

Fan, Yan; Härdle, Wolfgang Karl; Wang, Weining; Zhu, Lixing

**Working Paper**

## Composite quantile regression for the single-index model

SFB 649 Discussion Paper, No. 2013-010

**Provided in Cooperation with:**

Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

*Suggested Citation:* Fan, Yan; Härdle, Wolfgang Karl; Wang, Weining; Zhu, Lixing (2013) : Composite quantile regression for the single-index model, SFB 649 Discussion Paper, No. 2013-010, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:

<http://hdl.handle.net/10419/79568>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Composite Quantile Regression for the Single- Index Model

Yan Fan\*

Wolfgang Karl Härdle\*\*

Weining Wang\*\*

Lixing Zhu\*\*\*



\* Renmin University of China, Beijing, China

\*\* Humboldt-Universität zu Berlin, Germany

\*\*\* Hong Kong Baptist University

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>

ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin  
Spandauer Straße 1, D-10178 Berlin



# COMPOSITE QUANTILE REGRESSION FOR THE SINGLE-INDEX MODEL\*

BY YAN FAN<sup>†</sup>, WOLFGANG KARL HÄRDLE<sup>‡</sup>, WEINING WANG<sup>‡</sup> AND LIXING ZHU<sup>§</sup>

*Renmin University of China, Beijing, China<sup>†</sup>, Humboldt-Universität zu Berlin, Germany<sup>‡</sup> and Hong Kong Baptist University, Hong Kong, China<sup>§</sup>*

*Abstract* Quantile regression is in the focus of many estimation techniques and is an important tool in data analysis. When it comes to nonparametric specifications of the conditional quantile (or more generally tail) curve one faces, as in mean regression, a dimensionality problem. We propose a projection based single index model specification. For very high dimensional regressors  $X$  one faces yet another dimensionality problem and needs to balance precision vs. dimension. Such a balance may be achieved by combining semiparametric ideas with variable selection techniques.

**1. Introduction.** Regression between response  $Y$  and covariates  $X$  is a standard element of statistical data analysis. When the regression function is supposed to be estimated in a nonparametric context, the dimensionality of  $X$  plays a crucial role. Among the many dimension reduction techniques the single index approach has a unique feature: the index that yields interpretability and low dimension simultaneously. In the case of ultra high dimensional regressors  $X$  though it suffers, as any regression method, from singularity issues. Efficient variable selection is here the strategy to employ. Specifically we consider a composite regression with general weighted loss and possibly ultra high dimensional variables. Our setup is general, and includes quantile, expectile (and therefore mean) regression. We offer theoretical properties and demonstrate our method with applications to firm risk analysis in a CoVaR context.

Quantile regression (QR) is one of the major statistical tools and is “gradually developing into a comprehensive strategy for completing the regression prediction” [13]. In many fields of applications like quantitative finance, econometrics, marketing and also in medical and biological sciences, QR is a fundamental element for data analysis, modeling and

---

\*The financial support from the Deutsche Forschungsgemeinschaft via CRC 649 “Ökonomisches Risiko”, Humboldt-Universität zu Berlin, and the University Grants Council of Hong Kong is gratefully acknowledged. We also gratefully acknowledge the funding from DAAD ID 50746311.

*Keywords and phrases:* Quantile Single-index Regression, Minimum Average Contrast Estimation, CoVaR estimation, Composite quasi-maximum likelihood estimation, Lasso, Model selection, JEL Classification: C00, C14, C50, C58, AMS codes: 62G08; 62G20; 62P20

inference. An application in finance is the analysis of conditional Value-at-Risk (VaR). [5] proposed the CaViaR framework to model VaR dynamically. [12] used their QR techniques to test heteroscedasticity in the field of labor market discrimination. Like expectile analysis it models the conditional tail behavior.

The QR estimation implicitly assumes an asymmetric ALD (asymmetric Laplace distribution) likelihood, and may not be efficient in the QMLE case. Therefore, different types of flexible loss functions are considered in the literature to improve the estimation efficiency, such as, composite quantile regression, [29], [9] and [10]. Moreover, [3] proposed 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 special type of loss considered in [17] corresponds to expectile regression (ER) that is in spirit similar to QR but contains mean regression as its special case. Nonparametric expectile smoothing work with application to demography could be found in [19]. The ER curves are alternatives to the QR curves and give us an alternative picture of regression of  $Y$  on  $X$ .

The difficulty of characterizing an entire distribution partly arises from the high dimensionality of covariates, which asks for striking 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. [23] considered quantile regression via a single index model. However, to our knowledge there are no further literatures on generalized QR for the single-index model.

In addition to the dimension reduction, there is however the problem of choosing the right variables for projection. This motivates our second goal of this research: variable selection. [14], [22] and [27] focused on variable selection in mean regression for the single index model. Considering the uncertainty on the multi-index model structure, we restrict ourselves to the single-index model at the moment. An application of our research is presented in the relevant financial risk area: to investigate how the revenue distribution of companies depends on financial ratios describing risk factors for possible failure. Such kind of research has important consequences for rating and credit scoring.

When the dimension of  $X$  is high, severe nonlinear dependencies between  $X$  and the expectile (quantile) curves are expected. This triggers the nonparametric approach, but in its full gear, it runs into the “curse of dimensionality” trap, meaning that the convergence rate of the smoothing techniques is so slow that it is actually impractical to use in such situations. A balanced dimension reduction space for quantile regression is therefore needed. The MAVE technique, [24] provides us 1) with a dimension reduction and 2) good numerical properties for semiparametric function estimation. The set of ideas presented there, however, have never been applied to composite quantile framework or an even more general composite quasi-likelihood framework. The semiparametric multi-index approach that we consider herein will provide practitioners with a tool that combines flexibility in modeling with applicability for even very high dimensional data. Consequently the curse of

dimensionality is circumvented. The Lasso idea in combination with the minimum average contrast estimate (MACE) technique will provide a set of relevant practical techniques for a wide range of disciplines. The algorithms used in this project are published on the quantlet database [www.quantlet.org](http://www.quantlet.org).

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 estimation with an application in financial market. All the technical details are delegated to appendix.

**2. MACE for Single Index Model.** Let  $X$  and  $Y$  be  $p$  dimensional and univariate random elements respectively, ( $p$  can be very large, namely of the rate  $\exp(n^\delta)$ , where ( $\delta$  is a constant)). The single index model is defined to be:

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

where  $g(\cdot) : \mathbb{R}^1 \mapsto \mathbb{R}^1$  is an *unknown* smooth link function,  $\beta^*$  is the vector of index parameters,  $\varepsilon$  is a continuous variable with mean zero. The interest here is to simultaneously estimate  $\beta^*$  and  $g(\cdot)$ . Different assumptions on the error structure will give us quantile, expectile or mean regression scenarios.

**2.1. Quasi-Likelihood for The Single Index Model.** There exist several estimation technique for (2.1), among these the ADE method as one of the oldest ones [7]. The semi-parametric SIM (2.1) also permits a one-step projection pursuit interpretation, therefore estimation tools from this stream of literature might also be employed [8]. The MAVE technique aiming at simultaneous estimation of  $(\beta^*, g(\cdot))$  has been proposed by [24]. Here we will apply a minimum contrast approach, called MACE. The MACE technique uses similar to MAVE a double integration but is different since the squared loss function is replaced by a convex contrast. Here we generalize MAVE in 3 ways. First, we generalize the setting to weighted loss functions that allow us to identify and estimate conditional quantiles, expectiles and other tail specific objects. Second, we consider the situation where  $p \rightarrow \infty$  might be very large and therefore will add penalty terms that yield automatic model selection of e.g. Lasso or SCAD type. Third, we implement a composite estimation technique for estimating  $\beta^*$  that involves possibly many estimates.

In a quasi maximum likelihood (or equivalently a minimum contrast) framework the direction  $\beta$  (for known  $g(\cdot)$ ) is the solution of

$$(2.2) \quad \min_{\beta} \mathbb{E} \rho_w\{Y - g(X^\top \beta)\},$$

with the general quasi-likelihood loss function  $\rho_{\mathbf{w}}(\cdot) = \sum_{k=1}^K \mathbf{w}_k \rho_k(\cdot)$ , where  $\rho_1(\cdot), \dots, \rho_K(\cdot)$  are convex loss functions and  $\mathbf{w}_1, \dots, \mathbf{w}_K$  are positive weights. This weighted loss function form 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  $\beta$ ,  $\|\beta\|_2 = 1$  and the first component of  $\beta$  is positive.

For example when  $K = 1$ , the QR setting means to take the loss function as:

$$(2.3) \quad \rho_{\mathbf{w}}(u) = \rho_1(u) = \tau u \mathbf{1}(u > 0) + (1 - \tau) u \mathbf{1}(u < 0),$$

Moreover, for ER with  $K = 1$ , we have:

$$(2.4) \quad \rho_{\mathbf{w}}(u) = \rho_1(u) = \tau u^2 \mathbf{1}(u > 0) + (1 - \tau) u^2 \mathbf{1}(u < 0).$$

More general, the CQR setting employs  $K$  different quantiles  $\tau_1, \tau_2, \dots, \tau_K$ , with  $\mathbf{w}_k = 1/K, k = 1, \dots, K$  and

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

where  $b_k$  is the  $\tau_k$  quantile of the error distribution, see [3]. Let us now turn to the idea of MACE. First, we approximate  $g(X_i^\top \beta)$  for  $x$  near  $X_i$ :

$$(2.6) \quad g(X_i^\top \beta) \approx g(x^\top \beta) + g'(x^\top \beta)(X_i - x)^\top \beta,$$

In the context of local linear smoothing, a first order proxi of  $\beta$  (given  $x$ ) can therefore be constructed by minimizing:

$$(2.7) \quad L_x(\beta) \stackrel{\text{def}}{=} \mathbb{E} \rho_{\mathbf{w}}\{Y - g(x^\top \beta) - g'(x^\top \beta)(X_i - x)^\top \beta\},$$

The empirical version of (2.7) requires minimizing, with respect to  $\beta$ :

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

Employing now the double integration idea of MAVE, i.e. integrating with respect to the edf of the  $X$  variable yields as average contrast:

$$(2.9) \quad L_n(\beta) \stackrel{\text{def}}{=} n^{-2} \sum_{j=1}^n \sum_{i=1}^n \rho_{\mathbf{w}}\left\{Y_i - g(X_j^\top \beta) - g'(X_j^\top \beta)(X_i - X_j)^\top \beta\right\} K_h\{(X_i - X_j)^\top \beta\}$$

where  $K_h(\cdot)$  is the kernel function,  $K_h(u) = h^{-1}K(u/h)$ ,  $h$  a bandwidth.

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 minimization problems, motivated by the proposal in [15],

- Given  $\beta$ , the estimation of  $a(\cdot)$  and  $b(\cdot)$  are obtained through local linear minimization.
- Given  $a(\cdot)$  and  $b(\cdot)$ , the minimization with respect to  $\beta$  is carried out by the interior point method.

**2.2. Variable Selection for Single Index Model.** The dimension  $p$  of covariates is large, even  $p = \mathcal{O}\{\exp(n^\delta)\}$ , therefore selecting important covariates is a necessary step. Without loss of generality assume that the first  $q$  components of  $\beta^*$ , the true value of  $\beta$ , are non-zero. To point this out write  $\beta^* = (\beta_{(1)}^{*\top}, \beta_{(0)}^{*\top})^\top$  with  $\beta_{(1)}^* \stackrel{\text{def}}{=} (\beta_1, \dots, \beta_q)^\top \neq 0$  and  $\beta_{(0)}^* \stackrel{\text{def}}{=} (\beta_{q+1}, \dots, \beta_p)^\top = 0$  element-wise. Accordingly we denote  $X_{(1)}$  and  $X_{(0)}$  as the first  $q$  and the last  $p - q$  elements of  $X$ , respectively.

Suppose  $\{(X_i, Y_i)\}_{i=1}^n$  be  $n$  i.i.d. copies of  $(X, Y)$ . Consider now estimating the SIM coefficient  $\beta$  by solving the optimization problem

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

where  $X_{ij} \stackrel{\text{def}}{=} X_i - X_j$ ,  $\omega_{ij}(\beta) \stackrel{\text{def}}{=} K_h(X_{ij}^\top \beta) / \sum_{i=1}^n K_h(X_{ij}^\top \beta)$ . And  $\gamma_\lambda(t)$  is some non-negative function, and  $\hat{\beta}^{(0)}$  is an initial estimator of  $\beta^*$  (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 [21] with  $\gamma_\lambda(x) = \lambda$ , the SCAD [6] with

$$\gamma_\lambda(x) = \lambda \{ \mathbf{1}(x \leq \lambda) + \frac{(a\lambda - x)_+}{(a-1)\lambda} \mathbf{1}(x > \lambda) \}, \quad (a > 2)$$

and  $\gamma_\lambda(x) = \lambda|x|^{-a}$  for some  $a > 0$  corresponding to the adaptive Lasso [28].

We propose to estimate  $\beta$  in (2.10) with an iterative procedure described below. Denote  $\hat{\beta}_w$  the final estimate of  $\beta$ . Specifically, for  $t = 1, 2, \dots$ , iterate the following two steps. Denote  $\hat{\beta}^{(t)}$  as the estimate at step  $t$ .

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

$$(2.11) \quad (\hat{a}_j^{(t)}, \hat{b}_j^{(t)}) \stackrel{\text{def}}{=} \arg \min_{(a_j, b_j)'s} \sum_{i=1}^n \rho_w(Y_i - a_j - b_j X_{ij}^\top \hat{\beta}^{(t)}) \omega_{ij}(\hat{\beta}^{(t)}).$$

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

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

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

When choosing the penalty parameter  $\lambda$ , we adopt a  $C_p$ -type criterion as in [26] instead of the computationally involved cross validation method. We choose the optimal weights of the convex loss functions  $\rho_{\mathbf{w}}$  by minimizing the asymptotic variance of the resulting estimator of  $\beta$ , and the bandwidth by criteria minimizing the integrated mean squared errors of the estimator of  $g(\cdot)$ .

**3. Main Theorems.** Define  $\hat{\beta}_{\mathbf{w}} \stackrel{\text{def}}{=} (\hat{\beta}_{\mathbf{w}(1)}^\top, \hat{\beta}_{\mathbf{w}(2)}^\top)^\top$  as the estimator for  $\beta^*$  attained by the procedure in (2.11) and (2.12). Let  $\hat{\beta}_{\mathbf{w}(1)}$  and  $\hat{\beta}_{\mathbf{w}(2)}$  be the first  $q$  components and the remaining  $p - q$  components of  $\hat{\beta}_{\mathbf{w}}$  respectively. If in the iterations, we have the initial estimator  $\hat{\beta}_{(1)}^{(0)}$  as a  $\sqrt{n/q}$  consistency one for  $\beta_{(1)}^*$  (2.12), we will obtain with a very high probability, an oracle estimator of the following type, say  $\tilde{\beta}_{\mathbf{w}} = (\tilde{\beta}_{\mathbf{w}(1)}^\top, \mathbf{0}^\top)^\top$ , since the oracle knows the true model  $\mathcal{M}_* \stackrel{\text{def}}{=} \{l : \beta_l^* \neq 0\}$ . 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 (2.2) but within subspace  $\{\beta \in \mathbb{R}^p : \beta_{\mathcal{M}_*^c} = \mathbf{0}\}$ .

**THEOREM 1.** *Under conditions 1-5, the estimators  $\hat{\beta}^0$  and  $\hat{\beta}_{\mathbf{w}}$  exist and coincide on a set with probability tending to 1. Moreover,*

$$(3.1) \quad \mathbb{P}(\hat{\beta}^0 = \hat{\beta}_{\mathbf{w}}) \geq 1 - (p - q) \exp(-C' n^\alpha)$$

for a positive constant  $C'$ .

**THEOREM 2.** *Under conditions 1-5, we have*

$$(3.2) \quad \|\hat{\beta}_{\mathbf{w}(1)} - \beta_{(1)}^*\| = \mathcal{O}_p\{(\lambda D_n + n^{-1/2})\sqrt{q}\}$$

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

$$(3.3) \quad \mathbf{b}^\top C_{0(1)}^{-1} \sqrt{n}(\hat{\beta}_{\mathbf{w}(1)} - \beta_{(1)}^*) \xrightarrow{\mathcal{L}} \mathbf{N}(0, \sigma_{\mathbf{w}}^2)$$

where  $C_{0(1)} \stackrel{\text{def}}{=} \mathbb{E}\{[g'(Z_i)]^2 [\mathbb{E}(X_{(1)}|Z_i) - X_{i(1)}][\mathbb{E}(X_{(1)}|Z_i) - X_{i(1)}]^\top\}$ ,  $Z_i \stackrel{\text{def}}{=} X_i^\top \beta^*$ ,  $\psi_{\mathbf{w}}(\varepsilon)$  is a choice of the subgradient of  $\rho_{\mathbf{w}}(\varepsilon)$  and  $\sigma_{\mathbf{w}}^2 \stackrel{\text{def}}{=} \mathbb{E}[\psi_{\mathbf{w}}(\varepsilon_i)]^2 / [\partial^2 \mathbb{E} \rho_{\mathbf{w}}(\varepsilon_i)]^2$ , where

$$(3.4) \quad \partial^2 \mathbb{E} \rho_{\mathbf{w}}(\cdot) = \frac{\partial^2 \mathbb{E} \rho_{\mathbf{w}}(\varepsilon_i - v)^2}{\partial v^2} \Big|_{v=0}.$$



Let us now look at the distribution of  $\hat{g}(\cdot)$  and  $\hat{g}'(\cdot)$ , the estimator of  $g'(\cdot)$ .

**THEOREM 3.** *Under conditions 1-5,*

*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  $Z_i$ ,  $i = 1, \dots, n$ , if  $nh^3 \rightarrow \infty$  and  $h \rightarrow 0$ , we have*

$$\sqrt{nh} \sqrt{f_Z(z)/(\nu_0 \sigma_w^2)} \left\{ \widehat{g}(x^\top \widehat{\beta}) - g(x^\top \beta^*) - \frac{1}{2} h^2 g''(x^\top \beta^*) \mu_2 \partial \mathbf{E} \psi_w(\varepsilon) \right\} \xrightarrow{\mathcal{L}} \mathbf{N}(0, 1),$$

*Also, we have*

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

**4. Simulation.** In this section, we evaluate our technique in several settings, involving different combinations of link functions  $g(\cdot)$ , distributions of  $\varepsilon$ , and different choices of  $(n, p, q, \tau)$ s, where  $n$  is the sample size,  $p$  is the dimension of the true parameter  $\beta^*$ ,  $q$  is the number of non-zero components in  $\beta^*$ , and  $\tau$  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.

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

$$(4.1) \quad Y_i = 5 \cos(D \cdot Z_i) + \exp(-D \cdot Z_i^2) + \varepsilon_i,$$

where  $Z_i = X_i^\top \beta^*$ ,  $D = 0.01$  is a scaling constant and  $\varepsilon_i$  is the error term. Model 2:

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

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

$$(4.3) \quad Y_i = 10 \sin(D \cdot Z_i) + \sqrt{|\sin(0.5 \cdot Z_i) + \varepsilon_i|},$$

**4.2. Criteria.** For estimation accuracy for  $\beta$  and  $g(\cdot)$ , we use following five criteria to measure:

1) Standardized  $L_2$  norm:

$$Dev \stackrel{\text{def}}{=} \frac{\|\beta^* - \widehat{\beta}\|_2}{\|\beta^*\|_2},$$

2) Sign consistency:

$$Acc \stackrel{\text{def}}{=} \sum_{l=1}^p |\text{sign}(\beta_l^*) - \text{sign}(\widehat{\beta}_l)|,$$

3) Least angle:

$$Angle \stackrel{\text{def}}{=} \frac{\langle \beta^*, \hat{\beta} \rangle}{\|\beta^*\|_2 \cdot \|\hat{\beta}\|_2},$$

4) Relative error:

$$Error \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^n \left| \frac{g(Z_i) - \hat{g}(\hat{Z}_i)}{g(Z_i)} \right|,$$

5) Average squared error:

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

**4.3.  $L_1$ -norm quantile regression.** In this subsection, consider the  $L_1$ -norm quantile regression described by [16]. The initial value of  $\beta$  can be calculated by the  $L_1$ -norm quantile regression, then the two-step iterations mentioned in theoretical part are performed. Recall that  $X$  is a  $p \times n$  matrix, and  $q$  is the number of non-zero components in  $\beta^*$ . The  $j$ th column of  $X$  is an i.i.d. sample from  $N(j/2, 1)$ . Two error distributions are considered:  $\varepsilon_i \sim N(0, 0.1)$  and  $t(5)$ . Note that  $\beta_{(1)}^*$  is the vector of the non-zero components in  $\beta^*$ . In the simulation, we consider different  $\beta_{(1)}^*$ :  $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$ ,  $\beta_{(1)}^{*\top} = (5, 4, 3, 2, 1)$  and  $\beta_{(1)}^{*\top} = (5, 2, 1, 0.8, 0.2)$ . Here the indices  $Z_i$ 's are re-scaled to  $[0, 1]$  for nonparametric estimation. The bandwidth is selected by as in [25]:

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

where  $h_{mean}$  can be calculated by using the direct plug-in methodology of a local linear regression described by [18]. 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 (black) and the estimated link function (red). 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 decreased bandwidth ( $h = 0.068$ ). It can be seen that the estimated curve is very rough. The right plot shows that the estimated link function with increased bandwidth ( $h = 1$ ), the deviation between the estimated link function curve and the true curve is very large. From this comparison we may consider the aforementioned bandwidth selection performs well.

Figure 1 about here

Table 1 shows the criteria evaluated with different models and quantile levels. Here  $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$ , the error  $\varepsilon$  follows a  $N(0, 0.1)$  distribution or follows a  $t(5)$  distribution. In 100 simulations we set  $n = 100$ ,  $p = 10$ ,  $q = 5$ . 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 estimation for the nonlinear model 2 are not as good as model 1 and model 3, the error is still moderate. Figures 2 to Figure 4 present the plots of the true link function against the estimated ones for different quantile levels.

Table 1 and Figures 2 to 4 about here

Table 2 reports 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  $\varepsilon$  follows a  $N(0, 0.1)$  distribution. In 100 simulations we set  $n = 100, p = 10, q = 5, \tau = 0.95$ . 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 would be affected as well. Figure 5 and Figure 6 are the plots of the estimated link functions in these two cases.

Table 2 and Figures 5 to 6 about here

Table 3 shows the criteria evaluated under  $p > n$  case. Here  $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$ , the error  $\varepsilon$  follows a  $N(0, 0.1)$  distribution. In 100 simulations we set  $n = 100, p = 200, q = 5, \tau = 0.05$ . Standard deviations are given in brackets. We find that the errors are still moderate in  $p > n$  situation compared with Table 1. Figure 7 shows the graphs in this case.

Table 3 and Figures 7 about here

**4.4. Composite  $L_1$ - $L_2$  Regression.** In this subsection, a combined  $L_1$  and  $L_2$  loss is concerned and thus, the corresponding optimization is formed as:

$$(4.4) \quad \arg \min_{\beta, g(\cdot)} \left[ w_1 \sum_{i=1}^n |Y_i - g(X_i^\top \beta)| \omega_i(\beta) + w_2 \sum_{i=1}^n \{Y_i - g(X_i^\top \beta)\}^2 \omega_i(\beta) + n \sum_{l=1}^p \gamma_\lambda(|\beta_l|) |\beta_l| \right].$$

It can be further formulated as

$$(4.5) \quad \arg \min_{\beta, g(\cdot)} \left[ \{w_1 \sum_{i=1}^n |Y_i - g(X_i^\top \beta)|^{-1} \omega_i(\beta) + w_2\} |Y_i - g(X_i^\top \beta)|^2 \omega_i(\beta) + n \sum_{l=1}^p \gamma_\lambda(|\beta_l|) |\beta_l| \right].$$

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 until convergence:

$$(4.6) \quad \arg \min_{\beta, g(\cdot)} \left[ \{w_1 \sum_{i=1}^n |Res_i^t|^{-1} \omega_i(\beta^t) + w_2\} |Y_i - g(X_i^\top \beta)|^2 \omega_i(\hat{\beta}^t) + n \sum_{l=1}^p \gamma_\lambda(|\beta_l|) |\beta_l| \right].$$

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 w.r.t. the error distribution and the increasing dimension of  $p$ , since only the dimension of  $q$  matters.

Table 4 and Figure 8 about here.

4.5. *Composite  $L_1$  Quantile Regression.* Use 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 5 and Figure 8 about here

**5. Application.** In this section, we apply the proposed methodology to analyze risk conditional on macroprudential and other firm variables for small financial firms. 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.

5.1. *Data.* The firm data are selected according to the ranking of NASDAQ. We take as example city national corp. CYN as our objective. The remaining 199 financial institutions together with 7 lagged macroprudential variables are chosen as covariates. The list of these firms comes from the website.<sup>1</sup> The daily stock prices of these 200 firms are from Yahoo Finance for the period from January 6, 2006 to September 6, 2012. The descriptive statistics of the company, the description of the macroprudential variables and the list of the firms (Table 7 to Table 9) can be found in the Appendices. We consider a two step regression procedure. The first one is a quantile regression, where one regresses log returns of each covariate on all the lagged macroprudential variables:

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

where  $X_{i,t}$  represents the asset return of financial institution  $i$  at time  $t$ . We apply quantile regression proposed by [11]. Then the VaR of each firm with  $F_{\varepsilon_{i,t}}^{-1}(\tau|M_{t-1}) = 0$  can be obtained by:

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

---

<sup>1</sup> <http://www.nasdaq.com/screening/companies-by-industry.aspx?industry=Finance>.

Then the second regression is performed using our method, where the response variable is log returns of CYN, and the explanatory variable are the log returns of those covariates and the lagged macroprudential variables:

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

where  $S \stackrel{def}{=} [M_{t-1}, R]$ ,  $R$  is a vector of log returns for different firms.  $\beta_{j|S}$  is a  $p \times 1$  vector,  $p$  large. With  $F_{\varepsilon_{j,t}}^{-1}(\tau|S) = 0$  the CoVaR is estimated as:

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

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

Then we proceed the backtesting. The days on which the log returns are lower than the VaR or CoVaR can be called violations. The violation sequence is defined as follows:

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

Generally,  $I_t$  should be a martingale difference sequence. Then we apply CaViaR test, see by [2]. The CaViaR test model:

$$I_t = \alpha + \beta_1 I_{t-1} + \beta_2 VaR_t + u_t.$$

The test procedure is to estimate  $\beta_1$  and  $\beta_2$  by logistic regression, then Wald's test is applied with null hypothesis:  $\hat{\beta}_1 = \hat{\beta}_2 = 0$ .

**5.2. Results.** We use a moving window size of  $n = 126$  to calculate VaR of the log returns for the 199 firms. We also calculated the VaR of CYN. Figure 10 and Figure 11 show one example of the estimated VaR of one covariate (JPM) and the estimated VaR of CYN, respectively. It can be seen that the estimated VaR becomes more volatile when volatility of the returns is large.

Figures 10 and 11 about here.

Then the estimation of the CoVaR for CYN is conducted by applying a moving window size of  $n = 126$ .  $L_1$ -norm quantile regression is applied with  $\tau = 0.05$ . Recall there are  $p = 206$  covariates, the CoVaR is estimated with different variables selected in each window. Figure 12 shows the result.

Figures 12 about here.

Figures 13 and 14 show the estimated link functions against the indices in different window. We find some evidence on nonlinearity of the link function, although Figure 14 looks linear.

Figures 13 and 14 about here.

Figure 15 summarized the selection frequency of the firms and macroprudential variables for all the windows. The variable 187, "Radian Group Inc. (RDN)" is the most frequently selected variable with frequency 557.

Figure 15 about here.

Next we apply the backtesting. Figure 16 shows the  $\hat{I}_t$  sequence of  $\widehat{VaR}$  of CYN, there are totally 8 violations. Since  $T = 1543$ , we get that  $\hat{\tau} = 0.005$ . And the p value of wald test is 0.87, we can not reject the null hypotheses. From Figure 17 we get the  $\hat{I}_t$  sequence of  $\widehat{CoVaR}$  of CYN, there are 53 violations for  $T = 1543$ , and  $\hat{\tau} = 0.034$ . Since the p value of wald test is 0.36, the null hypotheses is not rejected. Therefore both VaR and CoVaR algorithms perform well.

Figures 16 and 17 about here.

## 6. Appendices.

6.1. *Proof of the main theorems.* We make the following assumptions for the proofs of the theorems in this paper.

*Condition 1.* The kernel  $K(\cdot)$  is a continuous symmetric pdf having at least four finite moments. The link  $g(\cdot)$  has a continuous second derivative.

*Condition 2.* Assume that  $\rho_k(x)$  are all strictly convex, and suppose  $\psi_k(x)$ , the derivative (or a subgradient of) of  $\rho_k(x)$ , satisfies  $\mathbf{E} \psi_k(\varepsilon_i) = 0$  and  $\inf_{|v| \leq c} \partial \mathbf{E} \psi_k(\varepsilon_i - v) = C_1$  where  $\partial \mathbf{E} \psi_k(\varepsilon_i - v)$  is the partial derivative with respect to  $v$ , and  $C_1$  is a constant.

*Condition 3.* In the case of composite quantile,  $K > 1$  assume that the error term  $\varepsilon_i$  is independent of  $X_i$ . For  $K = 1$  with a quantile and expectile loss relax to  $F_{y|x}^{-1}(\tau) = 0$ . Let  $X_{(1)i}$  denote the sub-vector of  $X_i$  consisting of its first  $q$  elements. Let  $Z_i \stackrel{\text{def}}{=} X_i^\top \beta^*$  and  $Z_{ij} = Z_i - Z_j$ . Define  $C_{0(1)} \stackrel{\text{def}}{=} \mathbf{E}\{[g'(Z_i)]^2(\mathbf{E}(X_{i(1)}|Z_i) - X_{i(1)})(\mathbf{E}(X_{i(1)}|Z_i - X_{i(1)}))^\top\}$ , 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  $c_0 > 0$  such that  $\sum_{i=1}^n \{\|X_{i(1)}\|/\sqrt{n}\}^{2+c_0} \rightarrow 0$ ,

with  $0 < c_0 < 1$ .  $v_{ij} \stackrel{\text{def}}{=} Y_i - a_j - b_j X_{ij}^\top \beta$ . Also

$$\left\| \sum_i \sum_j X_{(0)ij} \omega_{ij} X_{(1)ij}^\top \partial \mathbf{E} \psi_{\mathbf{w}}(v_{ij}) \right\|_{2,\infty} = \mathcal{O}(n^{1-\alpha_1}).$$

*Condition 4.* The penalty parameter  $\lambda$  is chosen such that  $\lambda D_n = \mathcal{O}\{n^{-1/2}\}$ , with  $D_n \stackrel{\text{def}}{=} \max\{d_l : l \in \mathcal{M}_*\} = \mathcal{O}(n^{\alpha_1-\alpha_2/2})$ ,  $d_l \stackrel{\text{def}}{=} \gamma_\lambda(|\beta_l^*|)$ ,  $\mathcal{M}_* = \{l : \beta_l^* \neq 0\}$  be the true model. Furthermore assume  $qh \rightarrow 0$  as  $n$  goes to infinity,  $q = \mathcal{O}(n^{\alpha_2})$ ,  $p = \mathcal{O}\{\exp(n^\delta)\}$ ,  $nh^3 \rightarrow \infty$  and  $h \rightarrow 0$ . Also,  $0 < \delta < \alpha < \alpha_2/2 < 1/2$ ,  $\alpha_2/2 < \alpha_1 < 1$ .

*Conditions 5.* The error term  $\varepsilon_i$  satisfies  $\mathbf{E} \varepsilon_i = 0$  and  $\text{Var}(\varepsilon_i) < \infty$ . Assume that

$$(6.1) \quad \mathbf{E} |\psi_{\mathbf{w}}^m(\varepsilon_i)/m!| \leq s_0 M^m$$

where  $s_0$  and  $M$  are constants.

Condition 1 is commonly-used and the standard normal pdf is a kernel satisfying this condition. Condition 2 is made on the weighted loss function so that it admits a quadratic approximation. Under condition 3, the matrix in the quadratic approximation is non-singular, so that the resulting estimate of  $\beta$  has an non-degenerate limiting distribution. Condition 4 guarantees that the proposed variable and estimation procedure for  $\beta$  is model-consistent. Condition 5 implies a certain tail behavior that we employ in all statistics argument.

Recall  $\hat{\beta}^0$  as the minimizer with the same loss in (2.2) but within the subspace  $\{\beta \in \mathbb{R}^p : \beta_{\mathcal{M}_*^c} = \mathbf{0}\}$ . The following lemma assures the consistency of  $\hat{\beta}^0$ ,

LEMMA 1. Under conditions 1-5, recall  $d_j = \gamma_\lambda(|\beta_j^*|)$ , we have that

$$(6.2) \quad \|\hat{\beta}^0 - \beta^*\| = \mathcal{O}_p(\sqrt{q/n} + \lambda \|\mathbf{d}_{(1)}\|)$$

where  $\mathbf{d}_{(1)}$  is the subvector of  $\mathbf{d} = (d_1, \dots, d_p)^\top$  which contains  $q$  elements corresponding to the nonzero  $\beta_{(1)}^*$ .

PROOF. Note that the last  $p - q$  elements of both  $\hat{\beta}^0$  and  $\beta^*$  are zero, so it is sufficient to prove  $\|\hat{\beta}_{(1)}^0 - \beta_{(1)}^*\| = \mathcal{O}_p(\sqrt{q/n} + \lambda \|\mathbf{d}_{(1)}\|)$ .

Write

$$\tilde{L}_n(\beta) = \sum_{j=1}^n \sum_{i=1}^n \rho_{\mathbf{w}}(Y_i - a_j - b_j X_{ij}^\top \beta) \omega_{ij}(\beta^*) + n \sum_{l=1}^p d_l |\beta_l|.$$

We first show for  $\gamma_n = \mathcal{O}(1)$ :

$$\mathbb{P} \left[ \inf_{\|\mathbf{u}\|=1} \{ \tilde{L}_n(\beta_{(1)}^* + \gamma_n \mathbf{u}, \mathbf{0}) > \tilde{L}_n(\beta^*) \} \right] \rightarrow 1$$

Construct  $\gamma_n \rightarrow 0$  so that for a sufficiently large constant  $B$ :  $\gamma_n > B \cdot (\sqrt{q/n} + \lambda \|\mathbf{d}_{(1)}\|)$ . We will show that by the local convexity of  $\tilde{L}_n(\beta_{(1)}, \mathbf{0})$  near  $\beta_{(1)}^*$ , there exists a unique minimizer inside the ball  $\{\beta_{(1)} : \|\beta_{(1)} - \beta_{(1)}^*\| \leq \gamma_n\}$  with probability tending to 1.

Let  $X_{(1)ij}$  denote the subvector of  $X_{ij}$  consisting of its first  $q$  components. By Taylor expansion at  $\gamma_n = 0$ :

$$\begin{aligned} & \tilde{L}_n(\beta_{(1)}^* + \gamma_n \mathbf{u}, \mathbf{0}) - \tilde{L}_n(\beta_{(1)}^*, \mathbf{0}) \\ &= \underbrace{\tilde{L}_n(\beta_{(1)}^* + \gamma_n \mathbf{u}, \mathbf{0}) - \tilde{L}_n(\beta_{(1)}^*, \mathbf{0}) - \mathbb{E}\{\tilde{L}_n(\beta_{(1)}^* + \gamma_n \mathbf{u}, \mathbf{0}) - \tilde{L}_n(\beta_{(1)}^*, \mathbf{0})\}}_{(T_{1n})} \\ & \quad + \underbrace{\mathbb{E}\{\tilde{L}_n(\beta_{(1)}^* + \gamma_n \mathbf{u}, \mathbf{0}) - \tilde{L}_n(\beta_{(1)}^*, \mathbf{0})\}}_{(T_{2n})} \end{aligned}$$

Taking Taylor expansion for the term  $T_{1n}$ ,  $T_{2n}$  respectively, where  $T_{1n}$  up to 1 order,  $T_{2n}$  up to 2 order:

$$\begin{aligned} & \tilde{L}_n(\beta_{(1)}^* + \gamma_n \mathbf{u}, \mathbf{0}) - \tilde{L}_n(\beta_{(1)}^*, \mathbf{0}) \\ &= -\gamma_n \sum_{i=1}^n \sum_{j=1}^n b_j \psi_{\mathbf{w}}(Y_i - a_j - b_j X_{(1)ij}^\top \beta_{(1)}^*) \omega_{ij}(\beta^*) X_{(1)ij}^\top \mathbf{u} \\ & \quad + \gamma_n \sum_{i=1}^n \sum_{j=1}^n b_j \partial \mathbb{E} \rho_{\mathbf{w}}(Y_i - a_j - b_j X_{(1)ij}^\top \beta_{(1)}^*) \omega_{ij}(\beta^*) X_{(1)ij}^\top \mathbf{u} \\ & \quad - \gamma_n \sum_{i=1}^n \sum_{j=1}^n b_j \partial \mathbb{E} \rho_{\mathbf{w}}(Y_i - a_j - b_j X_{(1)ij}^\top \beta_{(1)}^*) \omega_{ij}(\beta^*) X_{(1)ij}^\top \mathbf{u} \\ & \quad + \frac{1}{2} \gamma_n^2 \sum_{i=1}^n \sum_{j=1}^n b_j^2 \partial^2 \mathbb{E} \rho_{\mathbf{w}}(Y_i - a_j - b_j X_{(1)ij}^\top \beta_{(1)}^* - b_j \gamma_n X_{(1)ij}^\top \mathbf{u}) \omega_{ij}(\beta^*) (X_{(1)ij}^\top \mathbf{u})^2 \\ & \quad + n\lambda \sum_{l=1}^q d_l (|\beta_{(1)l}^*| + \gamma_n u_l - |\beta_{(1)l}^*|) + \mathcal{O}_p(\gamma_n) \end{aligned}$$



$$\begin{aligned}
&= -\gamma_n \sum_{i=1}^n \sum_{j=1}^n b_j \psi_{\mathbf{w}}(Y_i - a_j - b_j X_{(1)ij}^\top \beta_{(1)}^*) \omega_{ij}(\beta^*) X_{(1)ij}^\top \mathbf{u} \\
&\quad + \frac{1}{2} \gamma_n^2 \sum_{i=1}^n \sum_{j=1}^n b_j^2 \partial^2 \mathbb{E} \rho_{\mathbf{w}}(Y_i - a_j - b_j X_{(1)ij}^\top \beta_{(1)}^* - b_j \bar{\gamma}_n X_{(1)ij}^\top \mathbf{u}) \omega_{ij}(\beta^*) (X_{(1)ij}^\top \mathbf{u})^2 \\
&\quad + n\lambda \sum_{l=1}^q d_l (|\beta_{(1)l}^*| + \gamma_n u_l) - |\beta_{(1)l}^*| + \mathcal{O}_p(\gamma_n) \\
&\stackrel{\text{def}}{=} P_1 + P_2 + P_3 + \mathcal{O}_p(\gamma_n)
\end{aligned}$$

where  $\bar{\gamma}_n \in [0, \gamma_n]$ .

Define  $\omega_{ij} \stackrel{\text{def}}{=} \omega_{ij}(\beta^*)$ , it is not difficult to derive that  $\omega_{ij} = \frac{K_h(Z_{ij})}{nf_Z(Z_j)} \{1 + \mathcal{O}_p(1)\}$  where  $Z_i = X_i^\top \beta^*$ ,  $Z_{ij} = Z_i - Z_j$  and  $f_Z(\cdot)$  is the density of  $Z = X^\top \beta^*$ .

For  $P_1$ , because  $\|\mathbf{u}\| = 1$  and  $Y_i = a_i + \varepsilon_i$ , we get

$$\begin{aligned}
|P_1| &\leq \gamma_n \left\| \sum_{i=1}^n \sum_{j=1}^n b_j \psi_{\mathbf{w}}(Y_i - a_j - b_j X_{(1)ij}^\top \beta_{(1)}^*) \omega_{ij} X_{(1)ij} \right\| \{1 + \mathcal{O}_p(1)\} \\
&= \gamma_n \left\| \sum_{j=1}^n b_j \left\{ \frac{1}{n} \sum_{i=1}^n \psi_{\mathbf{w}}(\varepsilon_i + a_i - a_j - b_j Z_{ij}) \frac{K_h(Z_{ij})}{f_Z(Z_j)} X_{(1)ij} \right\} \right\| \{1 + \mathcal{O}_p(1)\} \\
&= \gamma_n \left\| \sum_{j=1}^n b_j \mathbb{E}_{\varepsilon_i, X_i} \left\{ \psi_{\mathbf{w}}(\varepsilon_i + a_i - a_j - b_j Z_{ij}) \frac{K_h(Z_{ij})}{f_Z(Z_j)} X_{(1)ij} \right\} \right\| \{1 + \mathcal{O}_p(1)\} \\
&= \gamma_n \left\| \sum_{j=1}^n b_j \mathbb{E}_{Z_i} \left\{ \mathbb{E}_{\varepsilon_i} [\psi_{\mathbf{w}}(\varepsilon_i + a_i - a_j - b_j Z_{ij})] \frac{K_h(Z_{ij})}{f_Z(Z_j)} \mathbb{E}(X_{(1)ij} | Z_i) \right\} \right\| \{1 + \mathcal{O}_p(1)\} \\
&= \gamma_n \left\| \sum_{j=1}^n b_j \mathbb{E} [\psi_{\mathbf{w}}(\varepsilon_j + a_j - a_j)] \{ \mathbb{E}(X_{(1)j} | Z_j) - X_{(1)j} \} \right\| \{1 + \mathcal{O}_p(1)\}
\end{aligned}$$

where  $\mathbb{E}_{\varepsilon_i, X_i}$  means taking expectation with respect to  $(\varepsilon_i, X_i)$ . Furthermore we have

$$\begin{aligned}
&\mathbb{E} \left\| \sum_{j=1}^n b_j \mathbb{E} [\psi_{\mathbf{w}}(\varepsilon_j + a_j - a_j)] \{ \mathbb{E}(X_{(1)j} | Z_j) - X_{(1)j} \} \right\| \\
&\leq \left\{ \mathbb{E} \psi_{\mathbf{w}}^2(\varepsilon_j + a_j - a_j) \mathbb{E} \sum_{j=1}^n b_j^2 [\mathbb{E}(X_{(1)j} | Z_j) - X_{(1)j}]^\top [\mathbb{E}(X_{(1)j} | Z_j) - X_{(1)j}] \right\}^{1/2} \\
&= \sqrt{n} \{ \mathbb{E} \psi_{\mathbf{w}}^2(\varepsilon_j + a_j - a_j) \text{tr}(C_{0(1)}) \}^{1/2},
\end{aligned}$$

recall  $C_{0(1)} \stackrel{\text{def}}{=} \mathbf{E}\{[g'(Z_j)]^2(\mathbf{E}(X_{(1)j}|Z_j) - X_{(1)j})(\mathbf{E}(X_{(1)j}|Z_j) - X_{(1)j})^\top\}$ . We thus arrive at

$$(6.3) \quad P_1 = \mathcal{O}_p(\gamma_n \sqrt{nq})$$

because  $\text{tr}(C_{0(1)}) = \mathcal{O}(q)$  and  $\mathbf{E} \psi_w^2(\varepsilon_j + a_j - a_j) < \infty$  by Condition 3.

For  $P_2$ , according to the property of kernel estimation, it can be seen that

$$\begin{aligned} P_2 &= \frac{1}{2} \gamma_n^2 \sum_{i=1}^n \sum_{j=1}^n b_j^2 \partial \mathbf{E} \psi_w(Y_i - a_j - b_j Z_{ij} - b_j \bar{\gamma}_n X_{(1)ij}^\top \mathbf{u}) \frac{K_h(Z_{ij})}{nf_Z(Z_j)} (X_{(1)ij}^\top \mathbf{u})^2 \{1 + \mathcal{O}_p(1)\} \\ &= \frac{1}{2} \gamma_n^2 \sum_{j=1}^n b_j^2 \partial \mathbf{E} \{ \psi_w(Y_i - a_j - b_j \bar{\gamma}_n X_{(1)ij}^\top \mathbf{u}) (X_{(1)ij}^\top \mathbf{u})^2 \} \{1 + \mathcal{O}_p(1)\} \end{aligned}$$

Let  $H_i(c) = \inf_{|v| \leq c} \partial \mathbf{E} \psi(\varepsilon_i - v)$ . By lemma 3.1 of Portnoy (1984), we have

$$(6.4) \quad P_2 \geq \frac{1}{2} \gamma_n^2 \sum_{i=1}^n \sum_{j=1}^n b_j^2 H(\gamma_n |X_{(1)ij}^\top \mathbf{u}|) (X_{(1)ij}^\top \mathbf{u})^2 \geq c \gamma_n^2 n$$

for some positive  $c$ .

For  $P_3$ , it is clear that

$$(6.5) \quad |P_3| \leq n \lambda \gamma_n \sum_{l=1}^q d_l |u_l| \leq n \lambda \gamma_n \|\mathbf{d}_{(1)}\|$$

Combining (6.3), (6.4) and (6.5), the following inequality holds with probability tending to 1 that

$$(6.6) \quad \tilde{L}_n(\beta_{(1)}^* + \gamma_n \mathbf{u}, \mathbf{0}) - \tilde{L}_n(\beta_{(1)}^*, \mathbf{0}) \geq n \gamma_n (c \gamma_n - \sqrt{q/n} - \lambda \|\mathbf{d}_{(1)}\|)$$

$\gamma_n = B(\sqrt{q/n} + \lambda \|\mathbf{d}_{(1)}\|)$  and  $B$  is a sufficiently large constant, so that the RHS of (6.6) is larger than 0. Owing to the local convexity of the objective function, there exists a unique minimizer  $\hat{\beta}_{(1)}^0$  such that

$$\|\hat{\beta}^0 - \beta^*\| = \|\hat{\beta}_{(1)}^0 - \beta_{(1)}^*\| = \mathcal{O}_p(\sqrt{q/n} + \lambda \|\mathbf{d}_{(1)}\|)$$

Therefore, (6.2) is proved.  $\square$

Recall that  $X = (X_{(1)}, X_{(2)})$  and  $\mathcal{M}_* = \{1, \dots, q\}$  is the set of indices at which  $\beta$  are nonzero.

LEMMA 2. Under conditions 1-5, the loss function (2.2) has a unique global minimizer  $\hat{\beta} = (\hat{\beta}_1^\top, \mathbf{0}^\top)^\top$ , if and only if

$$(6.7) \quad \sum_{j=1}^n \sum_{i=1}^n \psi_{\mathbf{w}}(Y_i - \hat{a}_j - \hat{b}_j X_{ij}^\top \hat{\beta}_{\mathbf{w}}) \hat{b}_j X_{(1)ij} \omega_{ij}(\beta^*) + n \mathbf{d}_{(1)} \circ \text{sign}(\hat{\beta}_{\mathbf{w}}) = 0$$

$$(6.8) \quad \|z(\hat{\beta}_{\mathbf{w}})\|_\infty \leq n\lambda,$$

where

$$(6.9) \quad z(\hat{\beta}_{\mathbf{w}}) \stackrel{\text{def}}{=} \mathbf{d}_{(0)}^{-1} \circ \left\{ \sum_{j=1}^n \sum_{i=1}^n b_j \psi_{\mathbf{w}}(Y_i - a_j - b_j X_{ij}^\top \hat{\beta}_{\mathbf{w}}) X_{(0)ij} \omega_{ij}(\hat{\beta}_{\mathbf{w}}) \right\}$$

where  $\circ$  stands for multiplication element-wise.

PROOF. According to the definition of  $\hat{\beta}_{\mathbf{w}}$ , it is clear that  $\hat{\beta}_{(1)}$  already satisfies condition (6.7). Therefore we only need to verify condition (6.8).

To prove (6.8), a bound for

$$(6.10) \quad \sum_{i=1}^n \sum_{j=1}^n b_j \psi_{\mathbf{w}}(Y_i - a_j - b_j X_{ij}^\top \beta^*) \omega_{ij} X_{(0)ij}$$

is needed. Define the following kernel function

$$\begin{aligned} & h_d(X_i, a_j, b_j, Y_i, X_j, a_i, b_i, Y_j) \\ &= \frac{n}{2} \left\{ b_j \psi_{\mathbf{w}}(Y_i - a_j - b_j X_{ij}^\top \beta^*) \omega_{ij} X_{(0)ij} + b_i \psi_{\mathbf{w}}(Y_j - a_i - b_i X_{ji}^\top \beta^*) \omega_{ji} X_{(0)ji} \right\}_d, \end{aligned}$$

where  $\{.\}_d$  denotes the  $d$ th element of a vector,  $d = 1, \dots, p - q$ .

According to [20], the proof of theorem B in page 201, and following the Conditions 5:

$$(6.11) \quad \mathbb{E}_F[\exp\{s \cdot h_d(X_i, a_j, b_j, Y_i, X_j, a_i, b_i, Y_j)\}] < \infty, \quad 0 < s < s_0,$$

where  $s_0$  is a constant.

Define  $U_{n,d} \stackrel{\text{def}}{=} \frac{1}{n(n-1)} \sum_{1 \leq i < j \leq n} h_d(X_i, a_j, b_j, Y_i, X_j, a_i, b_i, Y_j)$  as the  $U$ -statistics for (6.10).

Then, for  $\varepsilon > 0$ ,

$$\exp\{-s \cdot \mathbb{E} U_{n,d}\} \mathbb{E} \exp\{s \cdot h_d(\cdot)\} = 1 + \mathcal{O}(s^2), \quad s \rightarrow 0.$$

By taking  $s = \varepsilon/\{n^{2+\alpha}\}$ ,  $\varepsilon = n^{1/2+\alpha}$ , we have

$$\begin{aligned} \mathbb{P}(|U_{n,d} - \mathbb{E}U_{n,d}| > \varepsilon) &\leq 2 \left[ \exp(-s \cdot \varepsilon) \exp(-s \cdot \mathbb{E}U_{n,d}) \mathbb{E} \exp\{sh_d(\cdot)\} \right]^{[n/2]} \\ &\leq 2 \left[ \left\{1 + \mathcal{O}(s^2)\right\} \exp(-\varepsilon^2/n^{2+\alpha}) \right]^{[n/2]} \\ &\leq 2 \exp(-C_n n^\alpha), \end{aligned}$$

where  $C_n$  is a constant depending on  $n$ .

Define

$$F_{n,d} \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^n \sum_{j=1}^n b_j \psi_{\mathbf{w}}(Y_i - a_j - b_j X_{ij}^\top \beta^*) \omega_{ij} X_{(0)ij},$$

also it is not hard to derive that  $U_{n,d} = F_{n,d}$ .

It then follows that

$$\begin{aligned} \mathbb{P}(|F_{n,d} - \mathbb{E}F_{n,d}| > \varepsilon) &= \mathbb{P}(|U_{n,d} - \mathbb{E}U_{n,d}| > \varepsilon) \\ &\leq 2 \exp(-C'_n n^\alpha) \end{aligned}$$

Define  $\mathcal{A}_n = \{\|F_n - \mathbb{E}F_n\|_\infty \leq \varepsilon\}$ , thus

$$\mathbb{P}(\mathcal{A}_n) \geq 1 - \sum_{d=1}^{p-q} \mathbb{P}(|F_{n,d} - \mathbb{E}F_{n,d}| > \varepsilon) \geq 1 - 2(p-q) \exp(-C'_n n^\alpha).$$

Finally we get that on the set  $\mathcal{A}_n$ ,

$$\begin{aligned} \|z(\hat{\beta}^0)\|_\infty &\leq \|\mathbf{d}_{\mathcal{M}_*}^{-1} \circ F_n\|_\infty + \left\| \sum_{i=1}^n \sum_{j=1}^n b_j [\psi_{\mathbf{w}}(Y_i - a_j - b_j X_{ij}^\top \hat{\beta}^0) \right. \\ &\quad \left. - \psi_{\mathbf{w}}(Y_i - a_j - b_j X_{ij}^\top \beta^*)] \omega_{ij} X_{(0)ij} \right\|_\infty \\ &\leq \mathcal{O}(n^{1/2+\alpha}) + \left\| \sum_{i=1}^n \sum_{j=1}^n \partial \mathbb{E} \psi_{\mathbf{w}}(v_{ij}) b_j X_{(1)ij}^\top (\hat{\beta}_{(1)} - \beta_{(1)}^*) \omega_{ij} X_{(0)ij} \right\|_\infty, \end{aligned}$$

where  $v_{ij}$  is between  $Y_i - a_j - b_j X_{ij}^\top \beta^*$  and  $Y_i - a_j - b_j X_{ij}^\top \hat{\beta}^0$ . From Lemma 5.1,

$$\|\hat{\beta}^0 - \beta_{(1)}^*\|_2 = \mathcal{O}_p(\lambda \|\mathbf{d}_{(1)}\| + \sqrt{q}/\sqrt{n}).$$

Choosing  $\|\sum_i \sum_j X_{(0)ij} \omega_{ij} X_{(1)ij}^\top \partial \mathbb{E} \psi_{\mathbf{w}}(v_{ij})\|_{2,\infty} = \mathcal{O}(n^{1-\alpha_1})$ ,  $q = \mathcal{O}(n^{\alpha_2})$ ,  $\lambda = \mathcal{O}(\sqrt{n/q}) = n^{-1/2+\alpha_2/2}$ ,  $0 < \alpha_2 < 1$ ,  $\|\mathbf{d}_{(1)}\| = \mathcal{O}(\sqrt{q} D_n) = \mathcal{O}(n^{\alpha_2/2} D_n)$

$$\begin{aligned} (n\lambda)^{-1} \|z(\hat{\beta}^0)\|_\infty &= \mathcal{O}\{n^{-1} \lambda^{-1} (n^{1/2+\alpha} + n^{1-\alpha_1} \sqrt{q}/\sqrt{n} + \lambda \|\mathbf{d}_{(1)}\| n^{1-\alpha_1})\} \\ &= \mathcal{O}(n^{-\alpha_2/2+\alpha} + n^{-\alpha_1} + n^{-\alpha_1+\alpha_2/2} D_n), \end{aligned}$$

assume  $D_n = \mathcal{O}(n^{\alpha_1 - \alpha_2/2})$ , and let  $0 < \delta < \alpha < \alpha_2/2 < 1/2$ ,  $\alpha_2/2 < \alpha_1 < 1$ , with rate  $p = \mathcal{O}\{\exp(n^\delta)\}$ , then  $(n\lambda)^{-1}\|z(\hat{\beta}^0)\|_\infty = \mathcal{O}(1)$ .  $\square$

**Proof of Theorem 2.** By Theorem 3.1,  $\hat{\beta}_{\mathbf{w}(1)} = \beta_{(1)}$  almost surely. It then follows from Lemma 5.2 that

$$\|\hat{\beta}_{\mathbf{w}(1)} - \beta_{(1)}^*\| = \mathcal{O}_p\{(\lambda D_n + n^{-1/2})\sqrt{q}\}.$$

This completes the first part of the theorem.

Given  $a_j, b_j$ , consistent estimates of  $a_j$  and  $b_j$ , the estimate  $\hat{\beta}_{\mathbf{w}}$  is:

$$\hat{\beta}_{\mathbf{w}} = \arg \min_{\beta} \sum_{j=1}^n \sum_{i=1}^n \rho_{\mathbf{w}}(Y_i - a_j - b_j X_{ij}^\top \beta) \omega_{ij} + n \sum_{l=1}^p \gamma_\lambda(|\tilde{\beta}_l|) |\beta_l|$$

where  $\omega_{ij} = \omega_{ij}(\tilde{\beta})$  and  $X_{ij} = X_i - X_j$ .

Define  $\hat{\xi} \stackrel{\text{def}}{=} \sqrt{n}(\hat{\beta}_{\mathbf{w}} - \beta^*)$  and  $Y_{ij} = Y_i - a_j - b_j X_{ij}^\top \beta^*$ . As  $\omega_{ij}(\tilde{\beta}) = \frac{K_h(Z_{ij})}{nf_Z(Z_j)}\{1 + \mathcal{O}_p(1)\}$ , it follows that  $\hat{\xi}$  is the minimizer of

$$\begin{aligned} H_n(\xi) &= \sum_{i=1}^n \sum_{j=1}^n \left\{ \rho_{\mathbf{w}}(Y_{ij} - n^{-\frac{1}{2}} b_j X_{ij}^\top \xi) - \rho_{\mathbf{w}}(Y_{ij}) \right\} \frac{K_h(Z_{ij})}{nf_Z(Z_j)} \{1 + \mathcal{O}_p(1)\} \\ &\quad + n \sum_{l=1}^p \gamma_\lambda(|\beta_l^*|) (|\beta_l^* + n^{-1/2} \xi_l| - |\beta_l^*|) \\ &\stackrel{\text{def}}{=} Q_1(\xi) \{1 + \mathcal{O}_p(1)\} + Q_2(\xi), \end{aligned}$$

recall  $f_Z(z)$  is the density function of  $Z = X_i^\top \beta^*, i = 1, \dots, n$ . We study  $Q_1(\xi)$  and  $Q_2(\xi)$  respectively.

Let  $\Delta_{ij}(\xi) \stackrel{\text{def}}{=} \rho_{\mathbf{w}}(Y_{ij} - n^{-\frac{1}{2}} g'(Z_j) X_{ij}^\top \xi) - \rho_{\mathbf{w}}(Y_{ij}) - n^{-\frac{1}{2}} \psi_{\mathbf{w}}(Y_{ij}) g'(Z_j) X_{ij}^\top \xi$ . It can be seen that

$$\begin{aligned} Q_1(\xi) &= \left\{ \sum_{i=1}^n \sum_{j=1}^n n^{-1/2} \psi_{\mathbf{w}}(Y_{ij}) g'(Z_j) \frac{K_h(Z_{ij})}{nf_Z(Z_j)} X_{ij}^\top \xi \right. \\ &\quad \left. + \sum_{i=1}^n \sum_{j=1}^n \Delta_{ij}(\xi) \frac{K_h(Z_{ij})}{nf_Z(Z_j)} \right\} \{1 + \mathcal{O}_p(1)\} \\ &\stackrel{\text{def}}{=} [A^\top \xi + I(\xi)] \{1 + \mathcal{O}_p(1)\} \end{aligned}$$

Recall that  $Y_{ij} = \varepsilon_i + a_i - a_j - b_j Z_{ij} + o_p(1)$ .

Therefore we have

$$\begin{aligned} A &= \sum_{i=1}^n \sum_{j=1}^n \frac{1}{\sqrt{n}} \psi_{\mathbf{w}}(\varepsilon_i + a_i - a_j - b_j Z_{ij}) g'(Z_j) \frac{K_h(Z_{ij})}{nf_Z(Z_j)} X_{ij} \{1 + o_p(1)\} \\ &= \sum_{i=1}^n \frac{1}{\sqrt{n}} \psi_{\mathbf{w}}(\varepsilon_i) g'(Z_i) \{\mathbb{E}(X|Z_i) - X_i\} \{1 + o_p(1)\}. \end{aligned}$$

Similarly we have

$$\begin{aligned} I(\xi) &= \sum_{i=1}^n \sum_{j=1}^n \Delta_{ij}(\xi) \frac{K_h(Z_{ij})}{nf_Z(Z_j)} \\ &= \sum_{i=1}^n \sum_{j=1}^n \left\{ \rho_{\mathbf{w}}(Y_{ij} - n^{-\frac{1}{2}} g'(Z_j) X_{ij}^\top \xi) - \rho_{\mathbf{w}}(Y_{ij}) \right. \\ &\quad \left. - n^{-\frac{1}{2}} \psi_{\mathbf{w}}(Y_{ij}) g'(Z_j) X_{ij}^\top \xi \right\} \frac{K_h(Z_{ij})}{nf_Z(Z_j)} \\ &= \sum_{j=1}^n \mathbb{E}_{\varepsilon_i, X_i} \left[ \left\{ \rho_{\mathbf{w}}(Y_{ij} - n^{-\frac{1}{2}} g'(Z_j) X_{ij}^\top \xi) - \rho_{\mathbf{w}}(Y_{ij}) \right. \right. \\ &\quad \left. \left. - n^{-\frac{1}{2}} \psi_{\mathbf{w}}(\varepsilon_i) g'(Z_j) X_{ij}^\top \xi \right\} \frac{K_h(Z_{ij})}{f_Z(Z_j)} \right] \\ &= \sum_{j=1}^n \mathbb{E}_{Z_i} \mathbb{E}_{\varepsilon_i, X_i | Z_i} \left[ \left\{ \rho_{\mathbf{w}}(Y_{ij} - n^{-\frac{1}{2}} g'(Z_j) X_{ij}^\top \xi) - \rho_{\mathbf{w}}(Y_{ij}) \right. \right. \\ &\quad \left. \left. - n^{-\frac{1}{2}} \psi_{\mathbf{w}}(\varepsilon_i) g'(Z_j) X_{ij}^\top \xi \right\} \frac{K_h(Z_{ij})}{f_Z(Z_j)} \right] \\ &= \sum_{j=1}^n \left\{ \mathbb{E}_{\varepsilon} \rho_{\mathbf{w}}(\varepsilon - n^{-\frac{1}{2}} g'(Z_j) (X_j - \mathbb{E}(X_j|Z_j))^\top \xi) - \mathbb{E}_{\varepsilon} \rho_{\mathbf{w}}(\varepsilon) \right. \\ &\quad \left. - n^{-\frac{1}{2}} \mathbb{E}_{\varepsilon} [\psi_{\mathbf{w}}(\varepsilon)] g'(Z_j) (X_j - \mathbb{E}(X_j|Z_j))^\top \xi \right\} \\ &= (2n)^{-1} \sum_{j=1}^n \left\{ \partial^2 \mathbb{E} \rho_{\mathbf{w}}(\varepsilon_j) \right\} \mathbb{E}[g'(Z_j)]^2 \xi^\top \{X_j - \mathbb{E}(X_j|Z_j)\} \{X_j - \mathbb{E}(X_j|Z_j)\}^\top \xi \\ &= \frac{1}{2} \left\{ \partial^2 \mathbb{E} \rho_{\mathbf{w}}(\varepsilon_j) \right\} \xi^\top \mathbb{E} \{[g'(Z_j)]^2 \{X_j - \mathbb{E}(X_j|Z_j)\} \{X_j - \mathbb{E}(X_j|Z_j)\}^\top\} \xi \\ &\stackrel{\text{def}}{=} \frac{1}{2} \left\{ \partial^2 \mathbb{E} \rho_{\mathbf{w}}(\varepsilon_j) \right\} \xi^\top C_0 \xi [1 + o(1)] \end{aligned}$$

About  $Q_2(\xi)$ , we find that if  $\beta_l^* = 0$ , that is  $q < l \leq p$ ,  $\sqrt{n}(|\beta_l^* + n^{-1/2}\xi| - |\beta_l^*|) = |\xi|$ ,

otherwise when  $1 \leq l \leq q$ ,  $\sqrt{n}(|\beta_l^* + n^{-1/2}\xi| - |\beta_l^*|) \rightarrow \xi \text{sign}(\beta_l^*)$ . In the case of adaptive Lasso, we have  $\sqrt{n}\gamma_\lambda(|\beta_l|) \rightarrow 0$  for  $\beta_l^* \neq 0$  and  $\sqrt{n}\gamma_\lambda(|\beta_l^*|) \rightarrow \infty$  for  $\beta_l^* = 0$ . Thus  $n \sum_{l=1}^p \gamma_\lambda(|\beta_l^*|)(|\beta_l^* + n^{-1/2}\xi_l| - |\beta_l^*|)$  will converge to 0 only if  $\xi_l = 0$  for all  $q < l \leq p$ , otherwise it will converge to infinity. The loss function  $H_n(\xi)$  gives nontrivial solutions only when  $\xi_l = 0$  for all  $q < l \leq p$ . This implies that  $\hat{\beta}_{\mathbf{w}(2)}$  will converge to zero with probability one.

When  $\xi_{(2)} = 0$  the loss function

$$D_n(\xi) = [-\xi_{(1)}^\top A_{(1)} + \frac{1}{2}\{\partial^2 \mathbf{E} \rho_{\mathbf{w}}(\varepsilon_i)\} \xi_{(1)}^\top C_{0(1)} \xi_{(1)}]\{1 + \mathcal{O}(1)\}$$

where  $A_{(1)}$  is the sub-vector of  $A$  consisting of its first  $q$  components, and  $C_{0(1)}$  is the up-left  $q \times q$  sub-matrix of  $C_0$ . Following quadratic approximate lemma, we obtain  $\hat{\xi}_{(1)} = \sqrt{n}(\hat{\beta}_{(1)} - \beta_{(1)}^*) = [\partial^2 \mathbf{E} \rho_{\mathbf{w}}(\varepsilon_i)]^{-1} C_{0(1)}^{-1} A_{(1)} + \mathcal{O}_p(1)$ .

Recall that  $A_{(1)} = \sum_{i=1}^n \frac{1}{\sqrt{n}} \psi_{\mathbf{w}}(\varepsilon_i) g'(Z_i) \{\mathbf{E}(X_{(1)}|Z_i) - X_{i(1)}\} \{1 + \mathcal{O}_p(1)\}$ . Thus we have

$$\begin{aligned} \sqrt{n} \mathbf{b}^\top C_{0(1)}^{1/2} (\hat{\beta}_{(1)} - \beta_{(1)}^*) &= \mathbf{b}^\top [\partial^2 \mathbf{E} \rho_{\mathbf{w}}(\varepsilon_i)]^{-1} C_{0(1)}^{-1/2} \\ &\quad \times \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{\mathbf{w}}(\varepsilon_i) g'(Z_i) \{\mathbf{E}(X_{(1)}|Z_i) - X_{i(1)}\} \{1 + \mathcal{O}_p(1)\} \\ &\xrightarrow{\mathcal{L}} \mathbf{N}(0, \sigma_{\mathbf{w}}^2) \end{aligned}$$

where

$$\begin{aligned} \sigma_{\mathbf{w}}^2 &= \mathbf{b}^\top [\partial^2 \mathbf{E} \rho_{\mathbf{w}}(\varepsilon_i)]^{-2} \mathbf{E} \{[\psi_{\mathbf{w}}(\varepsilon_i)]^2\} C_{0(1)}^{-1/2} \\ &\quad \times \mathbf{E} \{[g'(Z_i)]^2 [\mathbf{E}(X_{(1)}|Z_i) - X_{i(1)}][\mathbf{E}(X_{(1)}|Z_i) - X_{i(1)}]^\top\} C_{0(1)}^{-1/2} \mathbf{b} \\ &= \mathbf{E} \{[\psi_{\mathbf{w}}(\varepsilon_i)]^2\} / [\partial^2 \mathbf{E} \rho_{\mathbf{w}}(\varepsilon_i)]^2. \end{aligned}$$

Here we have used the fact  $C_{0(1)} = \mathbf{E} \{[g'(Z_i)]^2 [\mathbf{E}(X_{(1)}|Z_i) - X_{i(1)}][\mathbf{E}(X_{(1)}|Z_i) - X_{i(1)}]^\top\}$ . The asymptotic normality can be proved.  $\square$

**Proof of Theorem 3.** We note that

$$\sqrt{nh} \left\{ \hat{g}(x^\top \hat{\beta}) - g(x^\top \beta^*) \right\} = \sqrt{nh} \left\{ \hat{g}(x^\top \hat{\beta}) - \hat{g}(x^\top \beta^*) \right\} + \sqrt{nh} \left\{ \hat{g}(x^\top \beta^*) - g(x^\top \beta^*) \right\}$$

As  $\rho_{\mathbf{w}}$  is strictly convex, then  $\hat{g}(\cdot)$  is continuous almost surely. As  $qh \rightarrow 0$ , the consistency of  $\hat{\beta}$  in Theorem 2 implies  $\sqrt{nh} \left\{ \hat{g}(x^\top \hat{\beta}) - \hat{g}(x^\top \beta^*) \right\} = \mathcal{O}_p(1)$ . Consequently it is sufficient to prove

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

We now prove equation (6.12). Let  $Z_i = X_i^\top \beta^*$  and  $z = x^\top \beta^*$ . Recall that  $\hat{g}(z) = \hat{a}$  and  $\hat{g}'(z) = \hat{b}$  where

$$(6.12) \quad (\hat{a}, \hat{b}) \stackrel{\text{def}}{=} \arg \min_{(a,b)} (nh) \sum_{i=1}^n \rho_w \{Y_i - a - b(Z_i - u)\} \omega_i(\beta^*)$$

where  $\omega_i(\beta^*) = K_h(Z_i - z) / \sum_{k=1}^n K_h(Z_k - z) = \{1 + \mathcal{O}_p(1)\} K_h(Z_i - z) / \{nf_Z(z)\}$

Denote  $\theta = [\sqrt{nh}\{a - g(z)\}, \sqrt{nh^3}\{b - g'(z)\}^\tau]^\top$  and  $\hat{\theta} = [\sqrt{nh}\{\hat{g}(z) - g(z)\}, \sqrt{nh^3}\{\hat{g}'(z) - g'(z)\}^\tau]^\top$ . Also let  $\mathbf{Z}_i = (1, (Z_i - z)/h)^\top$  and  $\delta_i = g(Z_i) - g(z) - g'(z)(Z_i - z)$ . Then  $\hat{\theta}$  is the minimizer of

$$(6.13) \quad L(\theta) = \sum_{i=1}^n [\rho_w(\varepsilon_i + \delta_i - (nh)^{-1/2} \mathbf{Z}_i^\top \theta) - \rho_w(\varepsilon_i + \delta_i)] \frac{K(Z_i - z)}{f_Z(z)} [1 + \mathcal{O}_p(1)].$$

By similar Taylor expansion to  $Q_1(\xi)$  in the previous proof, we have

$$\begin{aligned} L(\theta) &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n \psi_w(\varepsilon_i + \delta_i) \mathbf{Z}_i^\top \theta \frac{K(Z_i - z)}{f_Z(z)} [1 + \mathcal{O}_p(1)] \\ &\quad + \frac{1}{2nh} \sum_{i=1}^n \{\partial^2 \mathbb{E} \rho_w(\varepsilon_i + \delta_i)\} (\mathbf{Z}_i^\top \theta)^2 \frac{K(Z_i - z)}{f_Z(z)} [1 + \mathcal{O}_p(1)] \\ &= A_n^\top \theta + \frac{1}{2} \theta^\tau S_n \theta (1 + o_p(1)) \end{aligned}$$

where

$$\begin{aligned} A_n &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n \psi_w(\varepsilon_i + \delta_i) \mathbf{Z}_i \frac{K(Z_i - z)}{f_Z(z)} \\ S_n &= \frac{1}{nh} \sum_{i=1}^n \{\partial^2 \mathbb{E} \rho_w(\varepsilon_i + \delta_i)\} \mathbf{Z}_i \mathbf{Z}_i^\top \frac{K(Z_i - z)}{f_Z(z)}. \end{aligned}$$

It is easy to see that both  $A_n$  and  $S_n$  are sums of independent and identically distributed random elements. Thus the leading term of  $S_n$  is

$$\begin{aligned} S_n &= \frac{1}{h} \mathbb{E} \left\{ \partial^2 \mathbb{E} \rho_w(\varepsilon_i + \delta_i) \mathbf{Z}_i \mathbf{Z}_i^\top \frac{K(Z_i - z)}{f_Z(z)} | Z_i \right\} \\ &= \frac{1}{h} \mathbb{E} \left\{ \partial^2 \mathbb{E} \rho_w(\varepsilon_i + \delta_i) | Z_i \right\} \mathbf{Z}_i \mathbf{Z}_i^\top \frac{K(Z_i - z)}{f_Z(z)} \\ &= \partial^2 \mathbb{E} \rho_w(\varepsilon_i) \int \begin{pmatrix} 1 & (\xi - u)/h \\ (\xi - u)/h & (\xi - z)^2/h^2 \end{pmatrix} K\left(\frac{\xi - z}{h}\right) \frac{d\xi}{h} \\ &= [\partial^2 \mathbb{E} \rho_w(\varepsilon_i)] \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \{1 + \mathcal{O}(1)\} = S \{1 + \mathcal{O}(1)\} \end{aligned}$$



where  $\mu_2 = \int t^2 K(t) dt$ . As  $\rho_w$  is strictly convex, its second derivative  $\mathbf{E} \partial \psi(\cdot)$  is nonnegative, which implies  $S_n$  is non-negative definite with probability tending to one. And it follows that  $\hat{\theta} = -S^{-1}A_n\{1 + o_p(1)\}$ .

It is clear that  $A_n$  is asymptotically distributed as a bivariate normal distribution  $N(m_A, \Sigma)$  with mean

$$\begin{aligned} m_A &= \frac{1}{\sqrt{h}} \mathbf{E} \left\{ \psi_w(\varepsilon_i + \delta_i) \mathbf{Z}_i \frac{K(Z_i - z)}{f_Z(z)} \right\} \\ &= \frac{1}{\sqrt{h}} \mathbf{E} \left\{ \mathbf{E} [\psi_w(\varepsilon_i + \delta_i) | Z_i] \mathbf{Z}_i \frac{K(Z_i - z)}{f_Z(z)} \right\} \\ &= \frac{1}{\sqrt{h}} \iint \psi_w\{\varepsilon + g(\xi) - g(z) - g'(z)(\xi - z)\} f_E(\varepsilon) d\varepsilon \\ &\quad \times \left( \frac{1}{(\xi - z)/h} \right) K((\xi - z)/h) d\xi \\ &= \frac{1}{2} h^{5/2} g''(z) \mu_2 [\partial \mathbf{E} \psi_w(\varepsilon)] \begin{pmatrix} 1 \\ 0 \end{pmatrix} \{1 + o(1)\}. \end{aligned}$$

As  $m_A = o(1)$ , it can be seen that the asymptotic covariance matrix  $\Sigma$  is

$$\begin{aligned} \Sigma &= \frac{1}{h} \mathbf{E} \left\{ \psi_w^2(\varepsilon_i + \delta_i) \mathbf{Z}_i \mathbf{Z}_i^\top \left[ \frac{K(Z_i - z)}{f_Z(z)} \right]^2 \right\} \{1 + o(1)\} \\ &= \frac{1}{h} \int \int \psi_w^2\{\varepsilon + g(\xi) - g(z) - g'(z)(\xi - z)\} f_E(\varepsilon) d\varepsilon \\ &\quad \times \begin{pmatrix} 1 & (\xi - z)/h \\ (\xi - z)/h & (\xi - z)^2/h^2 \end{pmatrix} \frac{K^2(Z_i - z)}{f_Z(z)} d\xi \{1 + o(1)\} \\ &= \frac{1}{f_Z(z)} \mathbf{E}[\psi_w^2(\varepsilon)] \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \{1 + o(1)\}. \end{aligned}$$

where  $\nu_k = \int t^k K^2(t) dt$ .

Thus we finally obtain that as  $n$  tends to infinity,  $\hat{\theta} - S^{-1}m_A$  converges in distribution to  $N(0, S^{-1}\Sigma S^{-1})$ . Slight algebra gives

$$\begin{aligned} S^{-1}m_A &= \frac{1}{2} h^{5/2} g''(z) \mu_2 [\partial \mathbf{E} \psi_w(\varepsilon)] \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ S^{-1}\Sigma S^{-1} &= \frac{\mathbf{E}[\psi_w^2(\varepsilon)]}{\{[\partial^2 \mathbf{E} \rho_w(\varepsilon)]\}^2 f_Z(z)} \begin{pmatrix} 1 & 0 \\ 0 & \mu_2^{-1} \end{pmatrix} \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mu_2^{-1} \end{pmatrix} \\ &= \sigma_w^2 \frac{1}{f_Z(z)} \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2/\mu_2^2 \end{pmatrix}. \end{aligned}$$

This completes the proof. □

Considering the algorithm in (2.11) and (2.12), for both step involves a plug-in estimator, therefore it would be important for us to check the closeness between  $\tilde{L}_n(\beta) - L_n(\beta)$  and  $\tilde{L}_{n,x}(\beta) - L_{n,x}(\beta)$ . We have the following Lemma.

LEMMA 3. *Assuming Condition 1-5, for a ball with  $B_{\beta^*} \stackrel{\text{def}}{=} \{\beta : \|\beta - \beta^*\| \leq \sqrt{q/n}\}$ , (similar argument see Mammen (2012) Lemma 2)*

$$\begin{aligned} & \sup_{\beta \in B_{\beta^*}} |n^{-1} \sum_{i=1}^n \rho_{\mathbf{w}}(Y_i - a_j - b_j X_{ij}^\top \beta) \omega_{ij}(\beta) - n^{-1} \sum_{i=1}^n \rho_{\mathbf{w}}(Y_i - a_j - b_j X_{ij}^\top \beta^*) \omega_{ij}(\beta^*)| \\ &= \mathcal{O}_p(h^{-1} \sqrt{q/n} \Gamma_n), \end{aligned}$$

with  $\Gamma_n$  a slowly varying function. So we need to make sure  $h^{-1} \sqrt{q/n} = \mathcal{O}(1)$ .

$$\text{Also, } B_{a_j} \stackrel{\text{def}}{=} \{\tilde{a}_j : \|a_j - \tilde{a}_j\| \leq \sqrt{nh}^{-1/2}\}, B_{b_j} \stackrel{\text{def}}{=} \{\tilde{b}_j : \|b_j - \tilde{b}_j\| \leq \sqrt{nh}^{-1/2}\}$$

$$\begin{aligned} & \sup_{B_{a_j}, B_{b_j}} |n^{-1} \sum_{j=1}^n \sum_{i=1}^n \rho_{\mathbf{w}}(Y_i - \tilde{a}_j - \tilde{b}_j X_{ij}^\top \beta) \omega_{ij}(\beta) \\ & \quad - n^{-1} \sum_{j=1}^n \sum_{i=1}^n \rho_{\mathbf{w}}(Y_i - a_j - b_j X_{ij}^\top \beta) \omega_{ij}(\beta)| \\ &= \mathcal{O}_p(n^{-1/2} \Gamma_n + \sqrt{q/n} \Gamma_n). \end{aligned}$$

6.2. *Application.* The macroprudential variables are the same as suggested by [1] and applied and extended by [4]. The macro variables and the corresponding source are 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 Bloomberg database. The Treasury constant maturities data, 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 data period of these macro variables is from January 5, 2006 to September 5, 2012, the data frequency is daily.

Table 6 and Figure 9 about here

Table 6 shows the descriptive statistics of this series. While the mean of CYN before crisis (i.e. before September 15, 2008) is  $-0.000365$ , the mean of it in crisis (i.e. from September 15, 2008 to September 5, 2012) is a little higher, i.e.  $-0.000092$ . The standard deviation reverses, in crisis group is higher than before crisis group. The values of skewness in both groups are larger than 0. And the kurtosis of both groups are all higher than 3, which are steeper than normal distribution.

Robust Jarque Bera Test is performed. Since the  $p$  values of this test in both groups are smaller than 0.05.  $H_0$  is rejected. It indicates that log returns of CYN are not normally distributed. Stationarity is an important point in time series. Unit root test is performed. The result is that log returns of CYN are stationary. We also performed the mentioned two tests for the 20 larger firms, and found the same result, i.e. these series are all not normally distributed, but they follow stationary process.

Figure 9 is the line and symbol graph for the log returns of CYN. It can be found that the volatility between 2008 and 2010 is very high, and there are some clusters in this series.

### 6.3. Tables and Figures.

$g(\cdot)$	$\varepsilon$	$\tau$	Dev	Acc	Angle	Error	ASE
Model 1	N	0.95	1.22(0.36)	0.85(3.53)	9.874(0.079)	0.029(0.004)	0.044(0.014)
		0.50	0.74(0.25)	0.63(1.45)	9.969(0.023)	0.007(0.002)	0.003(0.002)
		0.05	1.75(0.59)	1.81(3.55)	9.829(0.123)	0.038(0.006)	0.064(0.021)
	$t$	0.95	1.79(0.76)	0.95(4.23)	9.720(0.191)	0.050(0.008)	0.111(0.037)
		0.50	0.91(0.39)	1.17(2.56)	9.951(0.039)	0.009(0.002)	0.004(0.003)
		0.05	1.92(0.79)	2.45(4.48)	9.746(0.128)	0.053(0.009)	0.122(0.049)
Model 2	N	0.95	1.68(1.88)	6.57(9.32)	9.691(0.666)	7.564(7.159)	4.769(8.771)
		0.50	1.49(1.46)	1.01(2.82)	9.780(0.401)	5.916(4.874)	1.363(2.305)
		0.05	1.50(1.73)	9.87(9.71)	9.556(0.985)	8.627(8.526)	6.145(9.168)
	$t$	0.95	2.31(1.88)	9.22(9.48)	9.610(0.668)	9.158(9.561)	5.643(6.561)
		0.50	1.77(1.59)	5.00(3.58)	9.712(0.487)	8.152(7.278)	1.785(2.814)
		0.05	3.07(1.06)	7.60(9.28)	9.695(0.551)	9.750(7.464)	4.643(4.462)
Model 3	N	0.95	0.37(0.27)	0.40(2.19)	9.989(0.016)	0.141(0.069)	0.574(0.624)
		0.50	0.11(0.08)	0.25(0.79)	9.997(0.002)	0.051(0.029)	0.076(0.049)
		0.05	0.56(0.32)	0.39(2.28)	9.978(0.025)	0.229(0.063)	0.724(0.711)
	$t$	0.95	0.32(0.24)	0.50(2.11)	9.987(0.016)	0.235(0.117)	0.759(0.798)
		0.50	0.29(0.11)	0.31(0.90)	9.994(0.008)	0.077(0.052)	0.081(0.085)
		0.05	0.42(0.26)	0.58(2.26)	9.982(0.019)	0.326(0.201)	0.861(0.863)

TABLE 1

Criteria evaluated with different models and quantiles.  $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$ , N means the error  $\varepsilon$  follows a  $N(0, 0.1)$  distribution,  $t$  means the error  $\varepsilon$  follows a  $t(5)$  distribution. In 100 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  $10^{-2}$ .

$g(\cdot)$	$\beta_{(1)}^{*}$	Dev	Acc	Angle	Error	ASE
Model 1	(a)	1.51(0.36)	1.02(3.62)	9.861(0.092)	0.135(0.105)	0.152(0.124)
	(b)	1.72(0.38)	1.35(3.94)	9.892(0.099)	0.166(0.119)	0.359(0.223)
Model 2	(a)	1.85(1.95)	7.37(9.45)	9.541(0.752)	8.135(8.352)	5.731(8.928)
	(b)	2.34(2.21)	9.54(9.88)	9.432(0.856)	8.374(8.973)	7.212(9.134)
Model 3	(a)	0.41(0.26)	0.53(2.46)	9.981(0.019)	0.259(0.122)	0.786(0.812)
	(b)	0.53(0.28)	0.62(2.87)	9.973(0.021)	0.352(0.229)	0.814(0.921)

TABLE 2

Criteria evaluated with different models. Two different  $\beta_{(1)}^{*}$ : (a)  $\beta_{(1)}^{*\top} = (5, 4, 3, 2, 1)$ , (b)  $\beta_{(1)}^{*\top} = (5, 2, 1, 0.8, 0.2)$  the error  $\varepsilon$  follows a  $N(0, 0.1)$  distribution. In 100 simulations we set  $n = 100, p = 10, q = 5, \tau = 0.95$ . 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  $10^{-2}$ .

$g(\cdot)$	Dev	Acc	Angle	Error	ASE
Model 1	1.86(0.84)	5.61(6.92)	9.891(0.225)	0.046(0.009)	0.103(0.040)
Model 2	1.85(1.65)	9.72(8.51)	9.873(0.651)	9.731(9.516)	4.971(3.121)
Model 3	0.92(0.39)	6.20(5.72)	9.952(0.041)	3.851(0.108)	9.432(1.042)

TABLE 3

Criteria evaluated with different models under  $p > n$  case.  $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$ , the error  $\varepsilon$  follows a  $N(0, 0.1)$  distribution. In 100 simulations we set  $n = 100, p = 200, q = 5, \tau = 0.05$ . 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  $10^{-2}$ .

model	settings	$\varepsilon$	Dev	Acc	Angle	Error
Model 1	$p = 10, q = 2$	N	1.51%(0.003)	0.12(0.067)	0.999	0.50%(0.001)
		$t$	3.62%(0.062)	0.43(0.421)	0.989	0.49%(0.001)
	$p = 10, q = 7$	N	8.27%(0.145)	0.31(0.118)	0.989	0.81%(0.002)
		$t$	6.44%(0.163)	0.68(0.081)	0.996	0.72%(0.002)
	$p = 100, q = 5$	N	8.50%(0.394)	1.19(0.886)	0.905	0.73%(0.001)
		$t$	6.50%(0.237)	1.82(1.023)	0.920	0.84%(0.002)
Model 2	$p = 10, q = 2$	N	0.25%(0.003)	0.03(0.015)	0.999	0.50%(0.001)
		$t$	2.22%(0.200)	0.02(0.201)	0.979	0.50%(0.001)
	$p = 10, q = 7$	N	6.13%(0.034)	0.07(0.271)	0.998	0.69%(0.001)
		$t$	5.69%(0.033)	0.18(0.393)	0.998	0.70%(0.001)
	$p = 100, q = 5$	N	8.10%(0.227)	2.42(0.214)	0.910	0.73%(0.003)
		$t$	7.50%(0.225)	2.55(1.893)	0.912	0.70%(0.001)
Model 3	$p = 10, q = 2$	N	1.04%(0.003)	0.02(0.003)	0.999	0.50%(0.001)
		$t$	1.67%(0.079)	0.02(0.611)	0.994	0.49%(0.001)
	$p = 10, q = 7$	N	5.28%(0.062)	0.08(0.246)	0.997	0.68%(0.003)
		$t$	5.49%(0.136)	0.08(0.482)	0.997	0.72%(0.002)
	$p = 100, q = 5$	N	4.12%(0.323)	1.67(0.724)	0.944	0.84%(0.004)
		$t$	8.20%(0.814)	2.38(0.706)	0.910	0.80%(0.005)

TABLE 4

Simulation results under sparsity, non-sparsity and large  $p$  cases. N means errors follow i.i.d.  $N(0, 0.1)$ ,  $t$  means  $t$  distribution with degree of 5.

model	settings	$\varepsilon$	Dev	Acc	Angle	Error
Model 1	$p = 10, q = 2$	N	2.68%(0.003)	0.07(0.003)	0.999	0.44%(0.001)
		$t$	1.74%(0.004)	0.08(0.003)	0.999	0.51%(0.001)
	$p = 30, q = 3$	N	2.14%(0.025)	0.87(0.378)	0.982	0.69%(0.001)
		$t$	2.02%(0.034)	0.14(0.018)	0.997	0.77%(0.001)
	$p = 120, q = 5$	N	4.35%(0.009)	1.14(0.099)	0.984	0.74%(0.001)
		$t$	5.40%(0.016)	1.33(0.054)	0.977	0.60%(0.001)
Model 2	$p = 10, q = 2$	N	2.79%(0.005)	0.00(0.000)	0.999	0.50%(0.001)
		$t$	1.88%(0.004)	0.00(0.000)	0.999	0.50%(0.001)
	$p = 30, q = 3$	N	2.76%(0.028)	0.04(0.204)	0.995	0.72%(0.001)
		$t$	2.46%(0.023)	0.03(0.007)	0.997	0.73%(0.001)
	$p = 120, q = 5$	N	2.80%(0.017)	0.67(0.707)	0.991	0.46%(0.001)
		$t$	1.20%(0.014)	1.45(1.213)	0.978	0.88%(0.001)
Model 3	$p = 10, q = 2$	N	3.64%(0.003)	0.00(0.000)	0.999	0.47%(0.001)
		$t$	0.91%(0.004)	0.02(0.003)	0.999	0.49%(0.001)
	$p = 30, q = 3$	N	5.84%(0.012)	0.16(0.081)	0.996	0.62%(0.001)
		$t$	2.26%(0.013)	0.64(0.003)	0.997	0.72%(0.001)
	$p = 120, q = 5$	N	9.89%(0.013)	0.44(0.387)	0.993	0.75%(0.001)
		$t$	3.20%(0.034)	0.95(0.993)	0.986	0.81%(0.001)

TABLE 5

Simulation results for Composite  $L_1$  Quantile Regression. N means errors follow i.i.d.  $N(0, 0.1)$ ,  $t$  means  $t$  distribution with degree of 5.

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

TABLE 6

Descriptive statistics of CYN

The financial firms	
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 7  
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 8  
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. (EFT)
130. Wintrust Financial Corporation (WTF)	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. (NAV)
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)	

TABLE 9  
The financial firms

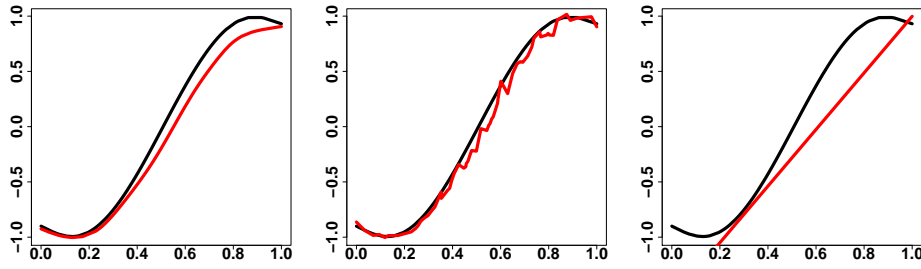


FIGURE 1. The true link functions (black) and the estimated link functions (red) in model 2 with  $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$ , and  $\varepsilon \sim N(0, 0.1)$ ,  $n = 100, p = 10, q = 5, \tau = 0.05$ , where  $h = 0.68$  (left),  $h = 0.068$  (middle) and  $h = 1$  (right).

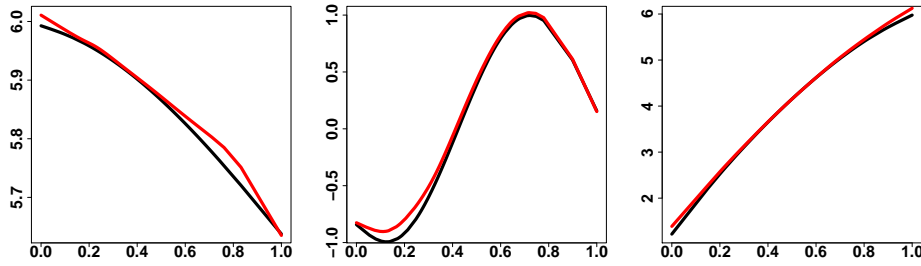


FIGURE 2. The true link functions (black) and the estimated link functions (red) with  $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$ , and  $\varepsilon \sim N(0, 0.1)$ ,  $n = 100, p = 10, q = 5, \tau = 0.95$ , model 1 (left) with  $h = 1.02$ , model 2 (middle) with  $h = 0.15$  and model 3 (right) with 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

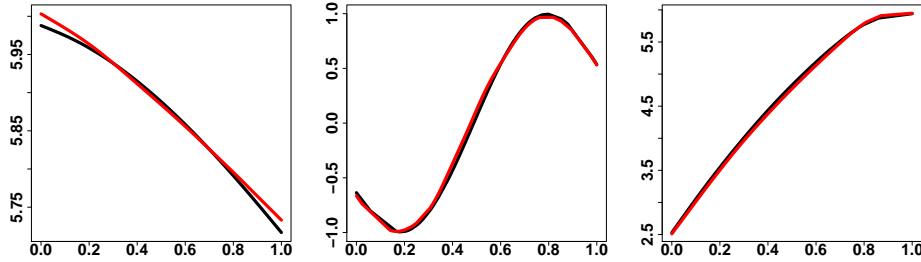


FIGURE 3. The true link functions (black) and the estimated link functions (red) with  $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$ , and  $\varepsilon \sim N(0, 0.1)$ ,  $n = 100, p = 10, q = 5, \tau = 0.5$ , model 1 (left) with  $h = 1.76$ , model 2 (middle) with  $h = 0.04$  and model 3 (right) with  $h = 0.65$ .

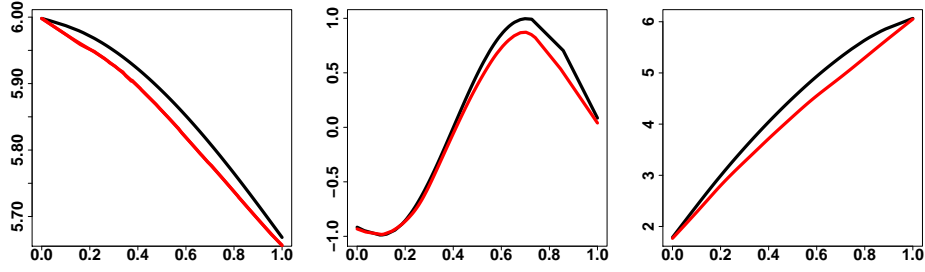


FIGURE 4. The true link functions (black) and the estimated link functions (red) with  $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$ , and  $\varepsilon \sim N(0, 0.1)$ ,  $n = 100, p = 10, q = 5, \tau = 0.05$ , model 1 (left) with  $h = 0.78$ , model 2 (middle) with  $h = 0.12$  and model 3 (right) with  $h = 1.0$ .

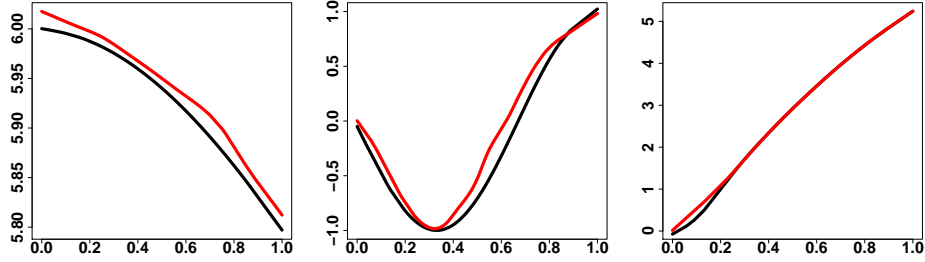


FIGURE 5. The true link functions (black) and the estimated link functions (red) with  $\beta_{(1)}^{*\top} = (5, 4, 3, 2, 1)$ , and  $\varepsilon \sim N(0, 0.1)$ ,  $n = 100, p = 10, q = 5, \tau = 0.95$ , model 1 (left) with  $h = 0.65$ , model 2 (middle) with  $h = 0.02$  and model 3 (right) with  $h = 0.33$ .

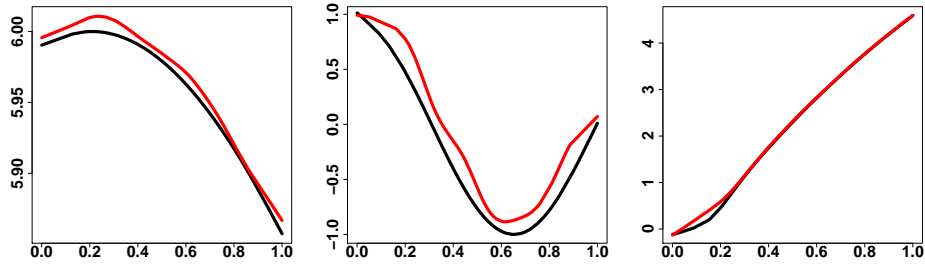


FIGURE 6. The true link functions (black) and the estimated link functions (red) with  $\beta_{(1)}^{*\top} = (5, 2, 1, 0.8, 0.2)$ , and  $\varepsilon \sim N(0, 0.1)$ ,  $n = 100, p = 10, q = 5, \tau = 0.95$ , model 1 (left) with  $h = 0.21$ , model 2 (middle) with  $h = 0.18$  and model 3 (right) with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

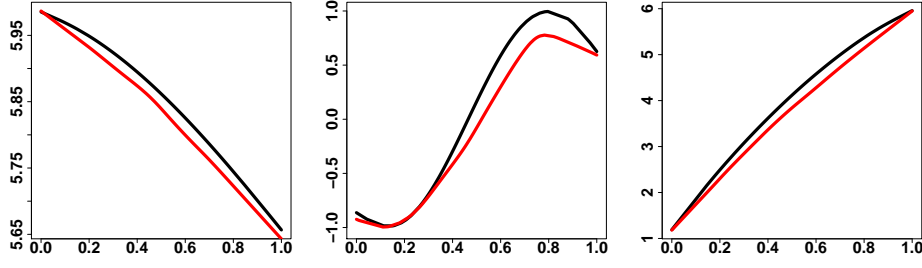


FIGURE 7. The true link functions (black) and the estimated link functions (red) with  $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$ , and  $\varepsilon \sim N(0, 0.1)$ ,  $n = 100, p = 200, q = 5, \tau = 0.05$ , model 1 (left) with  $h = 0.81$ , model 2 (middle) with  $h = 0.22$  and model 3 (right) with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

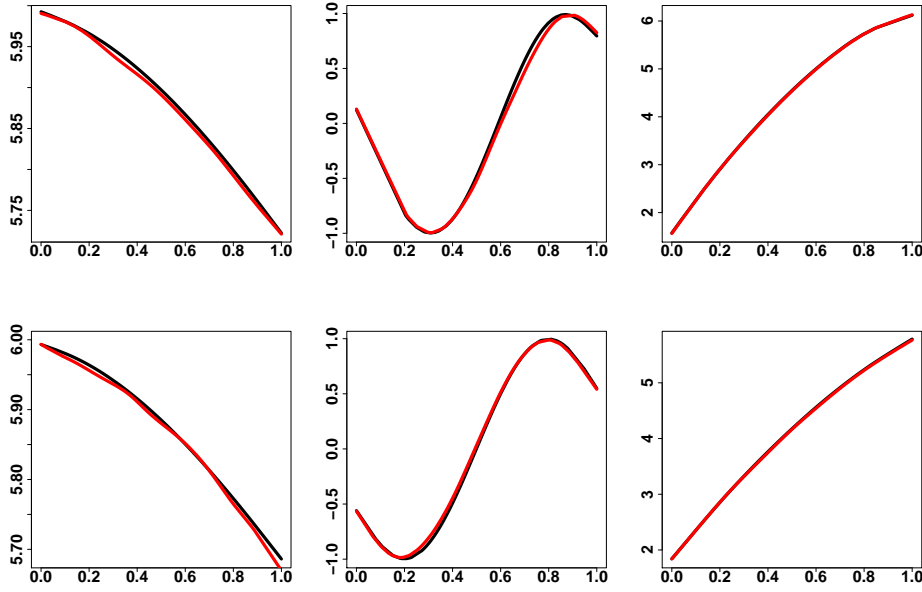


FIGURE 8. Plot of the true function  $g(\cdot)$  (black) and the estimation (red) with  $n = 100, p = 10, q = 5$  and  $\varepsilon \sim N(0, 0.1)$  in different  $g(\cdot)$  functions.  $L_1$ - $L_2$  regression (upper pannel), composite quantile (lower panel)

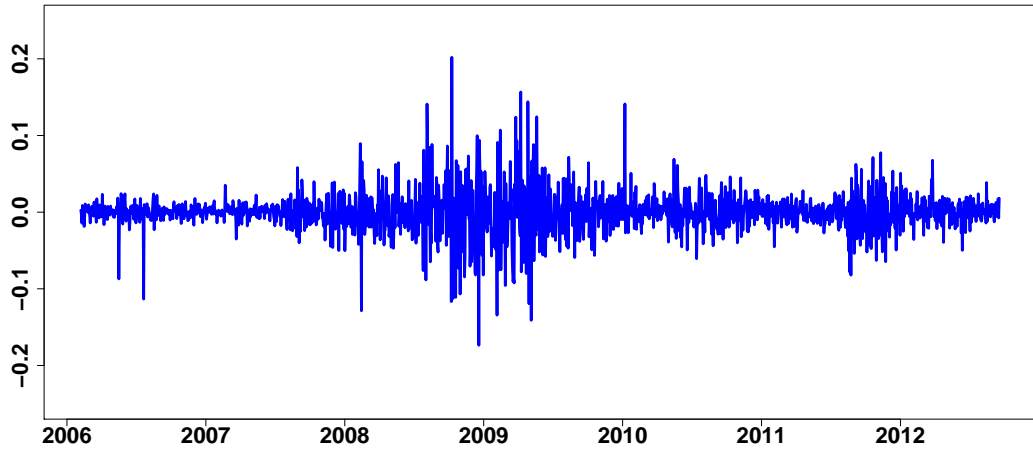


FIGURE 9. *Log returns of CYN*

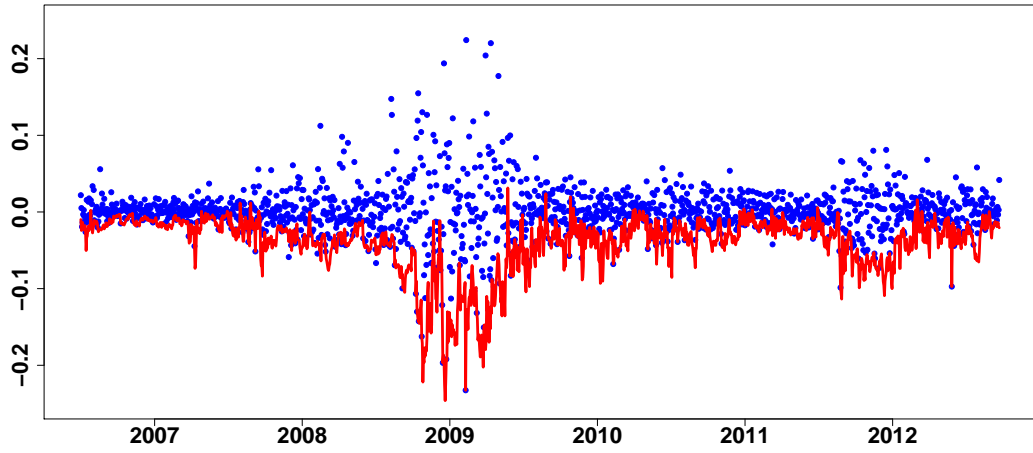


FIGURE 10. *Log returns of JPM (blue) and VaR of log returns of JPM (red),  $\tau = 0.05$ ,  $T = 1569$ , window size  $n = 100$ , refer to (5.2).*

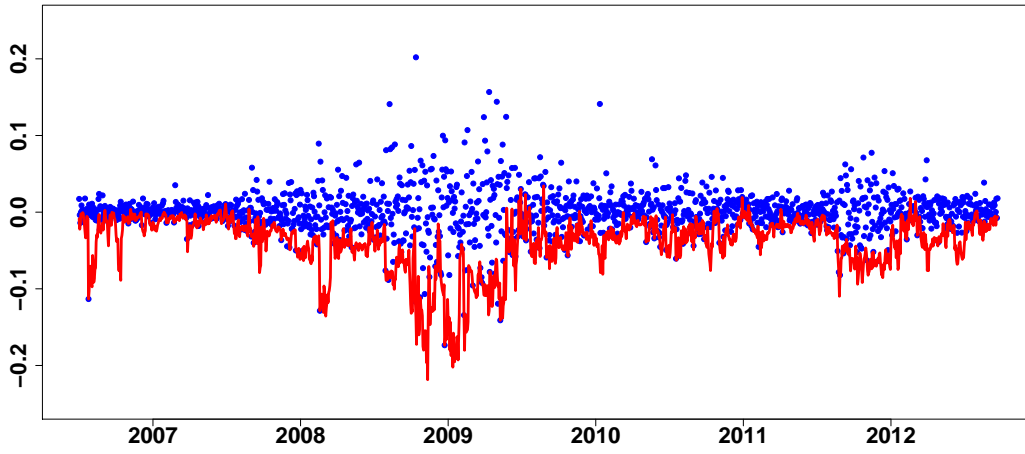


FIGURE 11. Log returns of CYN (blue) and VaR of log returns of CYN (red),  $\tau = 0.05$ ,  $T = 1569$ , window size  $n = 100$ , refer to (5.2).

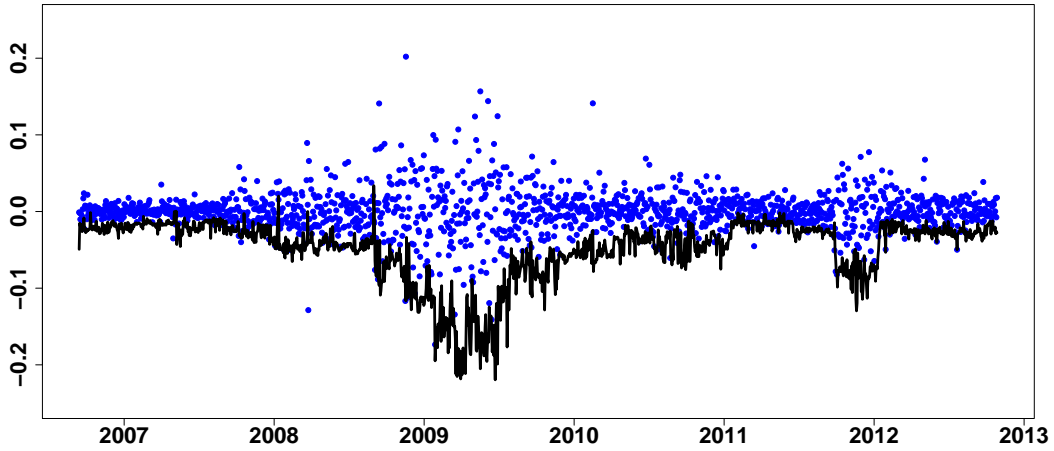


FIGURE 12. Log returns of CYN (blue) and the estimated CoVaR (black),  $\tau = 0.05$ ,  $T = 1543$ , window size  $n = 126$ , refer to (5.4).

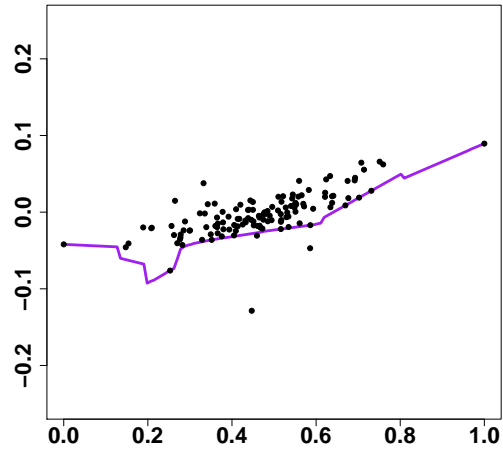


FIGURE 13. The estimated link function, window size  $n = 126$ , starting date: 20080707,  $\tau = 0.05$ ,  $h = 0.027$ ,  $p = 206$ ,  $\hat{q} = 3$  (where  $\hat{q}$  is the number of selected variables in this window): FHN, MBI, RDN.

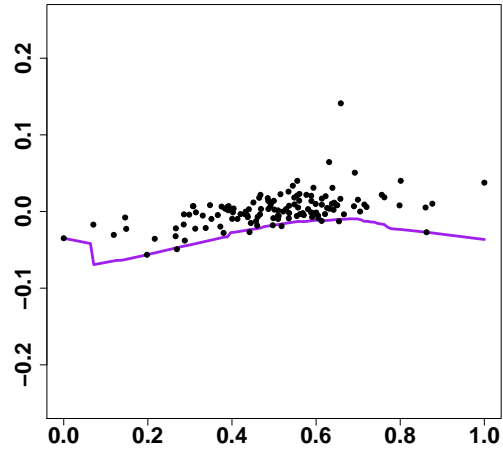


FIGURE 14. The estimated link function, window size  $n = 126$ , starting date: 20100308,  $\tau = 0.05$ ,  $h = 0.056$ ,  $p = 206$ ,  $\hat{q} = 5$ : ZION, EWBC, CNO, SNV, RDN.

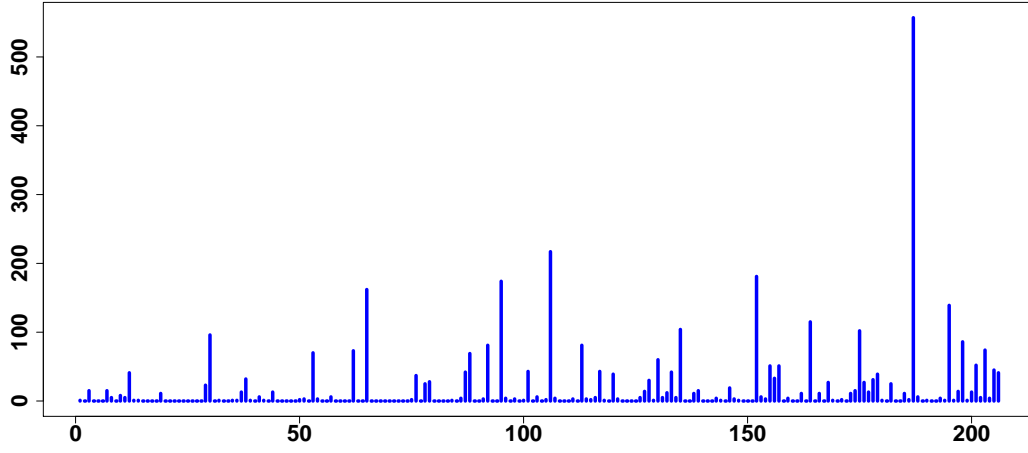


FIGURE 15. The frequency of the firms and macroprudential variables. The X-axis: 1 – 206 variables, and the Y-axis: the frequency of the variables selected in the moving window estimation. The variable 187, i.e. "Radian Group Inc. (RDN)" is the most frequently selected variable with frequency 557.

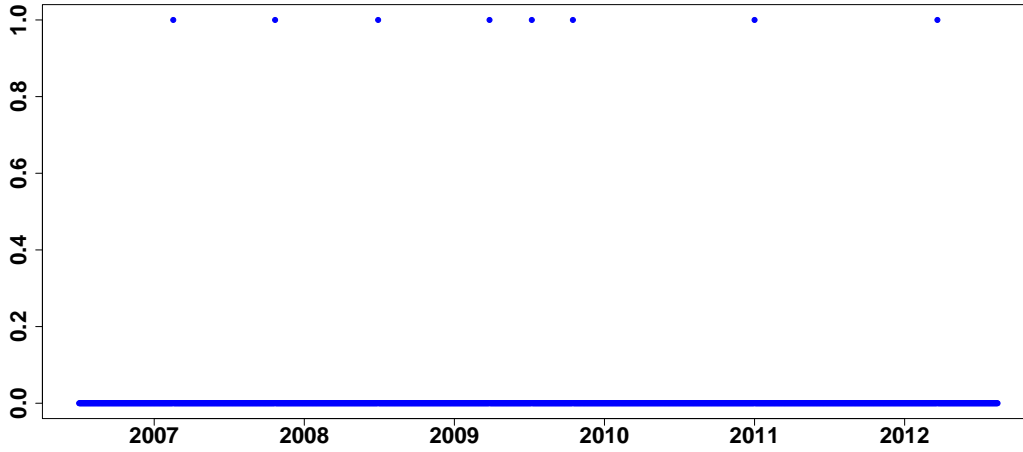


FIGURE 16. The top dots are the violations (i.e.  $\{t : I_t = 1\}$ ) of  $\widehat{VaR}$  of CYN, totally 8 violations,  $T = 1543$ ,  $\hat{\tau} = 0.005$ .



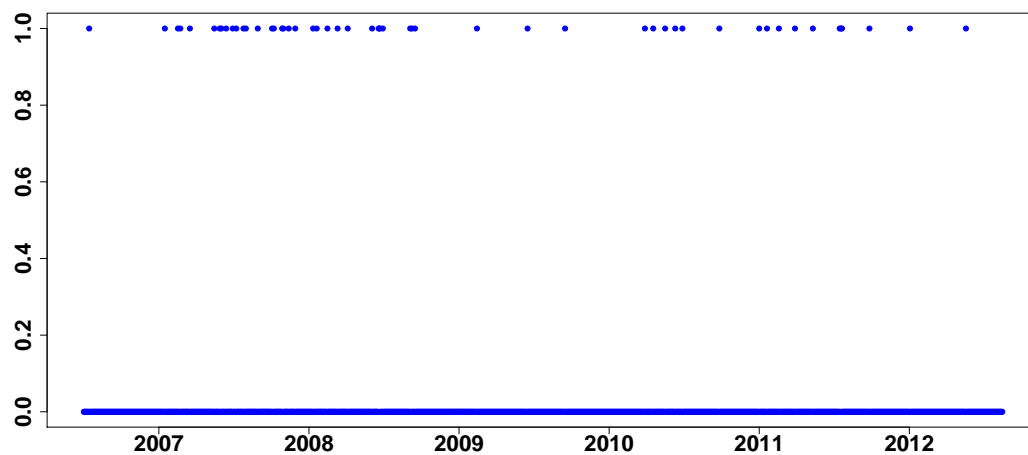


FIGURE 17. The top dots are the violations (i.e.  $\{t : I_t = 1\}$ ) of  $\widehat{\text{CoVaR}}$  of CYN, totally 53 violations,  $T = 1543$ ,  $\hat{\tau} = 0.034$ .

## References.

- [1] ADRIAN, T. and BRUNNERMEIER, M. K. (2011). CoVaR. *Staff Reports 348, Federal Reserve Bank of New York*.
- [2] BERKOWITZ, J., CHRISTOFFERSEN, P. and PELLETIER, D. (2009). Evaluating value-at-risk models with desk-level data. *Working Paper 010, North Carolina State University, Department of Economics*.
- [3] BRADIC, J., FAN, J. and WANG, W. (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *J. R. Statist. Soc. B.* **73** (3) 325–349.
- [4] CHAO, S. K., HÄRDLE, W. K. and WANG, W. (2012). *Quantile regression in Risk Calibration*. In *Handbook for Financial Econometrics and Statistics (Cheng-Few Lee, ed.)*. Springer Verlag, forthcoming, SFB 649 DP 2012-006.
- [5] ENGLE, R. F. and MANGANELLI, S. (2004). CaViaR: Conditional autoregressive value at risk by regression quantiles. *J. Bus. Econ. Stat.* **22** 367–381.
- [6] FAN, J. and LI, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *J. Amer. Statist. Assoc.* **96** 1348–1360.
- [7] HÄRDLE, W. and STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** 986–995.
- [8] HUBER, P. J. (1985). Projection pursuit. *Ann. Statist.* **13** 435–475.
- [9] KAI, B., LI, R. and ZOU, H. (2010). Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *J. R. Statist. Soc. B.* **72** 49–69.
- [10] KAI, B., LI, R. and ZOU, H. (2011) New Efficient Estimation and Variable Selection Methods for Semiparametric Varying-Coefficient Partially Linear Models. *Ann. Statist.* **39** (1) 305–332.
- [11] KOENKER, R. and BASSETT, G. W. (1978). Regression quantiles. *Econometrica*. **46** 33–50.
- [12] KOENKER, R. and BASSETT, G. W. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*. **50** 43–61.
- [13] KOENKER, R. and HALLOCK, K. F. (2001). Quantile regression. *Journal of Econometric Perspectives*. **15** (4) 143–156.
- [14] KONG, E. and XIA, Y. (1994). Variable selection for the single-index model. *Biometrika*. **94** 217–229.
- [15] LENG, C., XIA, Y. and XU, J. (2008). An adaptive estimation method for semiparametric models and dimension reduction. *WSPC-Proceedings*.
- [16] LI, Y. and ZHU, J. (2008). L1- norm quantile regression. *J. Comput. Graph. Stat.* **17** 163–185.
- [17] NEWEY, W. and POWELL, J. (1987). Asymmetric least squares estimation and testing. *Econometrica*. **55** 819–847.
- [18] RUPPERT, D., SHEATHER, S. J. and WAND, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90** 1257–1270.
- [19] SCHNABEL, S. and EILERS, P. (2009). Optimal expectile smoothing. *Comput. Stat. Data. An.* **53** (12) 4168–4177.
- [20] SERING, R. J. (2001). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- [21] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Statist. Soc. B.* **58** (1) 267–288.
- [22] WANG, Q. and YIN, X. (2008). A nonlinear multi-dimensional variable selection methods for high-dimensional data: sparse mave. *Comput. Stat. Data. An.* **52** 4512–4520.
- [23] WU, T. Z., YU, K. and YU, Y. (2010). Single-index quantile regression. *J. Multivariate Anal.* **101** 1607–1621.
- [24] XIA, Y., TONG, H., LI, W. and ZHU, L. (2002). An adaptive estimation of dimension reduction space. *J. R. Statist. Soc. B.* **64** 363–410.
- [25] YU, K. and JONES, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93** 228–237.
- [26] YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Statist. Soc. B.* **68** (1) 49–67.
- [27] ZENG, P., HE, T. H. and ZHU, Y. (2012). A lasso-type approach for estimation and variable selection

- in single-index models. *J. Comput. Graph. Stat.*, **21** 92–109.
- [28] ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 476.
- [29] ZOU, H. and YUAN, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.* **36** (3) 1108–1126.

# SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Functional Data Analysis of Generalized Quantile Regressions" by Mengmeng Guo, Lhan Zhou, Jianhua Z. Huang and Wolfgang Karl Härdle, January 2013.
- 002 "Statistical properties and stability of ratings in a subset of US firms" by Alexander B. Matthies, January 2013.
- 003 "Empirical Research on Corporate Credit-Ratings: A Literature Review" by Alexander B. Matthies, January 2013.
- 004 "Preference for Randomization: Empirical and Experimental Evidence" by Nadja Dwenger, Dorothea Kübler and Georg Weizsäcker, January 2013.
- 005 "Pricing Rainfall Derivatives at the CME" by Brenda López Cabrera, Martin Odening and Matthias Ritter, January 2013.
- 006 "Inference for Multi-Dimensional High-Frequency Data: Equivalence of Methods, Central Limit Theorems, and an Application to Conditional Independence Testing" by Markus Bibinger and Per A. Mykland, January 2013.
- 007 "Crossing Network versus Dealer Market: Unique Equilibrium in the Allocation of Order Flow" by Jutta Dönges, Frank Heinemann and Tijmen R. Daniëls, January 2013.
- 008 "Forecasting systemic impact in financial networks" by Nikolaus Hautsch, Julia Schaumburg and Melanie Schienle, January 2013.
- 009 "'I'll do it by myself as I knew it all along': On the failure of hindsight-biased principals to delegate optimally" by David Danz, Frank Hüber, Dorothea Kübler, Lydia Mechtenberg and Julia Schmid, January 2013.
- 010 "Composite Quantile Regression for the Single-Index Model" by Yan Fan, Wolfgang Karl Härdle, Weining Wang and Lixing Zhu, February 2013.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
**<http://sfb649.wiwi.hu-berlin.de>**

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

