}\psi_{\mathsf{w}}^2(\varepsilon_i)\{[g'(Z_i)]^2(\mathsf{E}(X_{i(1)}) - X_{i(1)})(\mathsf{E}(X_{i(1)}) - X_{i(1)})^\top\}\right] \quad (3.1)$$

$$C_{0(1)} \stackrel{\text{def}}{=} \mathsf{E}\left[\partial\,\mathsf{E}_{\varepsilon_i|Z_i}\psi_{\mathsf{w}}(\varepsilon_i)\{[g'(Z_i)]^2(\mathsf{E}(X_{i(1)}|Z_i) - X_{i(1)})(\mathsf{E}(X_{i(1)}|Z_i) - X_{i(1)})^\top\}\right] \quad (3.2)$$

and the matrix $C_{0(1)}$ satisfies $0 < L_1 \leq \lambda_{\min}(C_{0(1)}) \leq \lambda_{\max}(C_{0(1)}) \leq L_2 < \infty$ for positive constants L_1 and L_2. There exists a constant C_3 such that for all $\beta \in \{\|\beta - \beta^*\| \leq C_3\}$,

$$\|\,\mathsf{E}\left[\partial\,\mathsf{E}_{\varepsilon|Z_i}\{\psi_{\mathsf{w}}(\varepsilon)\}g'(Z_i)\{(X_{(0)}|Z_i) - X_{i(0)}\}\{(X_{(1)}|Z_i) - X_{i(1)}\}^\top\right]\|_{2,\infty} = \mathcal{O}(1),$$

where for a matrix B, $\|B\|_{2,\infty} = \max_{\|u\|=1}\|Bu\|_\infty$.

Condition 4. Let $d_l \stackrel{\text{def}}{=} \gamma_\lambda(|\beta_l^*|)$ with the penalty parameter $\liminf_{n\to\infty}\lambda \geq n^{-1/2+\alpha_2/2}$ and $D_n \stackrel{\text{def}}{=} \max\{d_l : l \in \mathcal{M}_*\} = o(n^{\alpha_1-\alpha_2/2}\lambda)$ where $\mathcal{M}_* = \{l : \beta_l^* \neq 0\}$ be the true model. Assume that $\liminf_{n\to\infty}\min_j\{d_j/\lambda : j \in \mathcal{M}_*^c\} > 0$. Furthermore assume $qh \to 0$ and $h^{-1}\sqrt{q/n} = o(1)$ as n goes to infinity, $q = \mathcal{O}(n^{\alpha_2})$, $p = \mathcal{O}\{\exp(n^\delta)\}$, $nh^3 \to \infty$ and $h \to 0$. Also, $0 < \delta < \alpha < \alpha_2/2 < 1/2$, $\alpha_2/2 < \alpha_1 < 1$.

Condition 5. The error term ε_i satisfies $\mathrm{Var}(\varepsilon_i) < \infty$. Assume that for any integer $m \geq 1$

$$\mathsf{E}\left|\psi_{\mathsf{w}}^m(\varepsilon_i)/m!\right| \leq s_0 M^m \quad (3.3)$$

where s_0 and M are constants, and $\psi_{\mathsf{w}}(\cdot)$ is the derivative (a subgradient) of $\rho_{\mathsf{w}}(\cdot)$.

Condition 6. The conditional density function $f(\varepsilon|Z_i = u)$ is bounded and absolutely continuous differentiable.

Condition 1 is commonly-used and the standard normal probability density function is a kernel satisfying this condition. Condition 2 is made on the weighted loss function so that it admits a quadratic approximation. Condition 2 assumes the dependence structure between errors and the covariates. For the CQR estimation in case of $K > 1$, it means that $F_{Y|X}^{-1}(\tau_k) = g(\beta^{*\top} X) + c(\tau_k)$ for all $\tau_1 \le \tau_k \le \tau_K$, where $c(\tau_k)$ is only a constant depending on τ_k, this is a similar condition as Wang et al. (2012). For $K = 1$ the assumption $\mathsf{E}\{\psi_{\mathsf{w}}(\varepsilon)|X\} = 0$ a.s. to $F_{\varepsilon|X}^{-1}(\tau) = 0$. Under Condition 3, the matrix in the quadratic approximation is non-singular, so that the resulting estimate of β has a non-degenerate limiting distribution. Condition 4 guarantees that the proposed variable selection and estimation procedure for β is model-consistent. Condition 5 implies a common tail behavior that we employ. Condition 6 is essential for the uniform Bahadur representation which we adopt in the proof.

Theorem 1. *Under Conditions 1-6, the estimators $\hat{\beta}^0$ and $\hat{\beta}_{\mathsf{w}}$ exist and coincide on a set with probability tending to 1. Moreover,*

$$\mathrm{P}(\hat{\beta}^0 = \hat{\beta}_{\mathsf{w}}) \ge 1 - (p - q)\exp(-C' n^{\alpha}) \tag{3.4}$$

for a positive constant C'.

It is worth noting that the above results imply the usual sign consistency, see e.g. Fan and Lv (2010). In addition, the theorem requires a relationship between the order of p, q, and the parameter α, see Condition 4.

Theorem 2. *Under Conditions 1-6, we have*

$$\|\hat{\beta}_{\mathsf{w}(1)} - \beta^*_{(1)}\| = \mathcal{O}_p\{(D_n + n^{-1/2})\sqrt{q}\} \tag{3.5}$$

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

$$\boldsymbol{b}^\top C_{0(1)}^{1/2} C_{1(1)}^{-1/2} C_{0(1)}^{1/2} \sqrt{n}(\hat{\beta}_{\mathsf{w}(1)} - \beta^*_{(1)}) \xrightarrow{\mathcal{L}} \mathsf{N}(0,1) \tag{3.6}$$

where racall that $C_{1(1)} \stackrel{\text{def}}{=} \mathsf{E}\{\mathsf{E}\{\psi^2_{\mathsf{w}}(\varepsilon_i)|Z_i\}[g'(Z_i)]^2[\mathsf{E}(X_{i(1)}|Z_i) - X_{i(1)}][\mathsf{E}(X_{i(1)}|Z_i) - X_{i(1)}]^\top\}$, *and* $C_{0(1)} \stackrel{\text{def}}{=} \mathsf{E}\{\partial\,\mathsf{E}\,\psi_{\mathsf{w}}(\varepsilon_i)|Z_i\}\{[g'(Z_i)]^2(\mathsf{E}(X_{i(1)}|Z_i) - X_{i(1)})(\mathsf{E}(X_{i(1)}|Z_i) - X_{i(1)})\}^\top$. *Note that* $\mathsf{E}(X_{i(1)}|Z_i)$ *denotes a* $q \times 1$ *dimension vector, and* $Z_i \stackrel{\text{def}}{=} X_i^\top \beta^*$, $\psi_{\mathbf{w}}(\varepsilon)$ *is a choice of the subgradient of* $\rho_{\mathbf{w}}(\varepsilon)$ *and*
$\sigma^2_{\mathbf{w}} \stackrel{\text{def}}{=} \mathsf{E}\{[\psi_{\mathsf{w}}(\varepsilon_i)]^2\}/[\partial\,\mathsf{E}\,\psi_{\mathsf{w}}(\varepsilon_i)]^2$, *where*

$$\partial\,\mathsf{E}\{\psi_{\mathsf{w}}(\cdot)|Z_i\} = \frac{\partial\,\mathsf{E}\{\psi_{\mathsf{w}}(\varepsilon_i - v)^2|Z_i\}}{\partial v}\Big|_{v=0}. \tag{3.7}$$

It is worth noting that in the case of quantile regression, $\sigma^2_{\mathbf{w}} = \tau(1-\tau)/f_{\varepsilon|Z}(0)^2$.

Let us now look at the distribution of the estimated link function $\hat{g}(x^\top\hat{\beta}_{\mathsf{w}})$ with the consistent estimate for β^* and the estimate $\hat{g}'(x^\top\hat{\beta}_{\mathsf{w}})$ with the consistent estimate of β^* plugged in.

Theorem 3. *Under conditions 1-6, let* $\mu_j \stackrel{\text{def}}{=} \int u^j K(u)du$ *and* $\nu_j \stackrel{\text{def}}{=} \int u^j K^2(u)du$, $j = 0, 1, 2$. *For any interior point* $z = x^\top\beta^*$, $f_Z(z)$ *is the density of* Z_i, $i = 1, \ldots, n$, *if $nh^3 \to \infty$ and $h \to 0$, we have*

$$\sqrt{nh}\sqrt{f_Z(z)/(\nu_0\sigma^2_{\mathsf{w}})}\left\{\hat{g}(x^\top\hat{\beta}_{\mathsf{w}}) - g(x^\top\beta^*) - \frac{1}{2}h^2 g''(x^\top\beta^*)\mu_2\partial\,\mathsf{E}\,\psi_{\mathsf{w}}(\varepsilon)\right\} \xrightarrow{\mathcal{L}} \mathsf{N}\,(0,\ 1)\,,$$

Also, we have

$$\sqrt{nh^3}\sqrt{\{f_Z(z)\mu_2^2\}/(\nu_2\sigma^2_{\mathsf{w}})}\left\{\hat{g}'(x^\top\hat{\beta}_{\mathsf{w}}) - g'(x^\top\beta^*)\right\} \xrightarrow{\mathcal{L}} \mathsf{N}\,(0,\ 1)\,,$$

not that $\sqrt{f_Z(z)/(\nu_0\sigma^2_{\mathsf{w}})}$ and $\sqrt{f_Z(z)\mu_2^2/(\nu_2\sigma^2_{\mathsf{w}})}$ are the scaling according to the standard deviations of the estimates, and recall $\sigma^2_{\mathsf{w}} \stackrel{\text{def}}{=} \mathsf{E}\{[\psi_{\mathsf{w}}(\varepsilon_i)]^2\}/[\partial\,\mathsf{E}\,\psi_{\mathsf{w}}(\varepsilon_i)]^2$.

All the proofs of the theorems can be found in Appendix (supplementary materials).

4 Simulation

In this section, we evaluate our technique in several settings, involving different combinations of link functions $g(\cdot)$, distributions of ε, and different choices of (n, p, q, τ)s, where n is the sample size, p is the dimension of the true parameter β^*, q is the number of non-zero components in β^*, and τ represents the quantile level. The evaluation is first done with a simple quantile loss function, and then with the composite $L_1 - L_2$ and the composite quantile cases. The weights $\mathsf{w}_1, \cdots, \mathsf{w}_K$ are preestimated by minimizing the object $\sum_l^K \sum_k^K \mathsf{w}_l \mathsf{w}_k \sum_{i=1}^n \psi_l(\hat{\varepsilon}_i^{(0)})\psi_k(\hat{\varepsilon}_i^{(0)})$, where $\hat{\varepsilon}_i^{(0)}$s are residuals for the initial estimator.

4.1 Link functions

Consider the following nonlinear link functions $g(\cdot)$s. Model 1:

$$Y_i = 5\cos(D_1 \cdot Z_i) + \exp(-D_1 \cdot Z_i^2) + \varepsilon_i, \tag{4.1}$$

where $Z_i = X_i^\top \beta^*$, $D_1 = 0.01$ is a scaling constant and ε_i is an error term. Model 2:

$$Y_i = 10\sin\{\pi(A \cdot Z_i - B)\} + \varepsilon_i, \tag{4.2}$$

with the parameters $A = 0.3$, $B = 3$. Finally Model 3 is with $D_2 = 0.1$:

$$Y_i = 10\sin(D_2 \cdot Z_i) + \sqrt{|\sin(0.5 \cdot Z_i) + \varepsilon_i|}. \tag{4.3}$$

4.2 Criteria

For estimation accuracy for β and $g(\cdot)$, we use the following five criteria to measure:

1) Standardized L_2 norm:

$$Dev \overset{\text{def}}{=} \frac{\|\beta^* - \widehat{\beta}\|}{\|\beta^*\|},$$

2) Sign consistency:

$$Acc \overset{\text{def}}{=} \sum_{l=1}^{p} |\mathbf{1}\{\beta_l^* \neq 0\} - \mathbf{1}\{\widehat{\beta}_l \neq 0\}|,$$

3) Least angle:

$$Angle \overset{\text{def}}{=} \frac{\beta^{*\top}\widehat{\beta}}{\|\beta^*\| \cdot \|\widehat{\beta}\|},$$

4) Average squared error:

$$ASE \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}\big\{g(Z_i) - \widehat{g}(\widehat{Z}_i)\big\}^2.$$

4.3 L_1-norm quantile regression

We adopt the algorithm for the L_1-norm quantile regression developed by Li and Zhu (2008). The initial estimate of β^* can be calculated by the L_1-norm quantile regression, and then we perform the two-step iterations mentioned in Section 2. Recall that $\mathbf{X}$ is a $p \times n$ matrix, and q is the number of non-zero components in β^*. The jth column of $\mathbf{X}$ is an i.i.d. sample from $\mathsf{N}(j/2, 1)$. Two error distributions are considered: $\varepsilon_i \sim \mathsf{N}(0, 0.1)$ and $t(5)$. Note that $\beta^*_{(1)}$ is the vector of the non-zero components in β^*. In the simulation, we consider different $\beta^*_{(1)}$: $\beta^{*\top}_{(1)} = (5, 5, 5, 5, 5)$, $\beta^{*\top}_{(1)} = (5, 4, 3, 2, 1)$ and $\beta^{*\top}_{(1)} = (5, 2, 1, 0.8, 0.2)$. Here the indices Z_is are re-scaled to $[0, 1]$ for nonparametric estimation. The bandwidth is selected as in Yu and Jones (1998):

$$h_\tau = h_{mean}\big[\tau(1-\tau)\varphi\{\Phi^{-1}(\tau)\}^{-2}\big]^{0.2}.$$

where h_{mean} can be calculated by using the direct plug-in methodology of a local linear regression described by Ruppert et al. (1995). To see the performance of the bandwidth selection, we compare the estimated link functions with different bandwidths. Figure 1 is an example showing the true link function (grey) and the estimated link function (black). The left plot in Figure 1 is with the bandwidth ($h = 0.68$) selected by applying the aforementioned bandwidth selection. We can see that the estimated link function curve is relatively smooth. The middle plot shows the estimated link function with a smaller bandwidth ($h = 0.068$). It can be seen that the estimated curve is wiggly shape. The right plot shows that the estimated link function with a larger bandwidth ($h = 0.8$), the deviation between the estimated link function curve and the true curve is very large.

Figure 1: The true link functions (grey) and the estimated link functions (black) in model 2 with $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$, and $\varepsilon \sim \mathsf{N}(0, 0.1)$, $n = 100, p = 10, q = 5, \tau = 0.05$, where $h = 0.68$ (left), $h = 0.068$ (middle) and $h = 0.8$ (right).

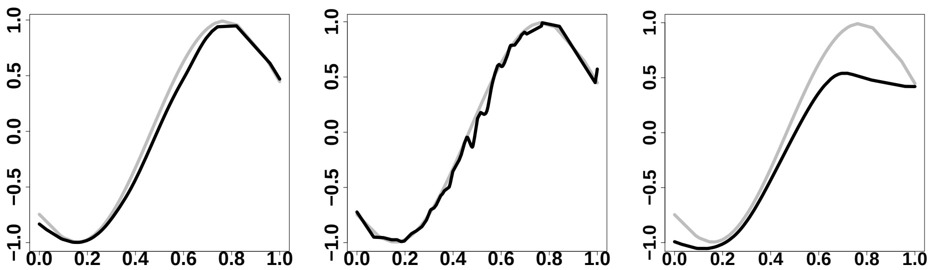


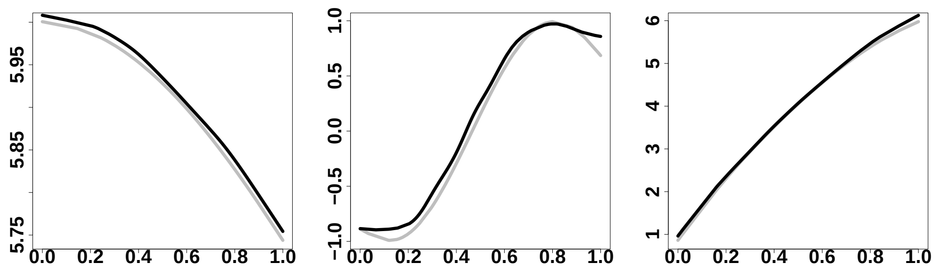
Table 1 shows the criteria evaluated with different models and quantile levels. Here $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$, the error ε follows a $\mathsf{N}(0, 0.1)$ distribution or follows a $t(5)$ distribution. In 10000 simulations we set $p = 10, q = 5$. Standard deviations are given in brackets. We find that for quantile levels 0.95 and 0.05 the errors are usually slightly larger than the median. Although the estimations for model 2 are not as good as for model 1 and model 3, the errors are still moderate. Figures 2 and Figure 3 present the plots of the true link functions against the estimated ones for different quantile levels.

Table 2 reports on the criteria evaluated under different $\beta_{(1)}^{*}$ cases. In this table two different $\beta_{(1)}^{*}$ are considered: (a) $\beta_{(1)}^{*\top} = (5, 4, 3, 2, 1)$, (b) $\beta_{(1)}^{*\top} = (5, 2, 1, 0.8, 0.2)$, the error ε follows a $\mathsf{N}(0, 0.1)$ distribution. In 10000 simulations we set $p = 10, q = 5, \tau = 0.95$. Standard deviations are given in brackets. We notice that for the case (b), the estimation results are not better than (a) since the smaller values of $\beta_{(1)}^{*}$ in case (b) would be estimated as zeros, and the estimation of the link function

Table 1: Criteria evaluated with different models and quantiles. $\beta_{(1)}^{*\top} = (5,5,5,5,5)$, N means the error ε follows a $\mathsf{N}\,(0, 0.1)$ distribution, t means the error ε follows a $t(5)$ distribution. In 10000 simulations we set $n = 100, p = 10, q = 5$. Standard deviations are given in brackets. *Dev*, *Acc*, *Angle*, *Error* and their standard deviations are reported in 10^{-1}. *ASE* and its standard deviations are reported in 10^{-2}.

$g(\cdot)$	ε	τ	Dev	Acc	Angle	ASE
Model 1	N	0.95	1.213(0.332)	0.949(0.327)	9.656(0.086)	0.357(0.085)
		0.50	1.132(0.137)	0.993(0.244)	9.736(0.022)	0.247(0.044)
		0.05	1.346(0.532)	1.335(0.443)	9.626(0.116)	0.260(0.076)
	t	0.95	1.736(0.744)	0.926(0.478)	9.548(0.135)	0.809(0.097)
		0.50	1.236(0.246)	1.157(0.357)	9.667(0.040)	0.448(0.093)
		0.05	1.536(0.737)	2.447(0.446)	9.570(0.126)	0.923(0.097)
Model 2	N	0.95	4.679(0.854)	6.579(0.643)	9.581(0.658)	1.768(0.247)
		0.50	1.489(0.458)	5.015(0.436)	9.455(0.274)	1.156(0.464)
		0.05	1.501(0.825)	6.858(0.747)	9.388(0.658)	2.015(0.274)
	t	0.95	5.325(0.960)	9.226(0.758)	9.360(0.567)	2.467(0.351)
		0.50	1.689(0.557)	7.004(0.879)	9.409(0.379)	1.279(0.473)
		0.05	2.065(0.847)	8.546(0.951)	9.475(0.531)	2.639(0.368)
Model 3	N	0.95	0.757(0.269)	1.702(0.248)	9.966(0.013)	0.569(0.162)
		0.50	0.618(0.175)	1.434(0.186)	9.867(0.021)	0.695(0.104)
		0.05	0.558(0.315)	1.845(0.173)	9.979(0.024)	0.758(0.173)
	t	0.95	0.625(0.287)	1.849(0.284)	9.836(0.038)	0.736(0.174)
		0.50	0.647(0.135)	1.655(0.303)	9.758(0.029)	0.789(0.115)
		0.05	0.918(0.260)	1.879(0.334)	9.879(0.036)	0.847(0.283)

Figure 2: The true link functions (grey) and the estimated link functions (black) with $\beta_{(1)}^{*\top} = (5,5,5,5,5)$, and $\varepsilon \sim \mathsf{N}(0, 0.1)$, $n = 100, p = 10, q = 5, \tau = 0.95$, model 1 (left) with $h = 1.02$, model 2 (middle) with $h = 0.15$ and model 3 (right) with $h = 0.76$.



would be affected as well. Figure 5 and Figure 6 are the plots of the estimated link functions in these two cases.

Figure 3: The true link functions (grey) and the estimated link functions (black) with $\beta_{(1)}^{*\top} = (5,5,5,5,5)$, and $\varepsilon \sim \mathsf{N}(0, 0.1)$, $n = 100, p = 10, q = 5, \tau = 0.05$, model 1 (left) with $h = 0.78$, model 2 (middle) with $h = 0.12$ and model 3 (right) with $h = 0.78$.

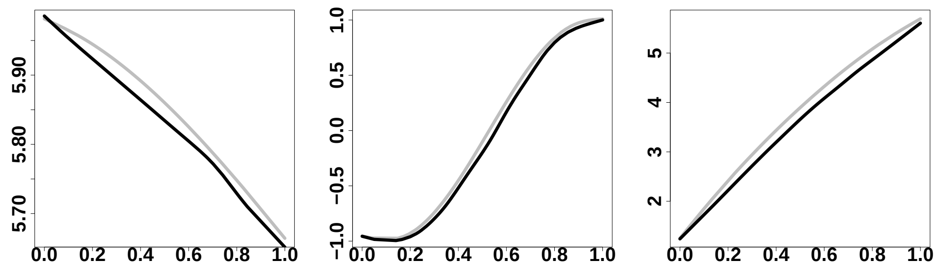


Figure 4: The true link functions (grey) and the estimated link functions (black) with $\beta_{(1)}^{*\top} = (5,5,5,5,5)$, and $\varepsilon \sim \mathsf{N}(0, 0.1)$, $n = 100, p = 10, q = 5, \tau = 0.5$, model 1 (left) with $h = 0.55$, model 2 (middle) with $h = 0.13$ and model 3 (right) with $h = 0.65$.

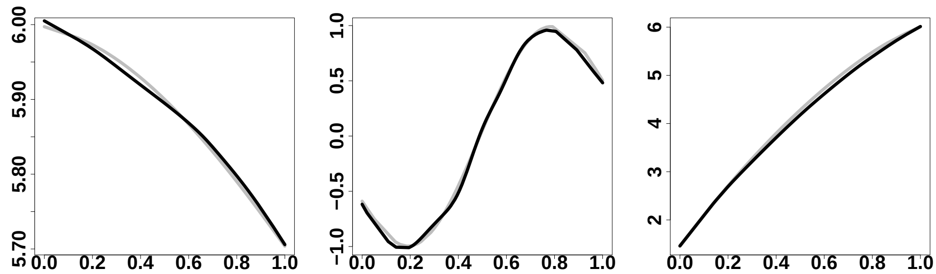


Table 2: Criteria evaluated with different models. Two different $\beta^*_{(1)}$: (a) $\beta^{*\top}_{(1)} = (5,4,3,2,1)$, (b) $\beta^{*\top}_{(1)} = (5,2,1,0.8,0.2)$ the error ε follows a $\mathsf{N}(0,0.1)$ distribution. In 10000 simulations we set $p = 10, q = 5, \tau = 0.95$. Standard deviations are given in brackets. *Dev*, *Acc*, *Angle* and their standard deviations are reported in 10^{-1}, *ASE* and its standard deviations are reported in 10^{-2}.

n	$g(\cdot)$	$\beta^*_{(1)}$	Dev	Acc	Angle	ASE
100	Model 1	(a)	1.402(0.351)	1.009(0.361)	9.735(0.071)	0.232(0.094)
		(b)	1.718(0.393)	1.313(0.391)	9.391(0.084)	0.353(0.119)
	Model 2	(a)	1.849(0.867)	7.367(0.944)	9.446(0.423)	1.451(0.852)
		(b)	2.304(0.913)	9.505(0.958)	9.341(0.556)	1.845(0.914)
	Model 3	(a)	0.406(0.256)	1.519(0.243)	9.643(0.066)	0.857(0.125)
		(b)	0.835(0.294)	1.781(0.289)	9.426(0.073)	0.906(0.136)
200	Model 1	(a)	1.318(0.368)	0.825(0.221)	9.756(0.062)	0.179(0.088)
		(b)	1.409(0.312)	0.956(0.252)	9.682(0.079)	0.302(0.073)
	Model 2	(a)	1.833(0.751)	5.126(0.936)	9.476(0.392)	1.338(0.701)
		(b)	2.257(0.887)	7.366(0.910)	9.385(0.460)	1.754(0.843)
	Model 3	(a)	0.389(0.231)	1.597(0.288)	9.632(0.052)	0.777(0.112)
		(b)	0.533(0.281)	1.624(0.290)	9.538(0.061)	0.864(0.129)
500	Model 1	(a)	1.012(0.287)	0.714(0.225)	9.846(0.061)	0.124(0.073)
		(b)	1.302(0.301)	0.854(0.245)	9.797(0.070)	0.287(0.061)
	Model 2	(a)	1.622(0.564)	5.024(0.821)	9.495(0.302)	1.204(0.592)
		(b)	2.176(0.636)	6.015(0.801)	9.452(0.363)	1.512(0.614)
	Model 3	(a)	0.361(0.211)	1.419(0.202)	9.781(0.029)	0.626(0.091)
		(b)	0.423(0.235)	1.612(0.236)	9.652(0.037)	0.751(0.111)

Figure 5: The true link functions (grey) and the estimated link functions (black) with $\beta_{(1)}^{*\top} = (5, 4, 3, 2, 1)$, and $\varepsilon \sim \mathsf{N}(0, 0.1)$, $n = 100, p = 10, q = 5, \tau = 0.95$, model 1 (left) with $h = 0.31$, model 2 (middle) with $h = 0.09$ and model 3 (right) with $h = 0.8$.

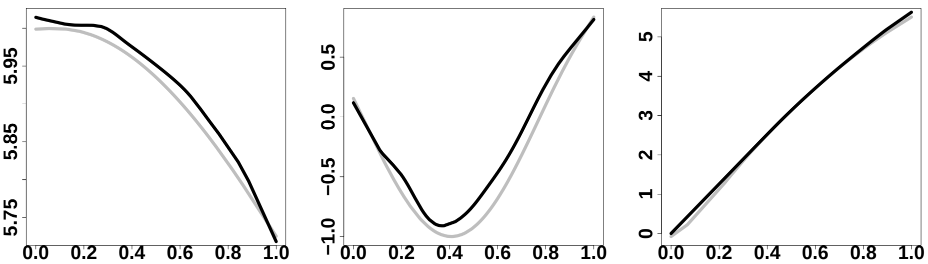


Figure 6: The true link functions (grey) and the estimated link functions (black) with $\beta_{(1)}^{*\top} = (5, 2, 1, 0.8, 0.2)$, and $\varepsilon \sim \mathsf{N}(0, 0.1)$, $n = 100, p = 10, q = 5, \tau = 0.95$, model 1 (left) with $h = 0.21$, model 2 (middle) with $h = 0.18$ and model 3 (right) with $h = 0.25$.

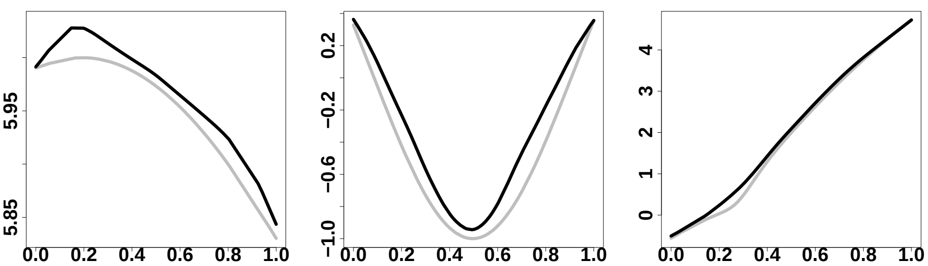
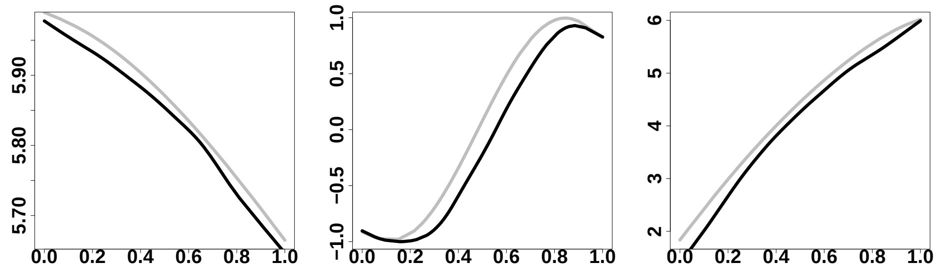


Table 3 shows the criteria evaluated under the $p > n$ case. Here $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$, the error ε follows a $\mathsf{N}(0, 0.1)$ distribution. In 10000 simulations we set $p = 200, q = 5, \tau = 0.05$. Standard deviations are given in brackets. We find that the errors are still moderate in the $p > n$ situation compared with Table 1. Figure 7 shows the graphs in this case.

Table 3: Criteria evaluated with different models $p \geq n$ case. $\beta_{(1)}^{*\top} = (5,5,5,5,5)$, the error ε follows a $\mathsf{N}(0, 0.1)$ distribution. In 10000 simulations we set $p = 200, q = 5, \tau = 0.05$. Standard deviations are given in brackets. *Dev*, *Acc*, *Angle* and their standard deviations are reported in 10^{-1}, *ASE* and its standard deviations are reported in 10^{-2}.

n	$g(\cdot)$	Dev	Acc	Angle	ASE
100	Model 1	1.880(0.753)	2.535(0.847)	9.303(0.157)	1.812(0.239)
	Model 2	2.859(0.954)	9.613(1.411)	9.035(0.835)	3.465(0.936)
	Model 3	1.554(0.635)	3.143(0.866)	9.265(0.095)	3.354(0.297)
200	Model 1	1.865(0.744)	1.818(0.724)	9.331(0.125)	1.103(0.233)
	Model 2	2.433(0.822)	8.499(1.222)	9.112(0.709)	2.224(0.931)
	Model 3	1.415(0.602)	2.001(0.713)	9.303(0.079)	2.915(0.203)

Figure 7: The true link functions (grey) and the estimated link functions (black) with $\beta_{(1)}^{*\top} = (5,5,5,5,5)$, and $\varepsilon \sim \mathsf{N}(0, 0.1)$, $n = 100, p = 200, q = 5, \tau = 0.05$, model 1 (left) with $h = 0.81$, model 2 (middle) with $h = 0.22$ and model 3 (right) with $h = 0.57$.



4.4 Composite L_1-L_2 Regression

In this subsection, a combined L_1 and L_2 loss is considered and thus, the corresponding optimization is formed as:

$$\arg\min_{\beta, g(\cdot)} \Big[\sum_{i=1}^{n} \mathsf{w}_1 |Y_i - g(X_i^\top \beta)| + \mathsf{w}_2 \sum_{i=1}^{n} \{Y_i - g(X_i^\top \beta)\}^2 \omega_i(\beta) + n \sum_{l=1}^{p} \gamma_\lambda(|\beta_l|)|\beta_l| \Big]. \quad (4.4)$$

It can be further formulated as

$$\arg\min_{\beta,g(\cdot)}\Big[\sum_{i=1}^{n}\{\mathsf{w}_1|Y_i-g(X_i^\top\beta)|^{-1}+\mathsf{w}_2\}|Y_i-g(X_i^\top\beta)|^2\omega_i(\beta)+n\sum_{l=1}^{p}\gamma_\lambda(|\beta_l|)|\beta_l|\Big]. \quad (4.5)$$

Let $Res_i^t \stackrel{\text{def}}{=} Y_i - \hat{g}^t(X_i^\top\hat{\beta}^t)$ be the residual at t-th step, and the final estimate can be acquired by the iteration between $g(\cdot)$ and β until convergence:

$$\arg\min_{\beta,g(\cdot)}\Big[\sum_{i=1}^{n}\{\mathsf{w}_1|Res_i^t|^{-1}+\mathsf{w}_2\}|Y_i-g(X_i^\top\beta)|^2\omega_i(\hat{\beta}^{(t)})+n\sum_{l=1}^{p}\gamma_\lambda(|\beta_l|)|\beta_l|\Big]. \quad (4.6)$$

Three different settings are conducted. The results are reported in Table 4. Figure 8 (the upper panel) shows the difference between the estimated and true $g(\cdot)$ functions. The level of estimation error is roughly the same as the previous level. Also the results would not change too much with respect to the error distributions and the increasing dimension of p, since only the dimension of q matters.

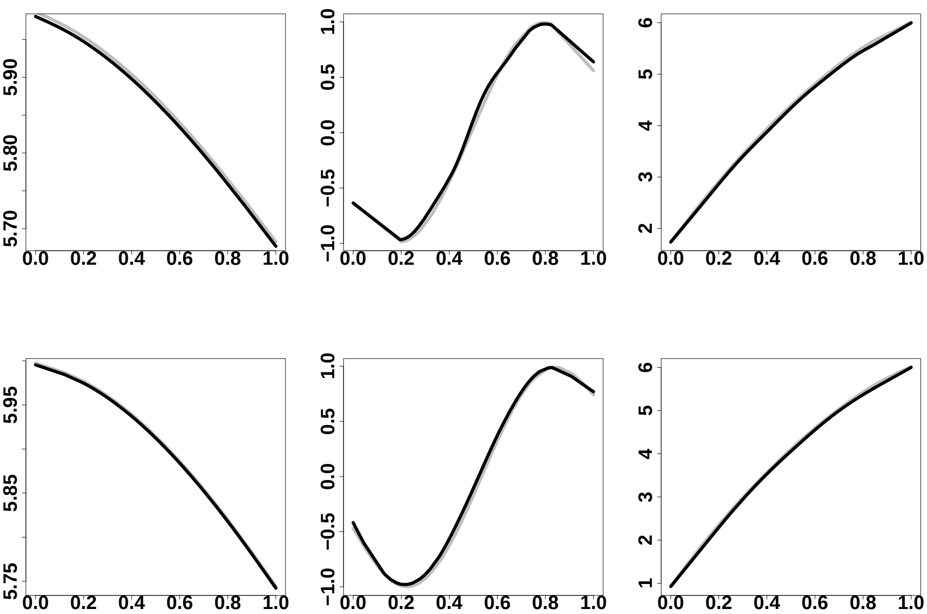
4.5 Composite L_1 Quantile Regression

We use Majorize-Minimization (MM) algorithm for a large scale regression problem. Table 5 shows the estimation quality. Compared with the results in Table 1, the estimation efficiency is improved, even in the case of $p > n$. Figure 8 presents the plots of the estimated link functions for different models using both the composite L_1 regression and the L_1-L_2 regression.

Table 4: Simulation results under sparsity, non-sparsity and large p cases. N means errors follow i.i.d. $\mathsf{N}(0, 0.1)$, t means t distribution with degree of 5. *Dev*, *Acc*, *Angle* and their standard deviations are reported in 10^{-1}, *ASE* and its standard deviations are reported in 10^{-2}.

n	model	settings	ε	Dev	Acc	Angle	ASE
100	Model 1	$p = 10, q = 2$	N	1.033(0.141)	1.037(0.231)	9.888(0.016)	0.223(0.031)
			t	1.223(0.230)	1.132(1.237)	9.860(0.021)	0.281(0.047)
		$p = 10, q = 7$	N	1.163(0.201)	1.219(0.211)	9.833(0.023)	0.290(0.049)
			t	1.444(0.232)	1.298(0.277)	9.805(0.050)	0.318(0.079)
		$p = 100, q = 5$	N	1.484(0.303)	1.624(1.426)	9.344(0.091)	0.473(0.216)
			t	1.576(0.365)	1.845(0.445)	9.311(0.106)	0.534(0.223)
	Model 2	$p = 10, q = 2$	N	1.134(0.277)	6.392(0.381)	9.399(0.125)	1.146(0.216)
			t	1.235(0.295)	6.442(0.412)	9.391(0.136)	1.241(0.227)
		$p = 10, q = 7$	N	1.323(0.346)	7.723(0.682)	9.281(0.287)	1.401(0.321)
			t	1.706(0.368)	7.953(0.704)	9.259(0.314)	1.577(0.361)
		$p = 100, q = 5$	N	1.207(0.483)	8.387(0.891)	9.230(0.359)	1.728(0.673)
			t	1.994(0.494)	8.543(0.903)	9.142(0.416)	1.751(0.701)
	Model 3	$p = 10, q = 2$	N	0.880(0.153)	1.254(0.143)	9.968(0.018)	0.550(0.091)
			t	1.077(0.175)	1.366(0.145)	9.951(0.023)	0.740(0.102)
		$p = 10, q = 7$	N	1.285(0.183)	1.553(0.197)	9.950(0.036)	0.838(0.127)
			t	1.334(0.195)	1.680(0.257)	9.947(0.048)	0.843(0.139)
		$p = 100, q = 5$	N	1.369(0.235)	2.023(0.636)	9.377(0.054)	1.304(0.182)
			t	1.494(0.383)	2.293(0.652)	9.344(0.063)	1.880(0.197)
500	Model 1	$p = 10, q = 2$	N	1.203(0.132)	0.999(0.193)	9.898(0.010)	0.214(0.031)
			t	1.338(0.147)	1.019(0.201)	9.835(0.009)	0.237(0.035)
		$p = 10, q = 7$	N	1.208(0.166)	1.118(0.218)	9.882(0.012)	0.309(0.046)
			t	1.457(0.178)	1.236(0.242)	9.802(0.018)	0.306(0.072)
		$p = 100, q = 5$	N	1.434(0.183)	1.478(0.396)	9.323(0.063)	0.332(0.152)
			t	1.482(0.217)	1.646(0.401)	9.315(0.088)	0.491(0.179)
	Model 2	$p = 10, q = 2$	N	1.036(0.222)	6.021(0.311)	9.598(0.133)	1.026(0.211)
			t	1.133(0.290)	6.129(0.411)	9.435(0.169)	1.198(0.231)
		$p = 10, q = 7$	N	1.152(0.229)	7.069(0.518)	9.402(0.212)	1.364(0.288)
			t	1.468(0.289)	7.188(0.625)	9.382(0.268)	1.473(0.306)
		$p = 100, q = 5$	N	1.773(0.461)	8.327(0.794)	9.207(0.281)	1.691(0.652)
			t	1.872(0.489)	8.376(0.864)	9.141(0.299)	1.706(0.691)
	Model 3	$p = 10, q = 2$	N	0.746(0.102)	1.023(0.103)	9.590(0.013)	0.498(0.081)
			t	0.865(0.169)	1.215(0.128)	9.481(0.020)	0.502(0.099)
		$p = 10, q = 7$	N	0.992(0.187)	1.436(0.137)	9.487(0.029)	0.578(0.112)
			t	1.003(0.193)	1.478(0.186)	9.459(0.032)	0.624(0.131)
		$p = 100, q = 5$	N	1.209(0.203)	1.646(0.468)	9.381(0.041)	0.847(0.165)
			t	1.402(0.353)	2.219(0.579)	9.343(0.053)	0.781(0.194)

Figure 8: Plot of the true function $g(\cdot)$ (grey) and the estimation (black) with $n = 100, p = 10, q = 5$ and $\varepsilon \sim \mathsf{N}(0, 0.1)$ in different $g(\cdot)$ functions. L_1-L_2 regression, $h = 0.6, 0.3, 0.4$ (upper pannel), composite quantile $h = 0.5, 0.2, 0.5$ (lower panel)



5 Application

In this section, we apply the proposed methodology to analyze risk for a specific firm conditioning on macro and other firm variables. More specifically, for small financial firms, we aim to detect the contagion effects and the potential risk contributions from larger firms and other market variables. As a result one identifies a risk index, which is expressed as a linear combination, composed of selected large firm returns and market prudential variables.

Table 5: Simulation results for Composite L_1 Quantile Regression. N means errors follow i.i.d. $\mathsf{N}(0, 0.1)$, t means t distribution with degree of 5. *Dev*, *Acc*, *Angle*, *Error* and their standard deviations are reported in 10^{-1}, *ASE* and its standard deviations are reported in 10^{-2}.

n	model	settings	ε	Dev	Acc	Angle	ASE
100	Model 1	$p = 10, q = 2$	N	2.638(0.053)	0.774(0.149)	9.993(0.013)	0.142(0.022)
			t	1.038(0.125)	0.899(0.156)	9.991(0.014)	0.145(0.031)
		$p = 30, q = 3$	N	1.148(0.141)	1.072(0.175)	9.828(0.011)	0.169(0.043)
			t	1.166(0.106)	1.197(0.193)	9.576(0.012)	0.257(0.063)
		$p = 120, q = 5$	N	1.183(0.186)	1.207(0.191)	9.421(0.040)	0.332(0.114)
			t	1.336(0.215)	1.219(0.201)	9.403(0.063)	0.367(0.119)
	Model 2	$p = 10, q = 2$	N	1.119(0.213)	4.001(0.282)	9.592(0.101)	1.112(0.212)
			t	1.215(0.241)	4.086(0.323)	9.499(0.117)	1.244(0.218)
		$p = 30, q = 3$	N	1.335(0.252)	5.154(0.393)	9.595(0.132)	1.304(0.311)
			t	1.359(0.282)	5.538(0.462)	9.583(0.168)	1.383(0.381)
		$p = 120, q = 5$	N	1.742(0.289)	6.703(0.504)	9.382(0.202)	1.453(0.412)
			t	1.946(0.320)	7.335(0.611)	9.363(0.310)	1.626(0.503)
	Model 3	$p = 10, q = 2$	N	0.415(0.086)	1.007(0.100)	9.974(0.011)	0.426(0.041)
			t	0.512(0.093)	1.032(0.113)	9.968(0.013)	0.493(0.059)
		$p = 30, q = 3$	N	0.841(0.143)	1.167(0.139)	9.965(0.013)	0.528(0.060)
			t	0.953(0.153)	1.235(0.155)	9.962(0.022)	0.560(0.069)
		$p = 120, q = 5$	N	0.883(0.161)	1.357(0.168)	9.575(0.034)	0.892(0.104)
			t	0.903(0.233)	1.946(0.273)	9.553(0.044)	0.949(0.113)
500	Model 1	$p = 10, q = 2$	N	0.935(0.102)	0.609(0.102)	9.998(0.003)	0.114(0.018)
			t	1.026(0.134)	0.774(0.124)	9.992(0.005)	0.125(0.029)
		$p = 30, q = 3$	N	1.132(0.142)	0.852(0.138)	9.993(0.005)	0.133(0.033)
			t	1.148(0.116)	0.945(0.165)	9.991(0.006)	0.174(0.049)
		$p = 120, q = 5$	N	1.157(0.125)	1.144(0.185)	9.543(0.030)	0.247(0.110)
			t	1.275(0.166)	1.232(0.196)	9.572(0.046)	0.303(0.115)
	Model 2	$p = 10, q = 2$	N	1.104(0.206)	3.908(0.260)	9.691(0.053)	1.009(0.116)
			t	1.185(0.214)	4.105(0.273)	9.685(0.055)	1.216(0.151)
		$p = 30, q = 3$	N	1.286(0.219)	4.239(0.294)	9.552(0.050)	1.309(0.216)
			t	1.294(0.278)	5.046(0.347)	9.504(0.127)	1.316(0.231)
		$p = 120, q = 5$	N	1.727(0.246)	5.675(0.405)	9.459(0.134)	1.448(0.317)
			t	1.824(0.289)	5.856(0.581)	9.443(0.168)	1.497(0.413)
	Model 3	$p = 10, q = 2$	N	0.380(0.076)	0.996(0.087)	9.993(0.010)	0.391(0.040)
			t	0.508(0.087)	1.022(0.116)	9.990(0.016)	0.446(0.048)
		$p = 30, q = 3$	N	0.763(0.092)	1.154(0.125)	9.982(0.016)	0.514(0.051)
			t	0.846(0.104)	1.265(0.142)	9.971(0.020)	0.546(0.064)
		$p = 120, q = 5$	N	0.966(0.113)	1.843(0.193)	9.833(0.022)	0.768(0.087)
			t	1.124(0.235)	1.898(0.237)	9.742(0.031)	0.830(0.104)

5.1 Data and Risk Calibration

The firm data are selected according to the ranking of NASDAQ. We take as an example, city national corp. (CYN) as our dependent variable. The remaining 199 financial institutions together with 7 lagged macro variables are chosen as covariates. The list of these firms comes from the website: http://www.nasdaq.com/screening

/companies-by-industry.aspx?industry=Finance. The daily stock prices of these 200 firms are from Yahoo Finance for the period from January 5, 2006 to October 30, 2015. The descriptive statistics of the company, the description of the macro variables and the list of the firms (Table 8 to Table 10) can be found in the Appendices. To evaluate the risk exposure of the firm CYN, we adopt a modified two-step quantile regression procedure which involves our quantile single index model in the second step. The first one is a quantile regression to calculate the VaR of all the covariates respectively. For this propose, one performs QR of log returns of each covariate on all the lagged macro variables:

$$X_{i,t} = \alpha_i + \gamma_i^\top M_{t-1} + \varepsilon_{i,t}, \tag{5.1}$$

where $X_{i,t}$ represents the asset return of financial institution i at time t. Then the VaR of each firm with $F^{-1}_{\varepsilon_{i,t}}(\tau|M_{t-1}) = 0$ is obtained by:

$$\widehat{VaR}^{\tau}_{i,t} = \widehat{\alpha}_i + \widehat{\gamma}_i^\top M_{t-1}, \tag{5.2}$$

Now the second regression is performed using the proposed MACE method. The response variable is log returns of CYN, and the explanatory variables are potential risk factors which includes the log returns of those covariates and the lagged macro variables:

$$X_{j,t} = g(S^\top \beta_{j|S}) + \varepsilon_{j,t}, \tag{5.3}$$

where $S \overset{def}{=} [M_{t-1}, R]$, R is a vector of log returns for different firms. $\beta_{j|S}$ is a $p \times 1$ vector. A detailed list of factors can be found in Table 8 to Table 10 in the

Appendix.

With $F^{-1}_{\varepsilon_{j,t}}(\tau|S) = 0$ the CoVaR for firm j is estimated as:

$$\widehat{CoVaR}^{\tau}_{j|\widehat{S}} \;=\; \widehat{g}(\widehat{S}^{\top}\widehat{\beta}_{j|S}), \tag{5.4}$$

where $\widehat{S} \stackrel{def}{=} [M_{t-1}, \widehat{V}]$, with $\widehat{V}$ as the estimated VaR in (5.2).

To evaluate the preciseness of the proposed CoVaR risk measure, we launch a back-testing procedure. First, one calculates the violations over time, which is defined as the days on which the log returns are lower than the estimated VaR or CoVaR:

$$\hat{I}_{i,t} = \begin{cases} 1, & X_{i,t} < \widehat{VaR}^{\tau}_{i,t}; \\ 0, & otherwise, \end{cases}$$

where theoretically $I_{i,t} - \tau$ should be a martingale difference sequence. Then we apply one version of the CaViaR test, see Berkowitz et al. (2011), which adopts a logit model:

$$I_{i,t} = \alpha + \beta_1 I_{i,t-1} + \beta_2 VaR_{i,t} + u_{i,t},$$

where $u_{i,t}$ has a logistic distribution. The Wald test is then applied with null hypothesis: $\widehat{\beta}_1 = \widehat{\beta}_2 = 0$, see Franke et al. (2004) for more details.

5.2 Results

We use a moving window size of $n = 126$ (corresponding approximately to half a year of trading days) to calculate VaR of the log returns for the 199 firms, macro variables, and CYN. Figure 9 and Figure 10 show one illustration of the estimated VaR of JPM (one covariate in the second step) and CYN respectively. It can be seen that the estimated VaR traces the low values of returns closely, and becomes more volatile when the volatility of the returns is large.

Figure 9: Log returns of JPM (grey) and VaR of log returns of JPM (black), $\tau = 0.05$, $T = 2335$, window size $n = 126$, refer to (5.2).

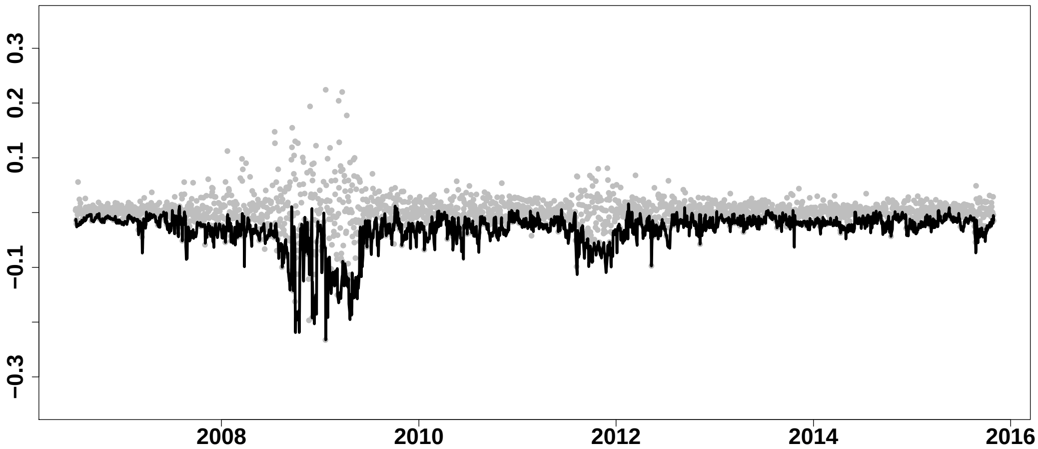
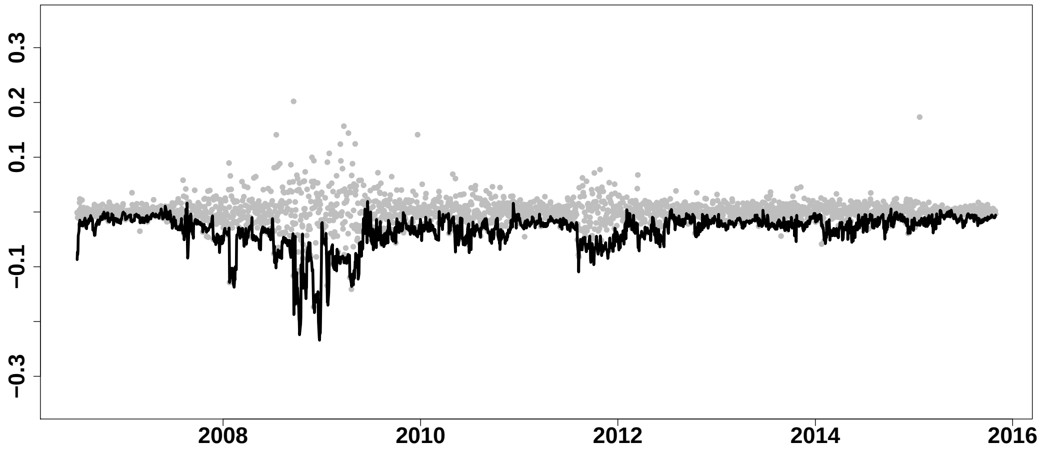


Figure 10: Log returns of CYN (grey) and VaR of log returns of CYN (black), $\tau = 0.05$, $T = 2335$, window size $n = 126$, refer to (5.2).



With the VaR estimation in previous step, we show further the estimation of the CoVaR for CYN. The estimation is conducted in a moving window of size 126. Our technique is applied with $\tau = 0.05$. We use $p = 206$ covariates, and the CoVaR for CYN is estimated with different variables selected in each window. Figure 11 shows the estimation results. We further summarize the selected variables in different windows.

Figure 12 summarizes the selection frequency of the firms and macro variables for all the windows. The variable 187, "Radian Group Inc. (RDN)" is the most frequently selected variable with frequency 752, which indicates the most relevant risk driver for CYN.

To compare the performance of our proposed measure with existing measures, we further apply CaViaR test for backtesting. Figure 13 shows the $\hat{I}_{i,t}$ sequence of

Figure 11: Log returns of CYN (grey) and the estimated CoVaR (black), $\tau = 0.05$, $T = 2335$, window size $n = 126$, refer to (5.4).

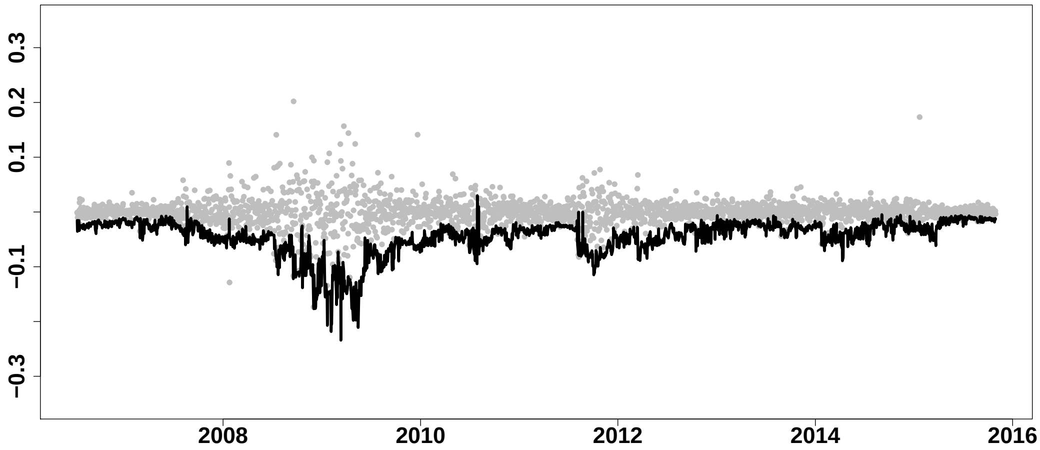
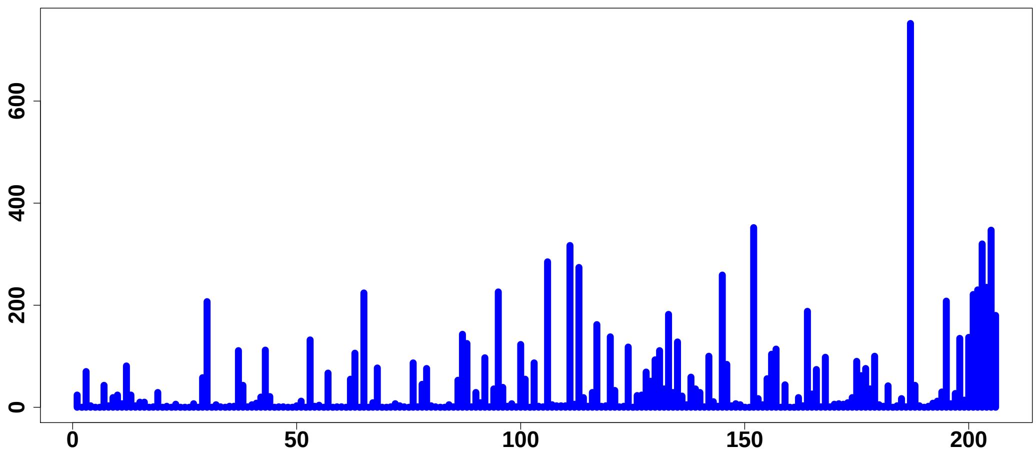


Figure 12: The frequency of the firms and macro variables. The X-axis: $1 - 206$ variables, and the Y-axis: the frequency of the variables selected in the moving window estimation. The variable 187, i.e. "Radian Group Inc. (RDN)" is the most frequently selected variable with frequency 752.



$\widehat{VaR}$ (estimated value at risk measure) of CYN, there are a total of 23 violations. With $T = 2335$, the violation proportion is then $\widehat{\tau} = 0.009$.

Table 6: The p-values for CaViaR test for $\widehat{VaR}$, $\widehat{CoVaR}_L$ and $\widehat{CoVaR}_{SIM}$ for CYN, $T = 2335$ in overall period $(20060710 - 20151030)$ and crisis period $(20080915 - 20100208)$.

p-value	Overall	Crisis
$\widehat{VaR}$	1.2×10^{-6}	0.99
$\widehat{CoVaR}_L$	0.01	3.2×10^{-5}
$\widehat{CoVaR}_{SIM}$	0.46	0.93

From Figure 14 we get the $\hat{I}_{i,t}$ sequence of $\widehat{CoVaR}$ of CYN, there are 28 violations out of $T = 2335$, which means $\widehat{\tau} = 0.012$.

The p-values of the CaViaR tests are then shown in Table 6, in which we compare our measure $\widehat{CoVaR}_{SIM}$ (CoVaR estimated from single index model) with the measure attained solely by doing linear quantile variable selection, i.e. $\widehat{CoVaR}_L$, see, for example, Belloni et al. (2011). For the overall period, only for $\widehat{CoVaR}_{SIM}$, the null hypothesis can not be rejected. Therefore, $\widehat{VaR}$ and $\widehat{CoVaR}_L$ algorithms do not perform so well in an overall period. During crisis times, the null hypothesis of $\widehat{VaR}$ and $\widehat{CoVaR}_{SIM}$ can not be rejected, therefore both $\widehat{VaR}$ and $\widehat{CoVaR}_{SIM}$ algorithms perform well during the crisis periods, but $\widehat{CoVaR}_L$'s performance is not favorable.

6 Appendices

6.1 Proof

Please find in the supplementary materials.

Figure 13: The violations (i.e. $\{t : \hat{I}_{i,t} = 1\}$) of $\widehat{VaR}$ for CYN(the dots above), in total 23 violations, $T = 2335$, $\widehat{\tau} = 0.009$.

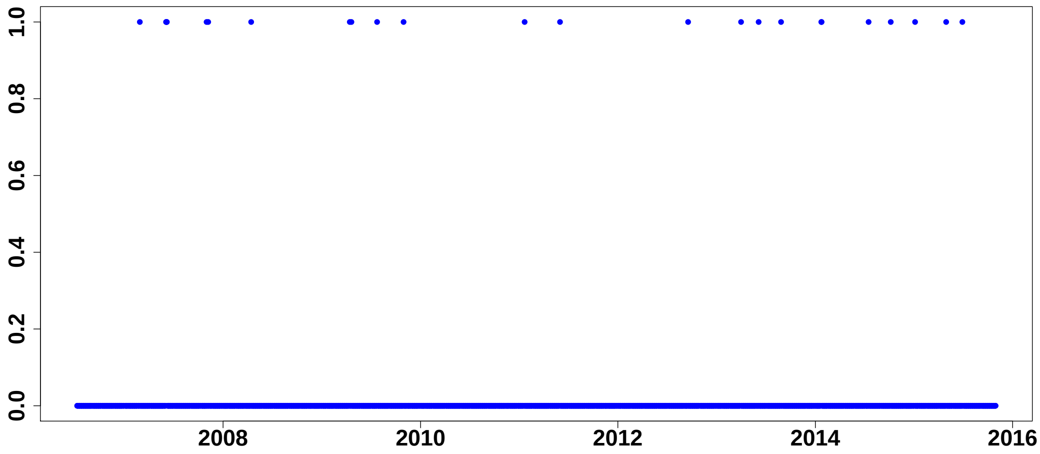
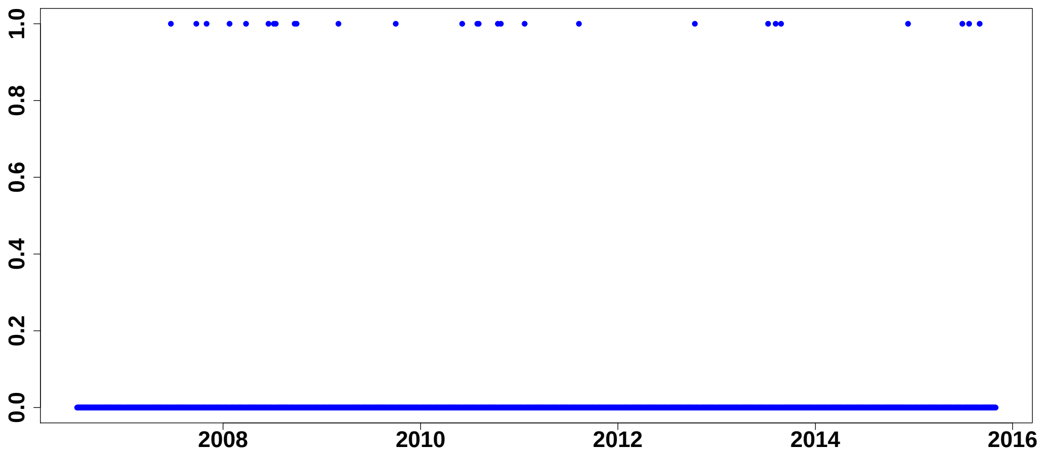


Figure 14: The violations (i.e. $\{t : \hat{I}_{i,t} = 1\}$) of $\widehat{CoVaR}_{sim}$ of CYN(the dots above), in total 28 violations, $T = 2335$, $\widehat{\tau} = 0.012$.



6.2 Application

The macro variables are the same as suggested by Adrian and Brunnermeier (2011) and Chao et al. (2012). The macro variables and the corresponding source are listed as follows:

1. VIX, which measures the implied volatility in the market.
2. The short term liquidity spread, which is calculated by the difference between the 3-month Treasury repo rate and 3-month Treasury constant maturities.
3. The daily change in the 3-month Treasury constant maturities, which can be defined as the difference between the current day and the previous day of 3-month Treasury constant maturities.
4. The change in the slope of the yield curve, which is defined by the difference between the 10 year Treasury constant maturities and the 3-month Treasury constant maturities.
5. The change in the credit spread between 10 years BAA corporate bonds and the 10 years Treasury constant maturities.
6. The daily S&P500 index returns.
7. The daily Dow Jones U.S. Real Estate index returns.

The repo data can be obtained from the Datastream database, and the 10 year Treasury constant maturities and BAA corporate bonds data can be found in the website of the Federal Reserve Board H.15:

http://www.federalreserve.gov/releases/h15/data.htm. Other data are available in Yahoo Finance. The macro variables' data are available from January 4, 2006 to October 29, 2015 with a daily frequency.

Table 7: Descriptive statistics of CYN

	Mean	SD	Skewness	Kurtosis
Overall period	-0.0001	0.0237	0.2821	14.0036
In crisis	-9.247×10^{-5}	0.0312	0.1326	8.9544

Table 7 shows the descriptive statistics of this series. The mean of CYN in the the overall period (i.e. January 06, 2006 to October 30, 2015) is -0.000118, which higher than it (-0.000092) in the crisis period (i.e. from September 15, 2008 to February 8, 2010). The volatility in the crisis period is higher than it in the overall period. The p values of the Jarque Bera test indicates that log returns of CYN are not normally distributed. We also perform a unit root test which suggests that the log returns of CYN are stationary. The mentioned two test results for the other firms show that all these series are not normally distributed, but are likely to be stationary.

Table 8: The financial firms

The financial frims	
1. Wells Fargo & Co (WFC)	15. Franklin Resources Inc. (BEN)
2. JP Morgan Chase & Co (JPM)	16. The Travelers Companies, Inc. (TRV)
3. Bank of America Corp (BAC)	17. AFLAC Inc. (AFL)
4. Citigroup Inc (C)	18. Prudential Financial, Inc. (PRU)
5. American Express Company (AXP)	19. State Street Corporation (STT)
6. U.S. Bancorp (USB)	20. The Chubb Corporation (CB)
7. The Goldman Sachs Group, Inc. (GS)	21. BB&T Corporation (BBT)
8. American International Group, Inc. (AIG)	22. Marsh & McLennan Companies, Inc. (MMC)
9. MetLife, Inc. (MET)	23. The Allstate Corporation (ALL)
10. Capital One Financial Corp. (COF)	24. Aon plc (AON)
11. BlackRock, Inc. (BLK)	25. CME Group Inc. (CME)
12. Morgan Stanley (MS)	26. The Charles Schwab Corporation (SCHW)
13. PNC Financial Services Group Inc. (PNC)	27. T. Rowe Price Group, Inc. (TROW)
14. The Bank of New York Mellon Corporation (BK)	28. Loews Corporation (L)
29. SunTrust Banks, Inc. (STI)	44. Lincoln National Corporation (LNC)
30. Fifth Third Bancorp (FITB)	45. Affiliated Managers Group Inc. (AMG)
31. Progressive Corp. (PGR)	46. Cincinnati Financial Corp. (CINF)
32. M&T Bank Corporation (MTB)	47. Equifax Inc. (EFX)
33. Ameriprise Financial Inc. (AMP)	48. Alleghany Corp. (Y)
34. Northern Trust Corporation (NTRS)	49. Unum Group (UNM)
35. Invesco Ltd. (IVZ)	50. Comerica Incorporated (CMA)
36. Moody's Corp. (MCO)	51. W.R. Berkley Corporation (WRB)
37. Regions Financial Corp. (RF)	52. Fidelity National Financial, Inc. (FNF)
38. The Hartford Financial Services Group, Inc. (HIG)	53. Huntington Bancshares Incorporated (HBAN)
39. TD Ameritrade Holding Corporation (AMTD)	54. Raymond James Financial Inc. (RJF)
40. Principal Financial Group Inc. (PFG)	55. Torchmark Corp. (TMK)
41. SLM Corporation (SLM)	56. Markel Corp. (MKL)
42. KeyCorp (KEY)	57. Ocwen Financial Corp. (OCN)
43. CNA Financial Corporation (CNA)	58. Arthur J Gallagher & Co. (AJG)

Table 9: The financial firms

The financial firms	
59. Hudson City Bancorp, Inc. (HCBK)	74. Commerce Bancshares, Inc. (CBSH)
60. People's United Financial Inc. (PBCT)	75. Signature Bank (SBNY)
61. SEI Investments Co. (SEIC)	76. Jefferies Group, Inc. (JEF)
62. Nasdaq OMX Group Inc. (NDAQ)	77. Rollins Inc. (ROL)
63. Brown & Brown Inc. (BRO)	78. Morningstar Inc. (MORN)
64. BOK Financial Corporation (BOKF)	79. East West Bancorp, Inc. (EWBC)
65. Zions Bancorp. (ZION)	80. Waddell & Reed Financial Inc. (WDR)
66. HCC Insurance Holdings Inc. (HCC)	81. Old Republic International Corporation (ORI)
67. Eaton Vance Corp. (EV)	82. ProAssurance Corporation (PRA)
68. Erie Indemnity Company (ERIE)	83. Assurant Inc. (AIZ)
69. American Financial Group Inc. (AFG)	84. Hancock Holding Company (HBHC)
70. Dun & Bradstreet Corp. (DNB)	85. First Niagara Financial Group Inc. (FNFG)
71. White Mountains Insurance Group, Ltd. (WTM)	86. SVB Financial Group (SIVB)
72. Cullen-Frost Bankers, Inc. (CFR)	87. First Horizon National Corporation (FHN)
73. Legg Mason Inc. (LM)	88. E-TRADE Financial Corporation (ETFC)
89. SunTrust Banks, Inc. (STI)	104. Valley National Bancorp (VLY)
90. Mercury General Corporation (MCY)	105. KKR Financial Holdings LLC (KFN)
91. Associated Banc-Corp (ASBC)	106. Synovus Financial Corporation (SNV)
92. Credit Acceptance Corp. (CACC)	107. Texas Capital BancShares Inc. (TCBI)
93. Protective Life Corporation (PL)	108. American National Insurance Co. (ANAT)
94. Federated Investors, Inc. (FII)	109. Washington Federal Inc. (WAFD)
95. CNO Financial Group, Inc. (CNO)	110. First Citizens Bancshares Inc. (FCNCA)
96. Popular, Inc. (BPOP)	111. Kemper Corporation (KMPR)
97. Bank of Hawaii Corporation (BOH)	112. UMB Financial Corporation (UMBF)
98. Fulton Financial Corporation (FULT)	113. Stifel Financial Corp. (SF)
99. AllianceBernstein Holding L.P. (AB)	114. CapitalSource Inc. (CSE)
100. TCF Financial Corporation (TCB)	115. Portfolio Recovery Associates Inc. (PRAA)
101. Susquehanna Bancshares, Inc. (SUSQ)	116. Janus Capital Group, Inc. (JNS)
102. Capitol Federal Financial, Inc. (CFFN)	117. MBIA Inc. (MBI)
103. Webster Financial Corp. (WBS)	118. Healthcare Services Group Inc. (HCSG)

Table 10: The financial firms

The financial firms	
119. The Hanover Insurance Group Inc. (THG)	134. BancorpSouth, Inc. (BXS)
120. F.N.B. Corporation (FNB)	135. Privatebancorp Inc. (PVTB)
121. FirstMerit Corporation (FMER)	136. United Bankshares Inc. (UBSI)
122. FirstMerit Corporation (FMER)	137. Old National Bancorp. (ONB)
123. RLI Corp. (RLI)	138. International Bancshares Corporation (IBOC)
124. StanCorp Financial Group Inc. (SFG)	139. First Financial Bankshares Inc. (FFIN)
125. Trustmark Corporation (TRMK)	140. Westamerica Bancorp. (WABC)
126. IberiaBank Corp. (IBKC)	141. Northwest Bancshares, Inc. (NWBI)
127. Cathay General Bancorp (CATY)	142. Bank of the Ozarks, Inc. (OZRK)
128. National Penn Bancshares Inc. (NPBC)	143. Huntington Bancshares Incorporated (HBAN)
129. Nelnet, Inc. (NNI)	144. Euronet Worldwide Inc. (EEFT)
130. Wintrust Financial Corporation (WTFC)	145. Community Bank System Inc. (CBU)
131. Umpqua Holdings Corporation (UMPQ)	146. CVB Financial Corp. (CVBF)
132. GAMCO Investors, Inc. (GBL)	147. MB Financial Inc. (MBFI)
133. Sterling Financial Corp. (STSA)	148. ABM Industries Incorporated (ABM)
149. Glacier Bancorp Inc. (GBCI)	164. Citizens Republic Bancorp, Inc (CRBC)
150. Selective Insurance Group Inc. (SIGI)	165. Horace Mann Educators Corp. (HMN)
151. Park National Corp. (PRK)	166. DFC Global Corp. (DLLR)
152. Flagstar Bancorp Inc. (FBC)	167. Navigators Group Inc. (NAVG)
153. FBL Financial Group Inc. (FFG)	168. Boston Private Financial Holdings, Inc. (BPFH)
154. Astoria Financial Corporation (AF)	169. American Equity Investment Life Holding Co. (AEL)
155. World Acceptance Corp. (WRLD)	170. BlackRock Limited Duration Income Trust (BLW)
156. First Midwest Bancorp Inc. (FMBI)	171. Columbia Banking System Inc. (COLB)
157. PacWest Bancorp (PACW))	172. Safety Insurance Group Inc. (SAFT)
158. First Financial Bancorp. (FFBC)	173. National Financial Partners Corp. (NFP)
159. BBCN Bancorp, Inc. (BBCN)	174. NBT Bancorp, Inc. (NBTB)
160. Provident Financial Services, Inc. (PFS)	175. Tower Group Inc. (TWGP)
161. FBL Financial Group Inc. (FFG)	176. Encore Capital Group, Inc. (ECPG)
162. WisdomTree Investments, Inc. (WETF)	177. Pinnacle Financial Partners Inc. (PNFP)
163. Hilltop Holdings Inc. (HTH)	178. First Commonwealth Financial Corp. (FCF)
179. BancFirst Corporation (BANF)	190. Berkshire Hills Bancorp Inc. (BHLB)
180. Independent Bank Corp. (INDB)	191. Brookline Bancorp, Inc. (BRKL)
181. Infinity Property and Casualty Corp. (IPCC)	192. National Western Life Insurance Company (NWLI)
182. Central Pacific Financial Corp. (CPF)	193. Tompkins Financial Corporation (TMP)
183. Kearny Financial Corp. (KRNY)	194. BGC Partners, Inc. (BGCP)
184. Chemical Financial Corporation (CHFC)	195. Epoch Investment Partners, Inc. (EPHC)
185. Banner Corporation (BANR)	196. United Fire Group, Inc (UFCS)
186. State Auto Financial Corp. (STFC)	197. 1st Source Corporation (SRCE)
187. Radian Group Inc. (RDN)	198. Citizens Inc. (CIA)
188. SCBT Financial Corporation (SCBT)	199. S&T Bancorp Inc. (STBA)
189. WesBanco Inc. (WSBC)	

References

Adrian, T. and Brunnermeier, M. K. (2011). CoVaR. Technical report, National Bureau of Economic Research.

Belloni, A., Chernozhukov, V., et al. (2011). L1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, 39(1):82–130.

Berkowitz, J., Christoffersen, P., and Pelletier, D. (2011). Evaluating value-at-risk models with desk-level data. *Management Science*, 57(12):2213–2227.

Bradic, J., Fan, J., and Wang, W. (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3):325–349.

Chao, S.-K., Härdle, W. K., and Wang, W. (2012). Quantile regression in risk calibration. Technical report, SFB 649 discussion paper.

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1):101.

Franke, J., Härdle, W., and Hafner, C. M. (2004). *Statistics of financial markets*, volume 2. Springer Heidelberg, Germany.

Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84(408):986–995.

Hautsch, N., Schaumburg, J., and Schienle, M. (2014). Financial network systemic risk contributions. *Review of Finance, forthcoming.*

Huber, P. J. (1985). Projection pursuit. *The annals of Statistics*, pages 435–475.

Kai, B., Li, R., and Zou, H. (2010). Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):49–69.

Kai, B., Li, R., and Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Annals of Statistics*, 39(1):305.

Kong, E. and Xia, Y. (2007). Variable selection for the single-index model. *Biometrika*, 94(1):217–229.

Kong, E. and Xia, Y. (2012). A single-index quantile regression model and its estimation. *Econometric Theory*, 28(04):730–768.

Leng, C., Xia, Y., and Xu, J. (2008). An adaptive estimation method for semiparametric models and dimension reduction. *WSPC-Proceedings.*

Li, Y. and Zhu, J. (2008). L1-norm quantile regression. *Journal of Computational and Graphical Statistics*, 17(1).

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90(432):1257–1270.

Schnabel, S. K. and Eilers, P. H. (2009). Optimal expectile smoothing. *Computational Statistics & Data Analysis*, 53(12):4168–4177.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Wang, H. J., Li, D., and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107(500):1453–1464.

Wang, Q. and Yin, X. (2008). A nonlinear multi-dimensional variable selection method for high dimensional data: Sparse MAVE. *Computational Statistics & Data Analysis*, 52(9):4512–4520.

Wu, T. Z., Yu, K., and Yu, Y. (2010). Single-index quantile regression. *Journal of Multivariate Analysis*, 101(7):1607–1621.

Xia, Y., Tong, H., Li, W., and Zhu, L.-X. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):363–410.

Yu, K. and Jones, M. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, 93(441):228–237.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zeng, P., He, T., and Zhu, Y. (2012). A lasso-type approach for estimation and variable selection in single index models. *Journal of Computational and Graphical Statistics*, 21(1):92–109.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.

Zou, H., Yuan, M., et al. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, 36(3):1108–1126.