Quantile Lasso Regression for Single Index Model

Master Thesis Submitted to

Prof. Dr. Wolfgang Karl Härdle

Prof. Dr. Weining Wang

Ladislaus von Bortkiewicz Chair of Statistics

C.A.S.E.- Centre for Applied Statistics and Economics

Humboldt-Universität zu Berlin

by

Lining Yu

(531992)

in partial fulfillment of the requirements

for the degree of

Master of Science in Statistics

Berlin, August 2, 2014

Abstract

In financial market there are many different risk factors surrounding a specified financial firm. For example, credit risk, liquidity risk and market risk. Other firms can affect this firm as well. To identify the relevant risk factors and to detect the possible contagion effects from other firms to this specified firm are important. Conditional value at risk (CoVaR) can measure these risks and will be applied in this paper. To estimate CoVaR quantile regresssion is a basic method. Since the impact from other risk factors to this specified financial firm is often nonlinear, single index model (SIM) as a semiparametric estimation plays an important role. Selecting the relevant risk factors can be solved by variable selection technique. Briefly, quantile regression for single index model associated with variable selection technique would be carried out in terms of financial data in this paper, the evaluation would be conducted by Backtesting.

Keywords: Value at Risk, Conditional Value at Risk, Semiparametric, Single index model, Backtesting

Acknowledgements

I appreciate Prof. Wolfgang Karl Härdle, he gave me the opportunity to do this paper, during this period, he always helps me to make this paper better and better.

Prof. Weining Wang supervises me and helps me both in theoretical part and empirical application. She is very patient to answer my questions.

Doctor Andrija Mihoci, PhD students Shi-kang Chao and Jing Huang also gave me a lot of valuable help on this paper, they are very nice.

I would like to thank in particular my husband Yang Wang, he always supports me both in my study and in my life, he continuously encourages me when I face some difficulties.

In addition, thanks my boy Yinchuan Wang' coming before the graduation of my Master degree. He is so cute, and makes my life more happy.

Finally, I would like to appreciate my parents and my husband's parents, they took good care of me and my boy in this special period.

Contents

1 Introduction

In the second half of 2008, the financial crisis started to hit the world. Many financial institutions are threatened, some of them even had to declare bankruptcy, for example, Lehman Brothers, National City Bank, Commerce Bancorp. These events alert other financial firms to look for the reasons of bankruptcy. First of all, the risk factors play the important role. There are some major risk factors including liquidity risk, credit risk, market risk and operational risks. Secondly, the contagion effects caused by other banks are very crucial. As a financial institution, to identify which kind of risks are more influential and which financial institutions have more impact to it are important. This paper applies the statistical model, and tries to give an effective proposal of identifying the influential factors for some financial institutions.

In the first step, the first quantile regression is conducted to show impact of some macro-prudential factors on some financial institutions.

In the second step, the second quantile regression by applying single index model associated with variable selection technique is carried out. It can reveal the impact not only of some macro factors, but also the contagion effects from some financial institutions on a specified financial firm.

This paper is organized as follows: in Section 2, the basic concepts are introduced. In Section 3, the estimation methodology is stated. In Section 4, simulation is conducted. In Section 5, application in terms of financial data is carried out. In Section 6, the research is concluded. Some details can be found in appendix.

2 Basic concepts

In this section some basic concepts used in this paper will be introduced.

2.1 VaR

VaR (Value at Risk) is a widely applied risk measure which can be intuitively understood. It was originally used by Dennis Weatherstone, CEO of J.P. Morgan and his staff. Nowadays many people applied VaR in financial market. The VaR of a financial institution i at $\tau \in (0,1)$:

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

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

2.2 CoVaR

Adrian, T. and Brunnermeier, M. K. (2011) proposed CoVaR (Conditional Value at Risk) which takes contagion effects and some conditional events into account. It can better explain the impact of different risk sources on a specified financial institution.

The CoVaR of a risk factor j given X_i at level $\tau \in (0,1)$ and at time t:

$$\mathrm{P}\big\{X_{j,t} \leq CoVaR_{j|i,t}^{\tau} | X_{i,t} = VaR^{\tau}(X_{i,t}), M_{t-1}\big\} \stackrel{def}{=} \tau,$$

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

2.3 Quantile regression

Since both VaR and CoVaR are τ-quantiles of asset return, this motivates the quantile regression estimation which was introduced by Koenker, R. and Bassett, G. W. (1978).

For any real valued random variable X with cumulative distribution function $F_X(x) = P(X \leq x)$. The τth quantile of X is given by:

$$Q_X(\tau) = F_X^{-1}(\tau) = \inf\left\{x : F_X(x) \geq \tau\right\}.$$

where $\tau \in (0, 1)$.

Suppose the τth conditional quantile function is $Q_y(\tau|x) = x^\top \beta(\tau)$, for given data (y_i, x_i), $\hat{\beta}(\tau)$ can be estimated by:

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top \beta).$$

where $\rho_\tau(u) = u\{\tau - \mathbf{1}(u < 0)\}$ and $\tau \in (0, 1)$.

Quantile regression has been commonly used in many fields in reality. Different from Linear regression estimation, Quantile regression can capture the outliers well. In financial area people focus on VaR, i.e. negative tail event, not the conditional mean calculated by linear regression model. That is why the Quantile regression is so attractive.

2.4 Single index model

Adrian, T. and Brunnermeier, M. K. (2011) used linear quantile regression model to estimate the CoVaR. In fact, the impact of some other firms is often nonlinear. Nonparametric method maybe a candidate, but the curse of dimensionality obstructs people to go further. A semiparametric model would be a good choice. Chao, S. K., Härdle, W. K. and Wang, W. (2012) proposed a partial linear model to estimate CoVaR, but it only can measure the impact from one firm to a specified firm, can not measure the impact from many other firms to a specified firm. In this paper this problem can be solved by applying Single index model (SIM), this model can not only solve the nonlinear problem from other firms, but also can show the impact from many other firms to a specified firm and at the same time select the most influential risk factors for this firm.

It is known that there are many different risks surrounding a specified financial firm. But these covariates are too much, the dimension of the explanatory variables needs to be reduced. In order to solve this problem single index model is therefore be applied. Single index model can reduce the ultra high dimensional explanatory variables into one dimensional index which can be intuitively interpreted.

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 (δ is a constant). The single index model is defined to be:

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

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

Fan, Y., Härdle, W. K., Wang, W., and Zhu, L. (2013) applied a minimum contrast approach (MACE) which would be used in this paper. In a quasi maximum likelihood (or equivalently a minimum contrast) framework the direction β (for known $g(\cdot)$) is the solution of

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

where for quantile regression:

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

First, $g(X_i^{\top}\beta)$ for x near X_i can be approximated by Taylor expansion:

$$g(X_i^{\top}\beta) \approx g(x^{\top}\beta) + g'(x^{\top}\beta)(X_i - x)^{\top}\beta, \tag{2.4}$$

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

$$L_x(\beta) \stackrel{\text{def}}{=} \mathsf{E}\, \rho_{\tau}\{Y - g(x^{\top}\beta) - g'(x^{\top}\beta)(X_i - x)^{\top}\beta\}, \tag{2.5}$$

The empirical version of (2.5) requires minimizing, with respect to β:

$$L_{n,x}(\beta) \stackrel{\text{def}}{=} n^{-1}\sum_{i=1}^{n} \rho_{\tau}\{Y_i - g(x^{\top}\beta) - g'(x^{\top}\beta)(X_i - x)^{\top}\beta\}K_h\{(X_i - x)^{\top}\beta\} \tag{2.6}$$

Employing now the double integration idea, i.e. integrating with respect to the empirical density function of the X variable yields as average contrast:

$$\begin{aligned} L_n(\beta) &\stackrel{\text{def}}{=} n^{-1}\sum_{j=1}^{n} L_{n,X_j}(\beta) \\ &= n^{-2}\sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau\left\{Y_i - g(\beta^\top X_j) - g'(\beta^\top X_j)\beta^\top (X_i - X_j)\right\} \\ &\quad K_h\{\beta^\top (X_i - X_j)\} \end{aligned} \tag{2.7}$$

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

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

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

$$\min_{(a_j,b_j)'\text{s},\beta} n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau\big(Y_i - a_j - b_j X_{ij}^\top \beta\big)\omega_{ij}(\beta), \tag{2.8}$$

where $X_{ij} \stackrel{\text{def}}{=} X_i - X_j$, $\omega_{ij}(\beta) \stackrel{\text{def}}{=} K_h(X_{ij}^\top \beta)/\sum_{i=1}^{n} K_h(X_{ij}^\top \beta)$. The calculation of the above minimization problem can be decomposed into two minimization problems:

- Given β, 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 β is carried out by the interior point method.

2.5 Variable selection for single index model

Although the dimension of the covariates can be reduced by single index model, some of these explanatory variables are still irrelevant to this specified financial firm. To select the most influential variables are also crucial. Variable selection is very necessary in this case.

The method introduced in Fan, Y., Härdle, W. K., Wang, W., and Zhu, L. (2013) will be applied here. Let β^* be the true value of β, $\beta^* = (\beta_{(1)}^{*\top}, \beta_{(0)}^{*\top})^\top$ with $\beta_{(1)}^* \stackrel{\text{def}}{=} (\beta_1, \ldots, \beta_q)^\top \neq 0$ and $\beta_{(0)}^* \stackrel{\text{def}}{=} (\beta_{q+1}, \ldots, \beta_p)^\top = 0$ element-wise. Accordingly 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 single index model coefficient β by solving the optimization problem, penalize the dimension p and estimate β by:

$$\min_{(a_j,b_j)'\text{s},\beta} n^{-1} \sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau\big(Y_i - a_j - b_j X_{ij}^\top \beta\big)\omega_{ij}(\beta) + \sum_{l=1}^{p} \gamma_\lambda(|\widehat{\beta}_l^{(0)}|)|\beta_l|, \tag{2.9}$$

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

Let $\widehat{\beta}^{(0)}$ be the initial estimator of β^* (linear quantile regression with variable selection). For $t = 0, 1, 2, \cdots$, given $\widehat{\beta}^{(t)}$, standardize $\widehat{\beta}^{(t)}$, such that $\|\widehat{\beta}^{(t)}\| = 1$ and first component of $\widehat{\beta}^{(t)}$ is positive, $\widehat{d}_l^{(t)} \overset{\text{def}}{=} \gamma_\lambda(|\widehat{\beta}_l^{(t)}|)$. Then compute:

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

For given $(\widehat{a}_j^{(t)}, \widehat{b}_j^{(t)})$, solve

$$\widehat{\beta}^{(t+1)} = \arg\min_{\beta} n^{-1} \sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau\big(Y_i - \widehat{a}_j^{(t)} - \widehat{b}_j^{(t)} X_{ij}^\top \beta\big)\omega_{ij}(\widehat{\beta}^{(t)}) + \sum_{l=1}^{p} \widehat{d}_l^{(t)} |\beta_l|.$$

Note that this iterative procedure is running until the above algorithm reaches certain degree of convergence.

In this paper let $\gamma_\lambda(t) = \lambda$, i.e. the Lasso penalization term proposed by Tibshirani, R. (1996) is applied, then (2.9) can be written as:

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

It can be found that (2.10) has a L_1 loss function and a L_1-norm penalty term. Therefore the optimization problem in (2.10) is simplified to be L_1-norm quantile regression estimation problem, see Li, Y. and Zhu, J. (2008). To choose the penalization parameter λ is very important step. The generalized approximate cross-validation criterion (GACV) suggested in Yuan, M. and Lin, Y. (2006) is applied:

$$GACV(\lambda) = \frac{\sum_{i=1}^{n} \rho_\tau\{y_i - f(x_i)\}}{n - df},$$

where df is a measure of the effective dimensionality of the fitted model.

3 Model setup

In this part, the construction of the models and the corresponding methodology are stated.

3.1 CoVaR estimation models

In financial market quantile regression for single index model can be applied in CoVaR estimation context. In order to show the performances of different CoVaR estimation methods two models are introduced here. One is quantile regression for single index model associated with variable selection technique, another is linear quantile regression model associated with variable selection technique.

3.1.1 Quantile lasso regression for single index model

Adrian, T. and Brunnermeier, M. K. (2011) applied two linear quantile regressions as follows:

$$X_{i,t} = \alpha_i + \gamma_i^\top M_{t-1} + \varepsilon_{i,t}, \tag{3.1}$$
$$X_{j,t} = \alpha_{j|i} + \beta_{j|i} X_{i,t} + \gamma_{j|i}^\top M_{t-1} + \varepsilon_{j,t}. \tag{3.2}$$

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

$$\widehat{VaR}^{\tau}_{i,t} = \widehat{\alpha}_i + \widehat{\gamma}_i^\top M_{t-1}, \tag{3.3}$$
$$\widehat{CoVaR}^{\tau}_{j|i,t} = \widehat{\alpha}_{j|i} + \widehat{\beta}_{j|i} \widehat{VaR}^{\tau}_{i,t} + \widehat{\gamma}_{j|i}^\top M_{t-1}. \tag{3.4}$$

In this paper two step regression procedure is considered as well. The first one is a quantile regression, where one regresses log returns of each covariate on all the lagged macroprudential variables, this step is the same as the first step in Adrian, T. and

Brunnermeier, M. K. (2011):

$$X_{i,t} \quad = \quad \alpha_i + \gamma_i^\top M_{t-1} + \varepsilon_{i,t}, \tag{3.5}$$

where $X_{i,t}$ represents the asset return of financial institution i at time t. The quantile regression proposed by Koenker, R. and Bassett, G. W. (1978) is applied. $(\widehat{\alpha}_i, \widehat{\gamma}_i)$ can be obtained. Then the VaR of each firm with $F^{-1}_{\varepsilon_{i,t}}(\tau|M_{t-1}) = 0$ can be calculated by:

$$\widehat{VaR}^{\tau}_{i,t} \quad = \quad \widehat{\alpha}_i + \widehat{\gamma}_i^\top M_{t-1}, \tag{3.6}$$

Then the second regression is performed using the method mentioned in single index model part which means that semiparametric estimation single index model associated with variable selection technique (L_1-norm penalty) is applied, where the response variable is log returns of one specified financial firm, the explanatory variables are the log returns of other financial firms and the lagged macroprudential variables:

$$X_{j,t} \quad = \quad g(S^\top \beta_{j|S}) + \varepsilon_{j,t}, \tag{3.7}$$

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 is very large. $g(\cdot)$ is a link function. With $F^{-1}_{\varepsilon_{j,t}}(\tau|S) = 0$ the CoVaR is estimated as:

$$\widehat{CoVaR}^{\tau}_{j|\widehat{S}} \quad = \quad \widehat{g}(\widehat{S}^\top \widehat{\beta}_{j|S}), \tag{3.8}$$

where $\widehat{S} \stackrel{def}{=} [M_{t-1}, \widehat{V}]$, where $\widehat{V}$ is the estimated VaR in (3.6).

3.1.2 Linear quantile lasso regression model

As comparison the linear L_1-norm quantile regression (i.e. linear quantile lasso regression) supposed by Li, Y. and Zhu, J. (2008) is conducted, which means that linear quantile regression model associated with L_1-norm penalty (i.e. lasso penalty) would be performed. Generalize (3.2), the two quantile regression functions are as follows:

$$X_{j,t} = S^\top \beta_{j|S} + \varepsilon_{j,t}, \tag{3.9}$$

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

$$\widehat{CoVaR}^{\tau}_{j|\widehat{S}} = \widehat{S}^\top \widehat{\beta}_{j|S}, \tag{3.10}$$

3.2 Bandwidth selection

For the single index model the bandwidth h_τ needs to be selected. Here the method proposed by Yu, K. and Jones, M. C. (1998) is implemented:

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

where h_{mean}: use direct plug-in methodology of a local linear regression described by Ruppert, D., Sheather, S. J. and Wand, M. P. (1995).

3.3 Backtesting

Then the backtesting is preformed. The days on which the log returns of a financial firm are lower than the VaR or CoVaR can be called violations. The violation sequence of financial institution i is defined as follows:

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

Generally, $I_{i,t}$ should be a martingale difference sequence. Then the CaViaR test is applied, see Berkowitz, J., Christoffersen, P. and Pelletier, D. (2009) and Chao, S. K., Härdle, W. K. and Wang, W. (2012). The CaViaR test model:

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

Note that $VaR_{i,t}$ can be replaced by $CoVaR_{i,t}$ in CoVaR estimation situation.

The test procedure is to estimate β_1 and β_2 by logistic regression, where

$$\begin{aligned} \mathrm{P}(I_{i,t} = 1 | I_{i,t-1}, VaR_{i,t}) &= \mathrm{P}(\alpha + \beta_1 I_{i,t-1} + \beta_2 VaR_{i,t} + u_{i,t} > 0 | I_{i,t-1}, VaR_{i,t}) \\ &= \Lambda(\alpha + \beta_1 I_{i,t-1} + \beta_2 VaR_{i,t}) \\ &= \frac{e^{\alpha+\beta_1 I_{i,t-1}+\beta_2 VaR_{i,t}}}{1 + e^{\alpha+\beta_1 I_{i,t-1}+\beta_2 VaR_{i,t}}} \end{aligned}$$

i.e.

$$\begin{aligned} logit(p) &= \log(\frac{p}{1-p}) \\ &= \alpha + \beta_1 I_{i,t-1} + \beta_2 VaR_{i,t} \end{aligned}$$

where $p = \mathrm{P}(I_{i,t} = 1 | I_{i,t-1}, VaR_{i,t})$.

Then Wald's test is applied with null hypothesis: $\widehat{\beta}_1 = \widehat{\beta}_2 = 0$, i.e. $I_{i,t}$ is a martingale difference sequence.

4 Simulation of quantile lasso regression for single index model

In this part, the simulation of single index models is conducted. L_1-norm quantile regression described by Li, Y. and Zhu, J. (2008) is applied here. The initial value of β can be calculated by the L_1-norm quantile regression, then the two-step iterations mentioned in single index model part are performed. Recall that X is a $p \times n$ matrix, and p is also the dimension of the true parameter β^*, $\beta^*_{(1)}$ denotes the non-zero components in β^*, q is the number of components in $\beta^*_{(1)}$, $g(\cdot)$ is the link function, n is the sample size, and τ represents the quantile level.

4.1 Different settings of the model

The evaluated model is:

$$Y_i = g(Z_i) + \varepsilon_i, \tag{4.1}$$

where $Z_i = X_i^\top \beta^*$. Assume that the jth column of X is an i.i.d. sample from $\mathrm{N}(j/2, 1)$, ε_i is the error term which follows a $\mathrm{N}(0, 0.1)$ distribution.

In the next step, different settings of $g(\cdot)$, n, p, q and τ will be considered in the simulation.

There are two $g(\cdot)$s, the first one is:

$$g(Z_i) = 5\cos(D \cdot Z_i) + \exp(-D \cdot Z_i^2), \tag{4.2}$$

where $D = 0.01$ is a scaling constant.

The second one is:

$$g(Z_i) = \sin\{\pi(A \cdot Z_i - B)\}, \tag{4.3}$$

with the parameters $A = 0.3$, $B = 3$.

Three different τ: $\tau = 0.95$, $\tau = 0.5$ and $\tau = 0.05$.

Three different $\beta^*_{(1)}$: $\beta^{*\top}_{(1)} = (5, 5, 5, 5, 5)$, $\beta^{*\top}_{(1)} = (5, 4, 3, 2, 1)$ and $\beta^{*\top}_{(1)} = (5, 2, 1, 0.8, 0.2)$.

Two different p: $p = 10$ and $p = 200$.

4.2 Criteria

To measure the accuracy for the estimation of β and $g(\cdot)$, five criteria are applied as follows:

a. Standardized L_2 norm:
$$Dev \stackrel{\text{def}}{=} \frac{\|\beta^* - \widehat{\beta}\|_2}{\|\beta^*\|_2},$$

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

c. Least angle:
$$Angle \stackrel{\text{def}}{=} \frac{< \beta^*, \widehat{\beta} >}{\|\beta^*\|_2 \cdot \|\widehat{\beta}\|_2},$$

d. Relative error:
$$Error \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^{n} \left| \frac{g(Z_i) - \widehat{g}(\widehat{Z}_i)}{g(Z_i)} \right|,$$

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

4.3 Evaluation

Different τ case is showed in Table 4.1, where $\beta^{*\top}_{(1)} = (5, 5, 5, 5, 5)$, in 100 simulations $n = 100, p = 10, q = 5$ is set. Standard deviations are given in brackets. *Dev*, *Acc*, *Angle*, *Error* and their standard deviations are reported in 10^{-1}. $ASE(h)$ and its standard deviations are reported in 10^{-2}. It can be seen that for quantile levels 0.95 and 0.05 the errors are usually slightly larger than quantile level 0.05. Although the estimation for the nonlinear model 2 are not as good as model 1, the error is still moderate. Figure 4.1 to Figure 4.3 present the plots of the true link function against the estimated ones for

$g(\cdot)$	τ	Dev	Acc	$Angle$	$Error$	$ASE(h)$
	0.95	1.22(0.36)	0.8(3.53)	9.874(0.079)	0.029(0.004)	0.044(0.014)
Model 1	0.50	0.74(0.25)	0.6(1.45)	9.969(0.023)	0.007(0.002)	0.003(0.002)
	0.05	1.75(0.59)	1.8(3.55)	9.829(0.123)	0.038(0.006)	0.064(0.021)
	0.95	1.68(1.88)	6.6(9.32)	9.691(0.666)	7.564(7.159)	4.769(8.771)
Model 2	0.50	1.49(1.46)	1.0(2.82)	9.780(0.401)	5.916(4.874)	1.363(2.305)
	0.05	1.50(1.73)	8.1(9.71)	9.556(0.985)	8.627(8.526)	6.145(9.168)

Table 4.1: Criteria evaluated under different models and quantiles.

different quantile levels.

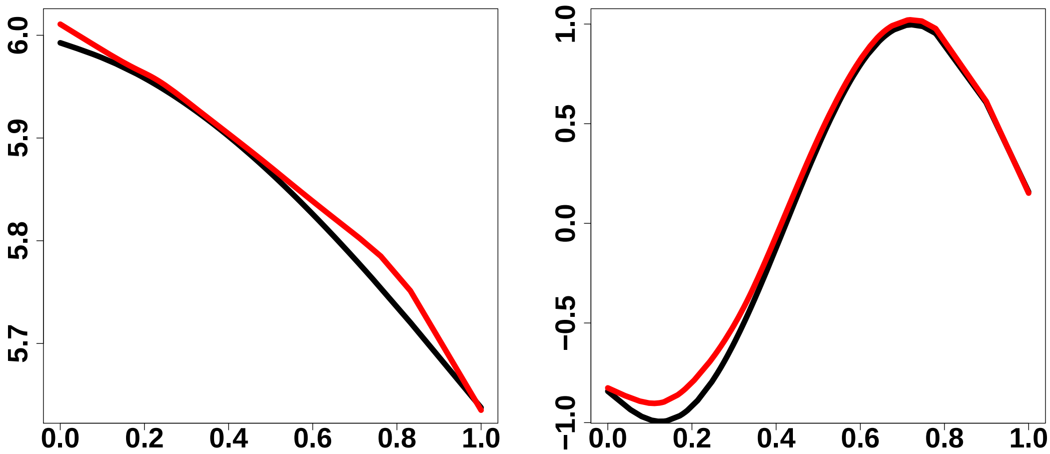

Figure 4.1: The true link functions (black) and the estimated link functions (red) with $\tau = 0.95$. Where $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$, $n = 100, p = 10, q = 5$, model 1 (left) with $h = 1.02$, model 2 (middle) with $h = 0.15$

In different $\beta_{(1)}^*$ case three different $\beta_{(1)}^*$s are given as follows: (a) $\beta_{(1)}^{*\top} = (5, 5, 5, 5, 5)$, (b) $\beta_{(1)}^{*\top} = (5, 4, 3, 2, 1)$, (c) $\beta_{(1)}^{*\top} = (5, 2, 1, 0.8, 0.2)$. See Table 4.2, in 100 simulations $n = 100, p = 10, q = 5, \tau = 0.95$ is set. Standard deviations are given in brackets. Dev, Acc, $Angle$, $Error$ and their standard deviations are reported in 10^{-1}. $ASE(h)$ and its standard deviations are reported in 10^{-2}. We notice that for the case (b) and (c), the estimation results are not better than (a) since the smaller values of $\beta_{(1)}^*$ in case (b) and (c) would be estimated as zeros, and the estimation of the link function would be affected as well. Figure 4.1, Figure 4.4 and Figure 4.5 are the plots of the estimated link functions in these three cases.

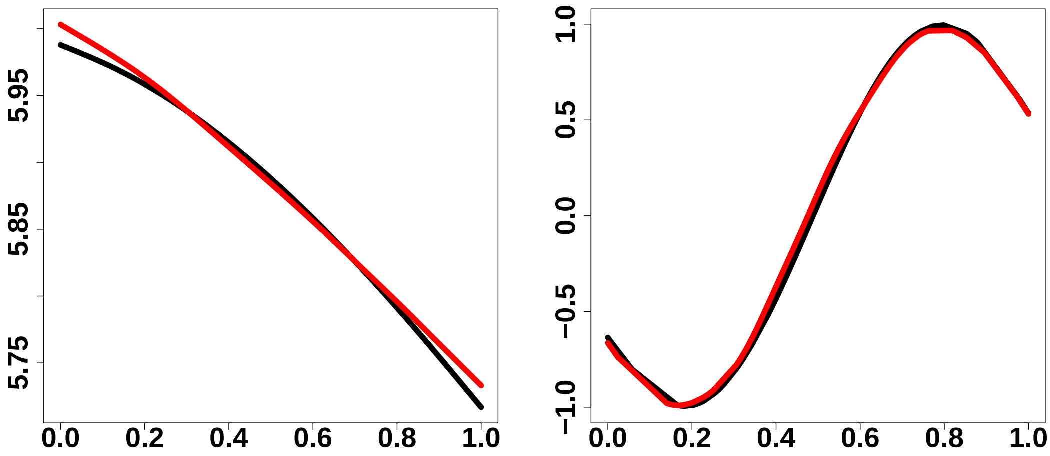

Figure 4.2: The true link functions (black) and the estimated link functions (red) with $\tau = 0.5$. Where $\beta_{(1)}^{*\top} = (5,5,5,5,5)$, $n = 100, p = 10, q = 5$, model 1 (left) with $h = 1.76$, model 2 (middle) with $h = 0.04$.

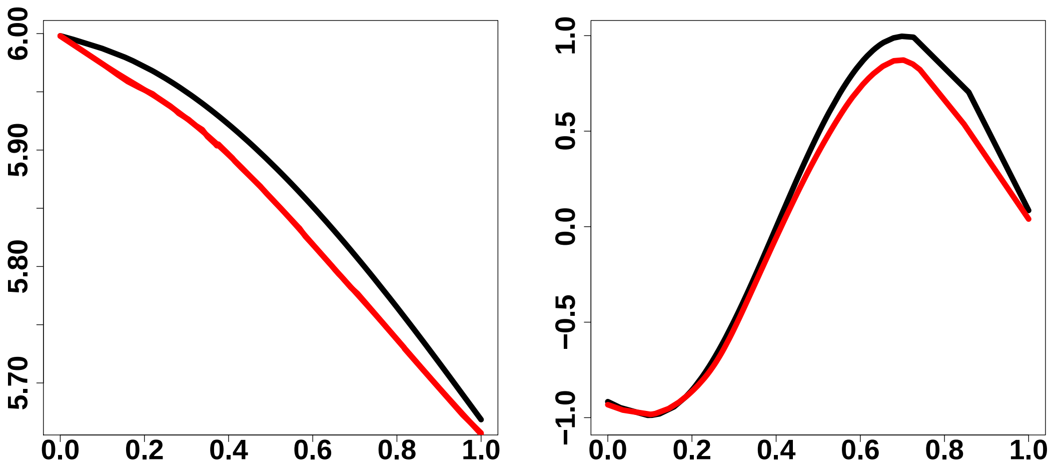

Figure 4.3: The true link functions (black) and the estimated link functions (red) with $\tau = 0.5$. Where $\beta_{(1)}^{*\top} = (5,5,5,5,5)$, $n = 100, p = 10, q = 5$, model 1 (left) with $h = 0.78$, model 2 (middle) with $h = 0.12$.

In large p case $p > n$ is detected, where $p = 200$. See Table 4.3, in 100 simulations, $\beta_{(1)}^{*\top} = (5,5,5,5,5)$, $n = 100, q = 5, \tau = 0.05$. Standard deviations are given in brackets. Dev, Acc, $Angle$, $Error$ and their standard deviations are reported in 10^{-1}, $ASE(h)$ and its standard deviations are reported in 10^{-2}. It can be found that the errors are still

$g(\cdot)$	$\beta^*_{(1)}$	Dev	Acc	$Angle$	$Error$	$ASE(h)$
	(a)	1.22(0.36)	0.8(3.53)	9.874(0.079)	0.029(0.004)	0.044(0.014)
Model 1	(b)	1.51(0.36)	1.0(3.62)	9.861(0.092)	0.035(0.005)	0.052(0.019)
	(c)	1.72(0.38)	1.3(3.94)	9.892(0.099)	0.036(0.005)	0.059(0.023)
	(a)	1.68(1.88)	6.6(9.32)	9.691(0.666)	7.564(7.159)	4.769(8.771)
Model 2	(b)	1.85(1.95)	7.4(9.45)	9.541(0.752)	8.135(8.352)	5.731(8.928)
	(c)	2.34(2.21)	9.5(9.88)	9.432(0.856)	8.374(8.973)	7.212(9.134)

Table 4.2: Criteria evaluated under three different $\beta^*_{(1)}$

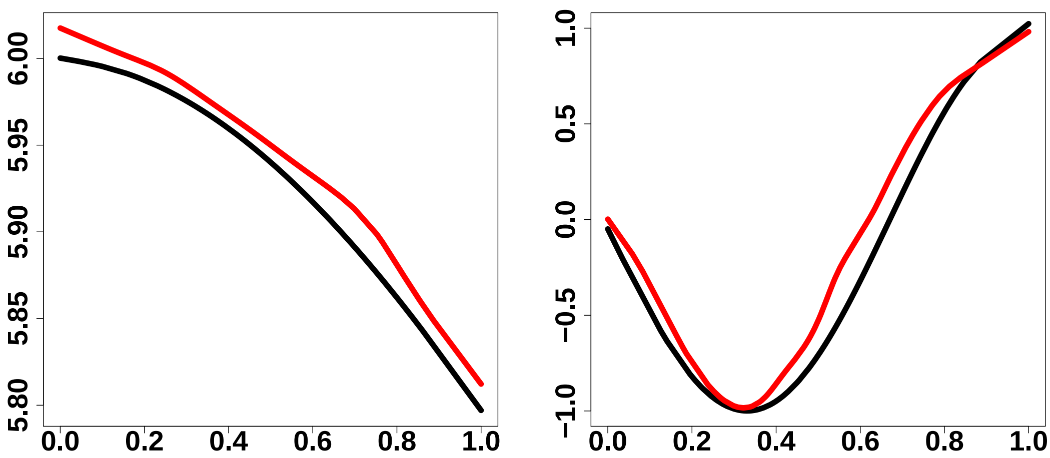

Figure 4.4: The true link functions (black) and the estimated link functions (red) with $\beta^{*\top}_{(1)} = (5, 4, 3, 2, 1)$. Where $n = 100, p = 10, q = 5, \tau = 0.95$, model 1 (left) with $h = 0.65$, model 2 (middle) with $h = 0.02$.

moderate in $p > n$ situation compared with Table 4.1, i.e. in $p = 10$ case. Figure 4.6 shows the graphs in this case.

$g(\cdot)$	Dev	Acc	$Angle$	$Error$	$ASE(h)$
Model 1	1.86(0.84)	5.6(6.92)	9.891(0.225)	0.046(0.009)	0.103(0.040)
Model 2	1.85(1.65)	9.7(8.51)	9.873(0.651)	9.731(9.516)	4.971(3.121)

Table 4.3: Criteria evaluated with different models under $p > n$ case.

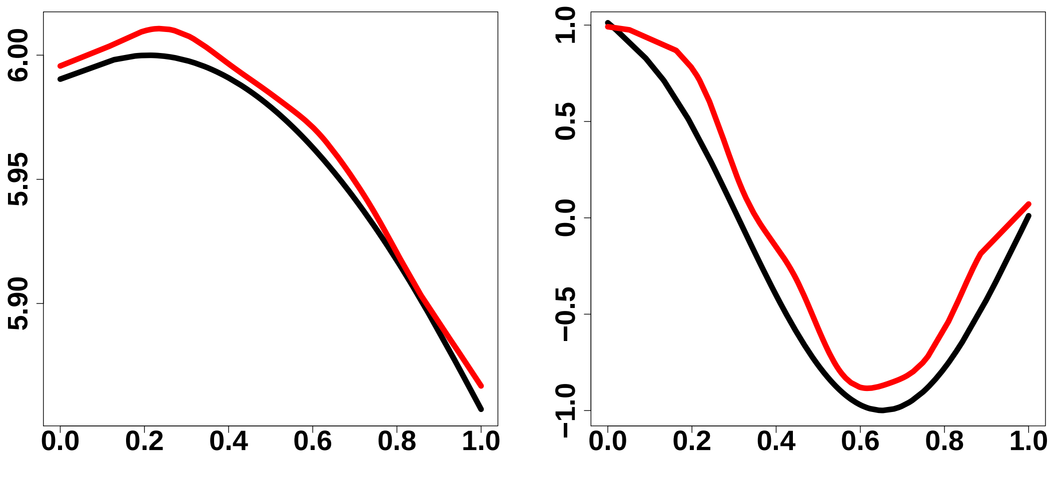

Figure 4.5: The true link functions (black) and the estimated link functions (red) with $\beta^{*\top}_{(1)} = (5, 2, 1, 0.8, 0.2)$. Where $n = 100, p = 10, q = 5, \tau = 0.95$, model 1 (left) with $h = 0.21$, model 2 (middle) with $h = 0.18$.

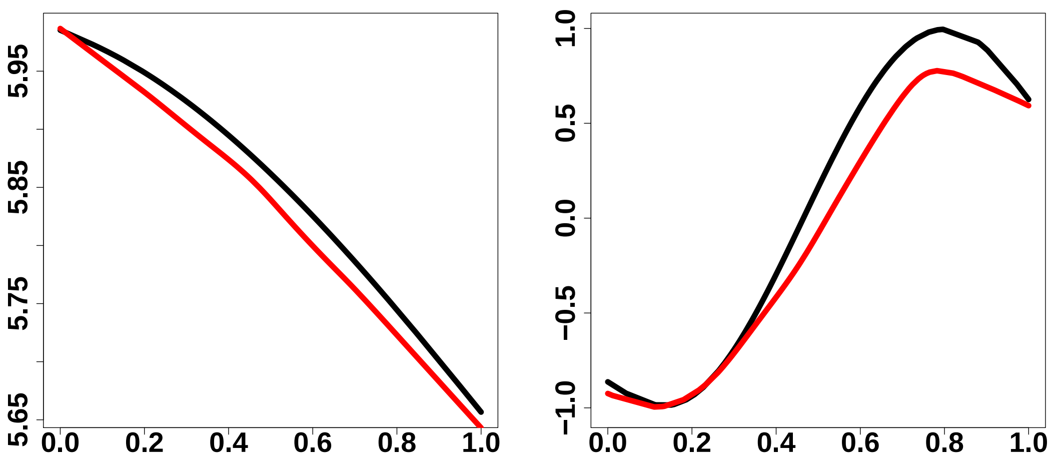

Figure 4.6: The true link functions (black) and the estimated link functions (red) with different models under $p > n$ case. Model 1 (left) with $h = 0.81$, model 2 (middle) with $h = 0.22$.

5 Empirical Applications

In this section, the methodology mentioned in model setup part will be applied in terms of financial data. One specified financial firm will be the objective and as an example. The other financial firms and some macroprudential variables are the covariates. The impact from the covariates to this specified firm will be detected.

5.1 Dataset

The firm data are selected according to the ranking of NASDAQ. The top 200 financial institutions are applied. The order of them is sorted according to their Market capitalization from high level to low level. City national corp. (CYN) which is ranked 85 among these firms is chosen as the objective. The remaining 199 financial institutions together with 7 lagged macroprudential variables are chosen as covariates, i.e. the number of the covariates $p = 206$.

The list of these financial firms comes from the website,[1] which can be found in appendix. Note that CYN is dependent variable and therefore is not in this list. The daily adjusted close stock prices of these 200 firms are from Yahoo Finance for the period from January 6, 2006 to September 6, 2012, i.e. the number of the observations $T = 1669$. Then the stock prices are transformed to log returns according the formula as follows:

$$r_t = p_t - p_{t-1}.$$

where p_t is the log stock price of one frim at time t. $p_t = lnP_t$, and P_t represents stock price of one firm at time t. r_t stands for the log return of one firm at time t.

The seven macroprudential variables are the same as suggested by Adrian, T. and Brunnermeier, M. K. (2011). Some of them were applied and extended by Chao, S. K., Härdle, W. K. and Wang, W. (2012) and Hautsch, N., Schaumburg, J. and Schienle, M. (2011).

[1] http://www.nasdaq.com/screening/companies-by-industry.aspx?industry=Finance.

These macroprudential variables and the corresponding source are as follows:

a. VIX, which measures the implied volatility in the market.

b. The short term liquidity spread, which is calculated by the difference between the 3-month Treasury repo rate and 3-month Treasury constant maturities.

c. 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.

d. 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.

e. The change in the credit spread between 10 years BAA corporate bonds and the 10 years Treasury constant maturities.

f. The daily S&P500 index returns.

g. 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 macroprudential variables is from January 5, 2006 to September 5, 2012, the data frequency is daily.

Note that for convenience of analysis the length of each variable is adjusted to be the same. For some variables which has more daily data than others have been already conducted. For example, in variable VIX there is a daily data on October 9, 2006, but there is no data in other variables on the same date, then the daily data in variable VIX is deleted.

5.2 Descriptive statistics of CYN

Table 5.1 shows the descriptive statistics of this series. While the mean of CYN in crisis period (i.e. from September 15, 2008 to February 08, 2010) is -1.7×10^{-4}, the mean of it in overall period (i.e. from July 06, 2006 to September 6, 2012) is

	Mean	SD	Skewness	Kurtosis
crisis period	-1.7×10^{-4}	0.04	0.24	5.9
overall period	-1.8×10^{-4}	0.03	0.16	10.6

Table 5.1: Descriptive statistics of CYN

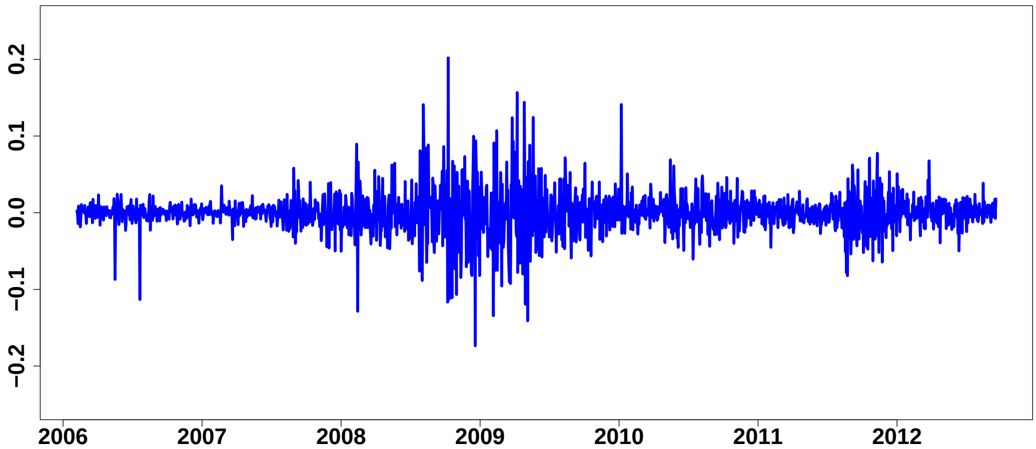

Figure 5.1: Log returns of CYN

a little lower, i.e. -1.8×10^{-4}. The standard deviation in crisis period is higher than in the overall period. It can be concluded that the log returns of CYN in the crisis time is very volatile. The values of skewness in both periods are larger than 0. And the kurtosis of both periods are all higher than 3, which are steeper than normal distribution.

Robust Jarque Bera Test is performed. Null hypothesis (H_0) of this test is: Data are normally distributed. And the alternative hypothesis (H_1) is: Data are not normally distributed. Since the p values of this test in both periods are smaller

	p value of Jarque Bera Test	p value of the Unit root test
crisis period	2.2×10^{-16}	1×10^{-4}
overall period	2.2×10^{-16}	1×10^{-6}

Table 5.2: Jarque Bera test and Unit root test of CYN

than 0.05, see Table 5.2. H_0s are rejected which indicates that log returns of CYN are not normally distributed.

Stationarity is an important point in time series. Unit root test is performed. H_0: log returns of CYN have a unit root. i.e. log returns of CYN is not stationary. H_1: log returns of CYN do not have a unit root. Table 5.2 shows the result of this test, i.e. H_0s are rejected which means that log returns of CYN are stationary in both period.

Figure 5.1 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.

5.3 Results

In this section the VaR, $CoVaR_{SIM}$ and $CoVaR_L$ would be estimated. Where $CoVaR_{SIM}$ is defined as $CoVaR$ calculated by quantile regression for single index model associated with lasso technique. And $CoVaR_L$ represents $CoVaR$ calculated by linear quantile lasso regression model. To compare the performance of these two models would be very interesting. The evaluation is carried out by backtesting.

5.3.1 Estimation of VaR, $CoVaR_{SIM}$ and $CoVaR_L$

A moving window size of $n = 126$ is set to calculate $T = 1543$ VaR of the log returns for the 199 firms. Recall (3.5), since log returns for the 199 firms are known, and $F^{-1}_{\varepsilon_{i,t}}(\tau|M_{t-1}) = 0$ is assumed, then $(\widehat{\alpha}_i, \widehat{\gamma}_i)$ can be calculated by quantile regression with $\tau = 0.05$. In (3.6) the lagged macroprudential variables are known, then VaR of the log returns for each firm can be simply predicted. Here the VaR of CYN is also calculated as comparison. Figure 5.2 and Figure 5.3 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.

Then $T = 1543$ $CoVaR_{SIM}$ of firm CYN is estimated according to (3.8). Where window size $n = 126$, the original covariates $p = 206$. Note that in this case $p > n$, i.e. there are more covariates than observations in each window size, the simple quantile regression can not solve this kind of problem, therefore quantile

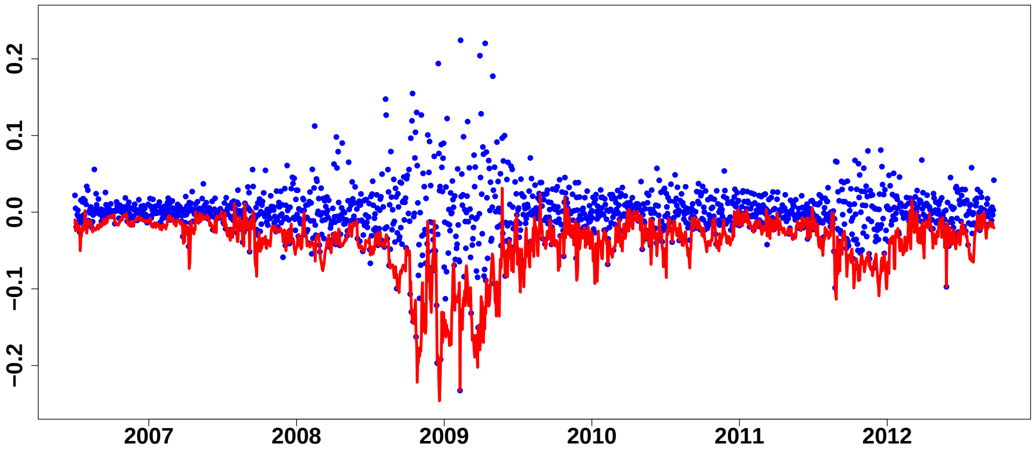

Figure 5.2: Log returns of JPM (blue) and VaR of log returns of JPM (red). Where $\tau = 0.05$, $T = 1543$, window size $n = 126$, refer to (3.6).

lasso regression for single index model is applied. Recall in (3.7 S is given here, log returns of CYN is known, $\widehat{\beta}_{j|S}$ can be estimated with $F^{-1}_{\varepsilon_{j,t}}(\tau|S) = 0$, where j represents the firm CYN in this application, and $\tau = 0.05$. Then for given $\widehat{\beta}_{j|S}$ the $CoVaR_{SIM}$ of firm CYN can be predicted by local linear smoothing as in (3.8). For the selection of penalization parameter and the bandwidth selection the methods mentioned in basic concepts part and model setup part are applied here.

There are different number of selected variables $\widehat{q}$ in each window and $T = 1543$ estimated $CoVaR_{SIM}$ by moving window estimation.

From Figure 5.4 it can be found that the estimated $CoVaR_{SIM}$ covers most lower values of log returns of CYN.

Figure 5.5 shows one example of $\widehat{\beta}$ for $CoVaR_{SIM}$ estimation in one window ($n = 126$) from 20090209 to 20090807. There are four non-zero $\widehat{\beta}$s, i.e. four selected covariates.

Figure 5.6 shows the estimated penalization parameter λ in overall period, the fluctuation seems to be time-dependent. i.e. during crisis period the fluctuation is higher than other period.

Figure 5.7 shows the frequency of the number of selected variables. $\widehat{q} = 3$, $\widehat{q} = 4$ and $\widehat{q} = 5$ have top three frequency. i.e. three, four and five selected risk factors

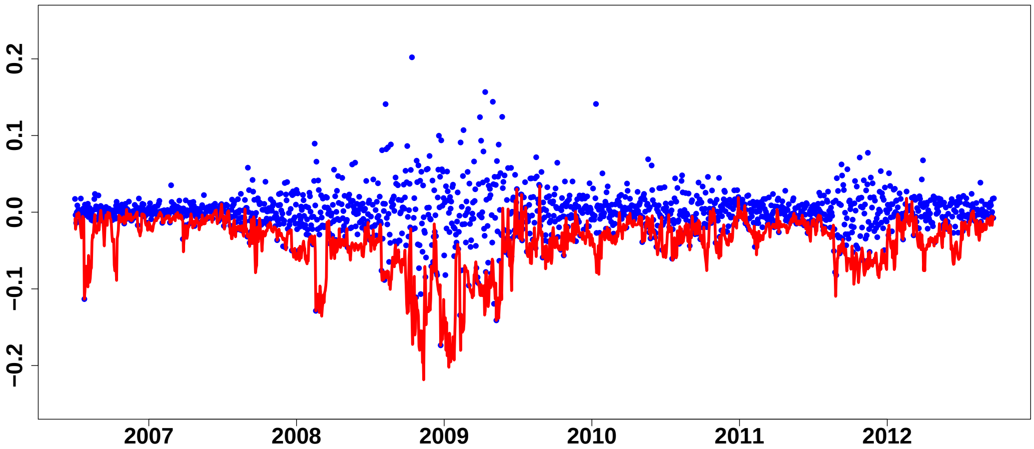

Figure 5.3: Log returns of CYN (blue) and VaR of log returns of CYN (red). Where $\tau = 0.05$, $T = 1543$, window size $n = 126$, refer to (3.6).

often have impact on firm CYN.

Figure 5.8 shows two examples of the estimated link function, nonlinear effect between the dependent variable CYN and its covariates can be therefore detected.

Figure 5.9 shows the frequency of the selected firms and macroprudential variables. Then the most frequently influential risk factors for firm CYN can be detected, and the top one influential financial firm is: Radian Group Inc. (RDN). The other selected influential covariates can be found in appendix, see Table 6.4

Next, the $CoVaR_L$ calculated by using (3.10) is estimated. Figure 5.10 shows that before 2009 almost half of the estimated $CoVaR_L$ can not cover the low log returns.

Figure 5.11 shows one example of $\hat{\beta}$ for $CoVaR_L$ estimation in one window ($n = 126$) from 20090209 to 20090807. It can be seen that there are many selected covariates which are more than in $CoVaR_{SIM}$ situation.

Figure 5.12 shows the frequency of the number of selected variables which are different from $CoVaR_{SIM}$ case.

Figure 5.13 shows the estimated penalization parameter λ in overall period. Time-dependent fluctuation is also an interesting character.

Figure 5.14 shows the frequency of the selected firms and macroprudential vari-

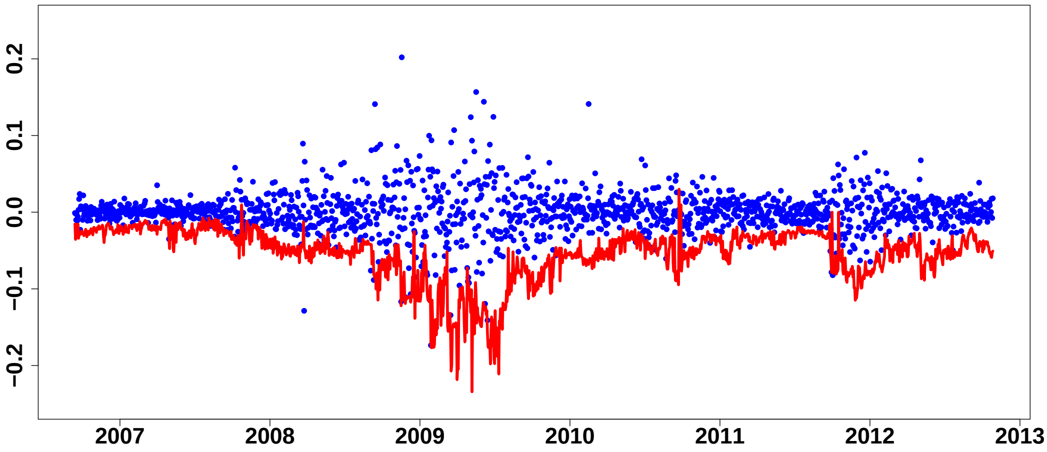

Figure 5.4: Log returns of CYN (blue) and the estimated $CoVaR_{SIM}$ (red). Where $\tau = 0.05$, $T = 1543$, window size $n = 126$, refer to (3.8).

ables. Then the most frequently influential risk factors for firm CYN by using $CoVaR_L$ estimation are different from $CoVaR_{SIM}$ estimation. The top one influential financial firm is : Flagstar Bancorp Inc. (FBC). More details of the selected firm can be found in Table 6.5

5.3.2 Backtesting for VaR, $CoVaR_{SIM}$ and $CoVaR_L$

The backtesting is preformed. For VaR estimation there are totally 14 violations, i.e. $\widehat{\tau} = 0.009$ where $T = 1543$. Figure 5.15 shows this result.

For $CoVaR_{SIM}$ estimation there are totally 19 violations (see Figure 5.16), $\widehat{\tau} = 0.012$.

For $CoVaR_L$ estimation there are even totally 231 violations, $\widehat{\tau} = 0.15$. Figure 5.17 shows this result.

The Wald's test is carried out. Table 5.3 shows the test results in the overall period. The p-value of $CoVaR_{SIM}$ is 0.54, therefore only for $\widehat{CoVaR}_{SIM}$, null hypothesis can not be rejected. $\widehat{VaR}$ and $\widehat{CoVaR}_L$ algorithms perform not so well in overall period.

Table 5.4 shows the test results during the crisis period. Null hypothesis of $\widehat{VaR}$ and

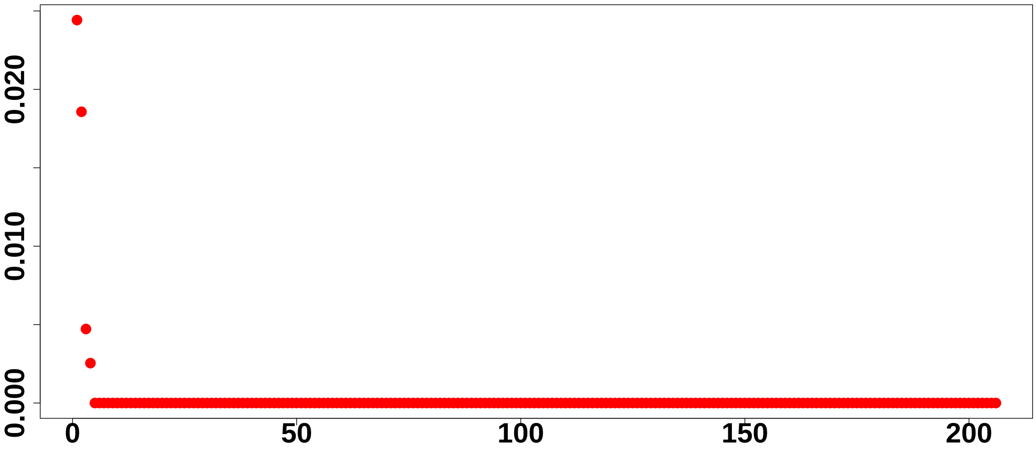

Figure 5.5: $\widehat{\beta}$ in $CoVaR_{SIM}$ estimation, 20090209-20090807.

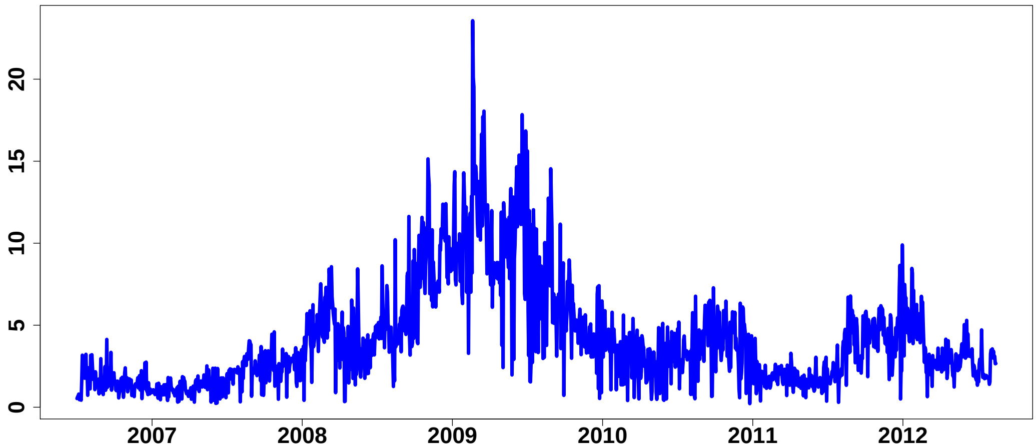

Figure 5.6: The $\widehat{\lambda}$ in $CoVaR_{SIM}$ estimation.

	p-value of Wald's test statistics
$\widehat{VaR}$	2.7×10^{-6}
$\widehat{CoVaR}_{SIM}$	0.54
$\widehat{CoVaR}_{L}$	0.00

Table 5.3: The CaViaR test for $\widehat{VaR}$, $\widehat{CoVaR}_{SIM}$ and $\widehat{CoVaR}_{L}$ for CYN, $T = 1543$, $20060706 - 20120906$.

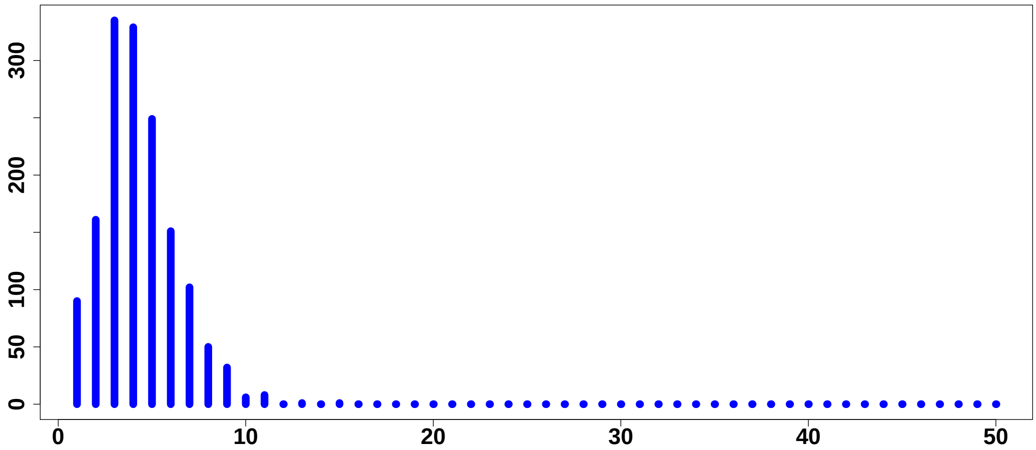

Figure 5.7: Frequency of the number of selected variables in $CoVaR_{SIM}$ estimation. Where $\widehat{q} = 3$ with frequency 335, $\widehat{q} = 4$ with frequency 329 and $\widehat{q} = 5$ with frequency 249.

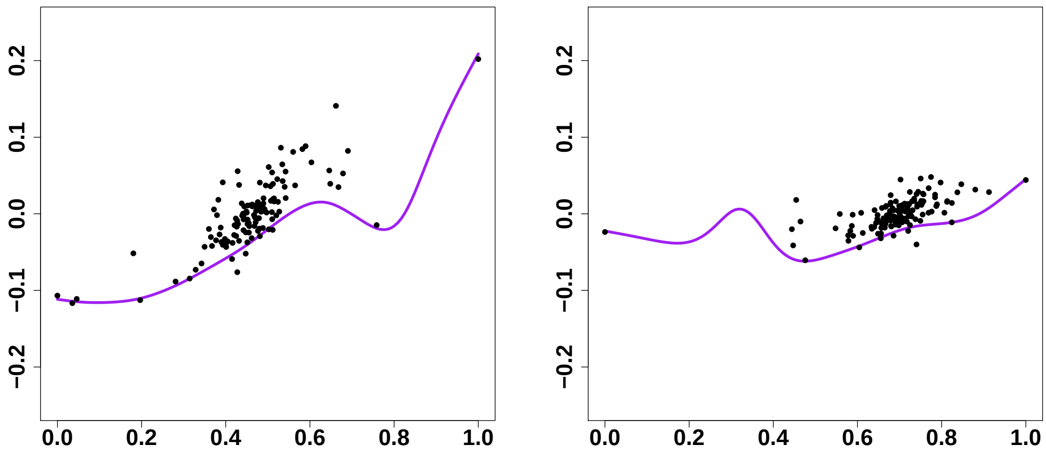

Figure 5.8: The estimated link functions in $CoVaR_{SIM}$ estimation. Where window size $n = 126$, $\tau = 0.05$, $p = 206$. For the left graph: starting date is 20081029, $h = 0.065$, $\widehat{q} = 3$: HBAN, CNO, STSA. For the right graph: starting date is 20101230, $h = 0.058$, $\widehat{q} = 2$: FBC, RDN.

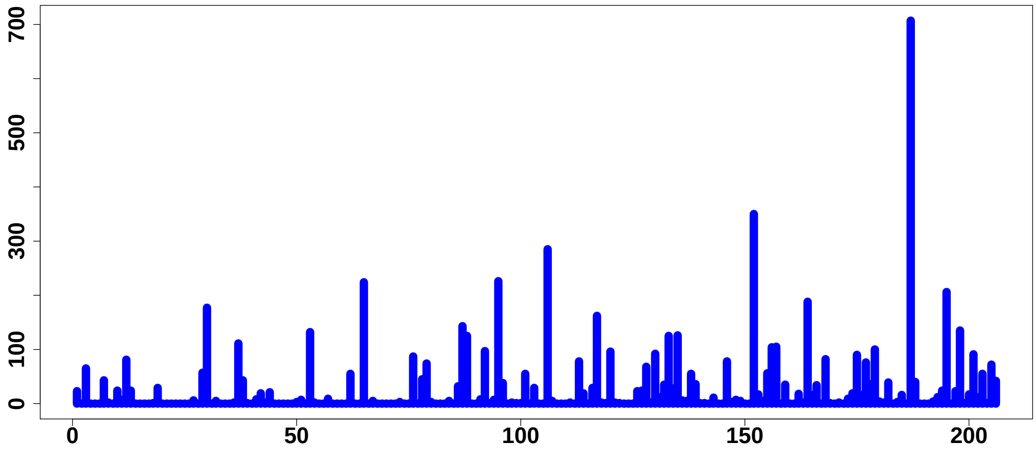

Figure 5.9: The frequency of the firms and macroprudential variables in $CoVaR_{SIM}$ estimation. The X-axis: $1-206$ variables, and the Y-axis: the frequency of the variables selected in the moving window estimation. The variable 187, i.e. "Radian Group Inc. (RDN)" is the most frequently selected variable with frequency 707.

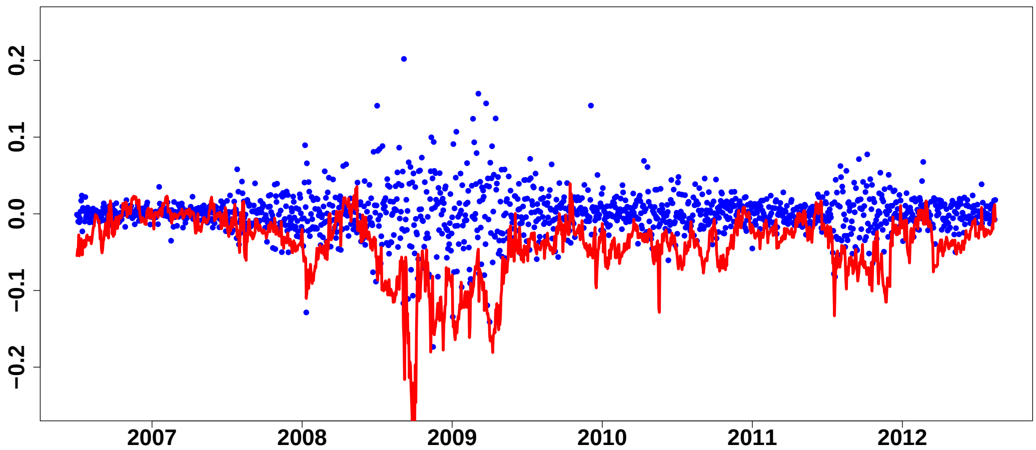

Figure 5.10: Log returns of CYN (blue) and the estimated $CoVaR_L$ (red). Where $\tau = 0.05$, $T = 1543$, window size $n = 126$, refer to (3.10).

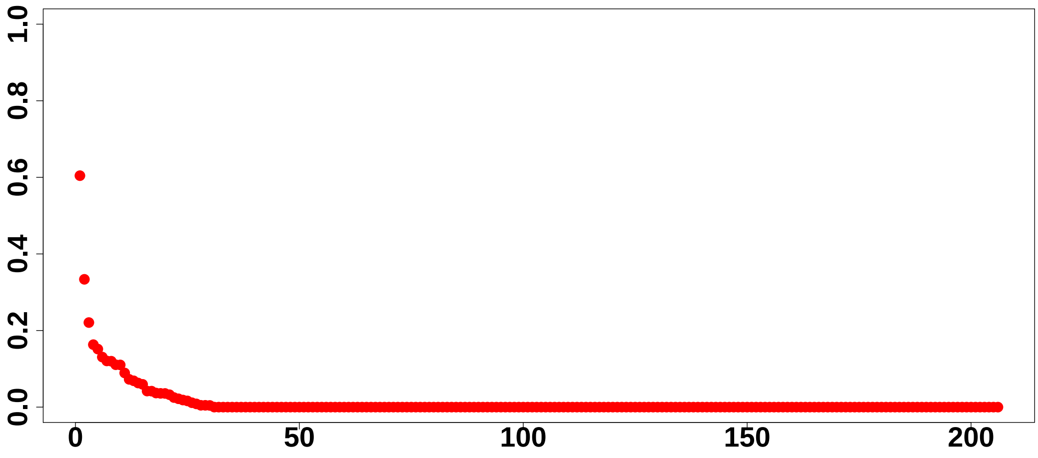

Figure 5.11: $\widehat{\beta}$ in $CoVaR_L$ estimation, 20090209-20090807.

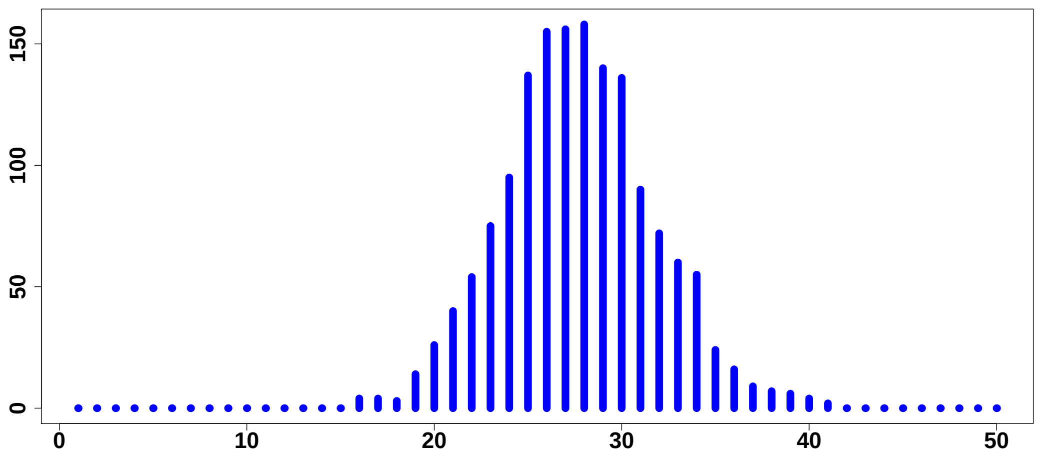

Figure 5.12: Frequency of the number of selected variables in $CoVaR_L$ estimation. Where $\widehat{q} = 28$ with frequency 158, $\widehat{q} = 27$ with frequency 156 and $\widehat{q} = 26$ with frequency 155.

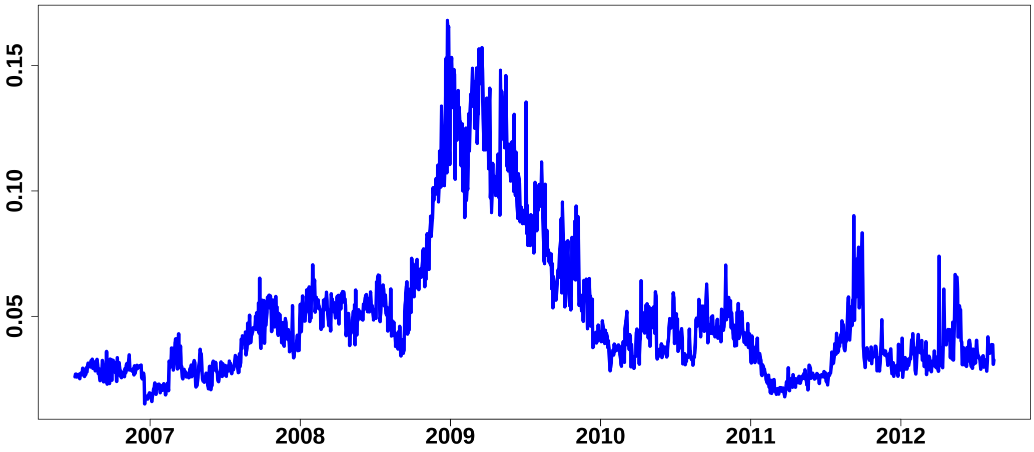

Figure 5.13: The $\widehat{\lambda}$ in $CoVaR_L$ estimation.

	p-value of Wald's test statistics
$\widehat{VaR}$	0.99
$\widehat{CoVaR}_{SIM}$	0.93
$\widehat{CoVaR}_L$	3.2×10^{-5}

Table 5.4: The CaViaR test for $\widehat{VaR}$, $\widehat{CoVaR}_{SIM}$ and $\widehat{CoVaR}_L$ for CYN, $T = 350$, $20080915 - 20100208$.

$\widehat{CoVaR}_{SIM}$ can not be rejected, therefore both $\widehat{VaR}$ and $\widehat{CoVaR}_{SIM}$ algorithms perform well during the crisis period, but $\widehat{CoVaR}_L$ performs not well.

Therefore, $\widehat{CoVaR}_{SIM}$ performs well in both crisis and overall period. From this comparison it can be concluded that $CoVaR_{SIM}$ risk measure is more precise, and $CoVaR_{SIM}$ can help people to find the most relevant influential firms.

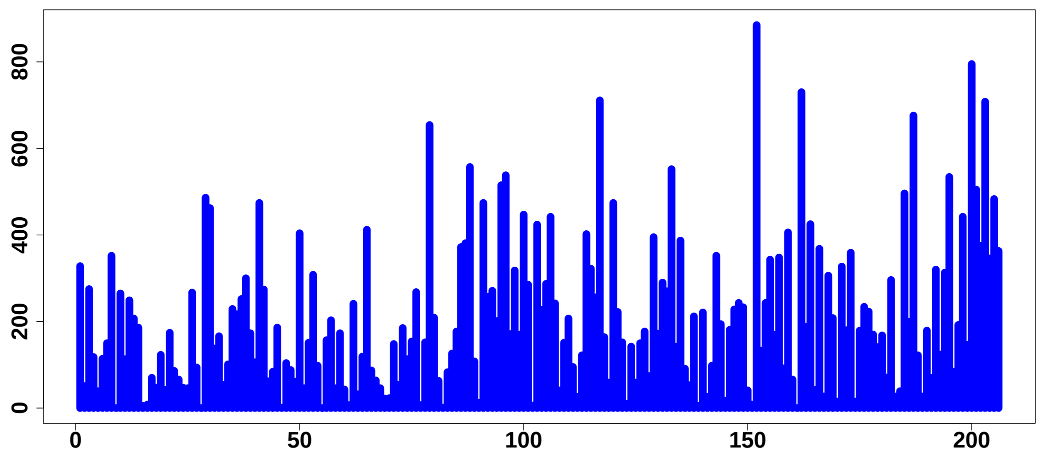

Figure 5.14: The frequency of the firms and macroprudential variables in $CoVaR_L$ estimation. The X-axis: $1 - 206$ variables, and the Y-axis: the frequency of the variables selected in the moving window estimation. The variable 152, i.e. "Flagstar Bancorp Inc. (FBC)" is the most frequently selected variable with frequency 885.

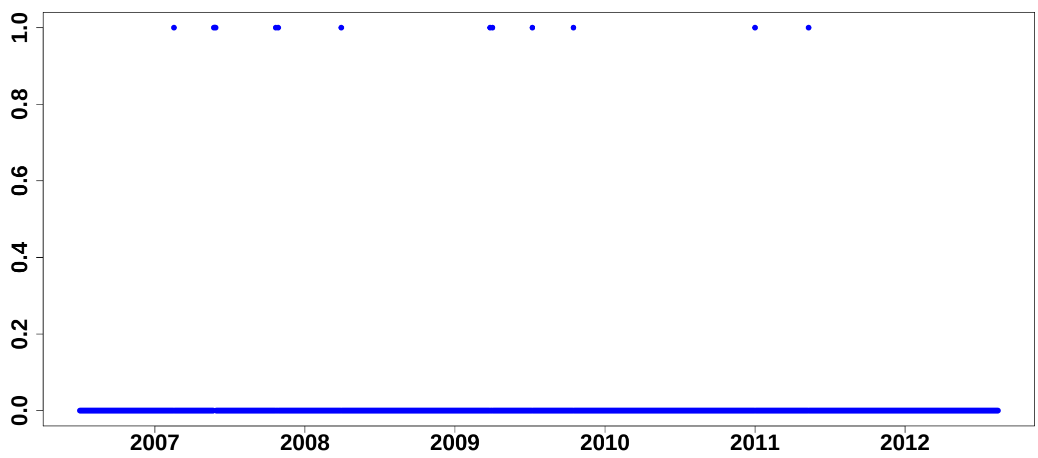

Figure 5.15: The dot plots in VaR estimation. The top dots are the violations (i.e. $\{t : I_{i,t} = 1\}$) of $\widehat{VaR}$ of CYN, totally 14 violations, $\widehat{\tau} = 0.009$, $T = 1543$.

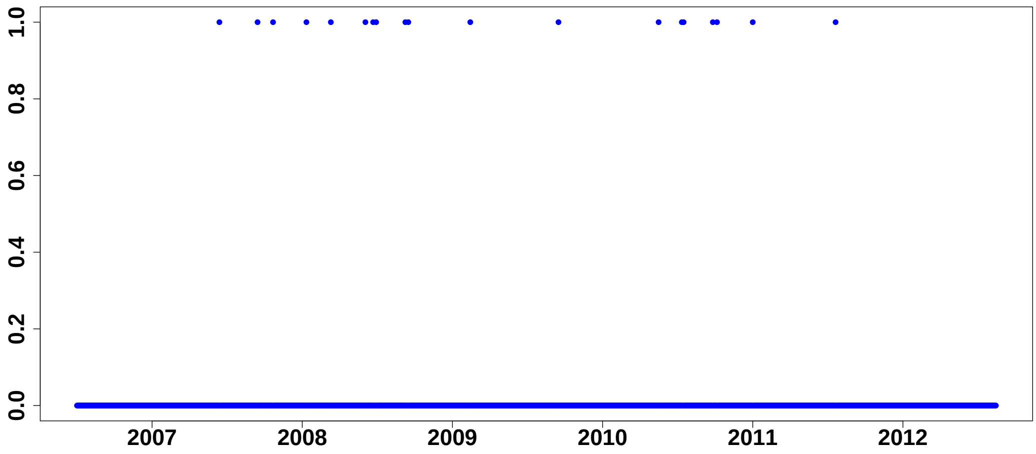

Figure 5.16: The dot plots in $CoVaR_{SIM}$ estimation. The top dots are the violations (i.e. $\{t : I_{i,t} = 1\}$) of $\widehat{CoVaR}_{SIM}$ of CYN, totally 19 violations, $\widehat{\tau} = 0.012$, $T = 1543$.

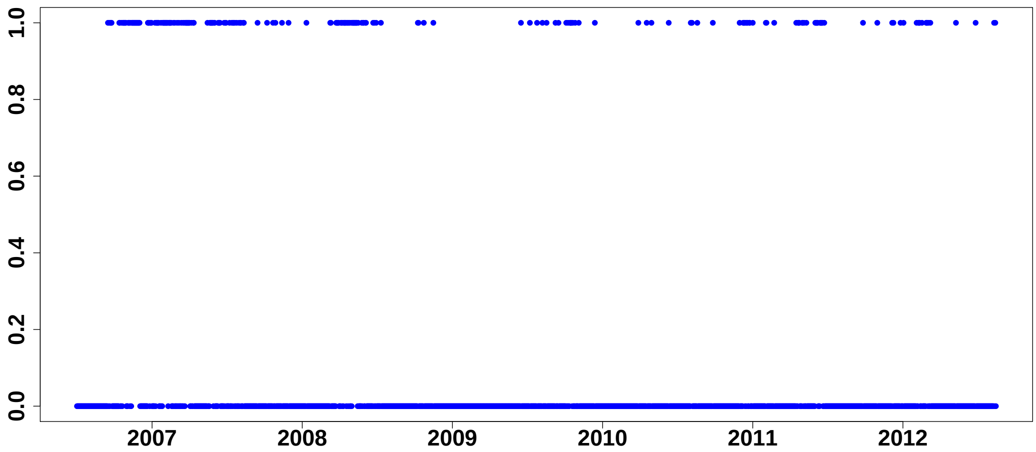

Figure 5.17: The dot plots in $CoVaR_L$ estimation. The top dots are the violations (i.e. $\{t : I_{i,t} = 1\}$) of $\widehat{CoVaR}_L$ of CYN, totally 231 violations, $\widehat{\tau} = 0.15$, $T = 1543$.

6 Conclusion

In this paper a quantile lasso regression for the single index models is introduced. And the methodology is implemented in terms of financial time series to estimate CoVaR of one specified firm, then two different methods are compared: quantile lasso regression for single index model and linear quantile lasso regression model. The evaluation is conducted by backtesting. From the result of backtesting, it can be concluded that $CoVaR_{SIM}$ risk measure is more precise, the most relevant influential firms can be found by applying this method.

List of Figures

List of Tables

Bibliography

[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] Fan, Y., Härdle, W. K., Wang, W. and Zhu, L. (2013). Composite Quntile Regression For the Single-Index Model. *SFB 649 Discussion Papers 2013-010*

[8] Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84**: 986-995.

[9] Hautsch, N., Schaumburg, J. and Schienle, M. (2011) Financial

Network Systemic Risk Contributions. *SFB 649 Discussion Papers 2011-072*

[10] Huber, P. J. (1985). Projection pursuit. *Ann. Math. Statist.* **13**: 435-475.

[11] 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.

[12] 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.

[13] Koenker, R. and Bassett, G. W. (1978). Regression quantiles. *Econometrica* **46**: 33-50.

[14] Koenker, R. and Bassett, G. W. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50**: 43-61.

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

[16] Kong, E. and Xia, Y. (1994). Variable selection for the single-index model. *Biometrika* **94**: 217-229.

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

[18] Li, Y. and Zhu, J. (2008). L1- norm quantile regression. *J. Comput. Graph. Stat.* **17**: 163-185.

[19] Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55**: 819-847.

[20] 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.

[21] Schnabel, S. and Eilers, P. (2009). Optimal expectile smoothing. *Comput. Stat. Data. An.* **53** (12): 4168-4177.

[22] Serfling, R. J. (2001). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

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

[24] 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.

[25] Wu, T. Z., Yu, K. and Yu, Y. (2010). Single-index quantile regression. *J. Multivariate Anal.* **101**: 1607-1621.

[26] 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.

[27] Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93**: 228-237.

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

[29] 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.

[30] Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101**: 476.

[31] Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.* **36** (3): 1108-1126.

Appendix

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 6.1: 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. Prosperity Bancshares, Inc. (PB)	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 6.2: The financial firms

The financial frims	
119. The Hanover Insurance Group Inc. (THG)	134. BancorpSouth, Inc. (BXS)
120. Pacific Capital Bancorp (PCBC)	135. Privatebancorp Inc. (PVTB)
121. F.N.B. Corporation (FNB)	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. MarketAxess Holdings, Inc. (MKTX)
129. Nelnet, Inc. (NNI)	144. Euronet Worldwide Inc. (EEFT)
130. Wintrust Financial Corporation (WTFC)	145. Community Bank System Inc. (CBU)
131. Umpqua Holdings Corporation (UMPQ)	146. CVB Financial Corp. (CVBF)
132. GAMCO Investors, Inc. (GBL)	147. MB Financial Inc. (MBFI)
133. Sterling Financial Corp. (STSA)	148. ABM Industries Incorporated (ABM)
149. Glacier Bancorp Inc. (GBCI)	164. Citizens Republic Bancorp, Inc (CRBC)
150. Selective Insurance Group Inc. (SIGI)	165. Horace Mann Educators Corp. (HMN)
151. Park National Corp. (PRK)	166. DFC Global Corp. (DLLR)
152. Flagstar Bancorp Inc. (FBC)	167. Navigators Group Inc. (NAVG)
153. Home BancShares, Inc. (HOMB)	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 6.3: The financial firms

Top 6 influential covariates	Frequency
No. 187 Radian Group Inc. (RDN)	707
No. 152 Flagstar Bancorp Inc. (FBC)	350
No. 106 Synovus Financial Corporation (SNV)	285
No. 95 CNO Financial Group, Inc. (CNO)	226
No. 65 Zions Bancorp. (ZION)	224
No. 195 Epoch Investment Partners, Inc. (EPHC)	206

Table 6.4: The selected risk factors ($CoVaR_{SIM}$)

Top 6 influential covariates	Frequency
No. 152 Flagstar Bancorp Inc. (FBC)	885
No. 200 VIX	795
No. 162 WisdomTree Investments, Inc. (WETF)	730
No. 117 MBIA Inc. (MBI)	711
No. 203 Change in the slope of the yield curve (yield)	703
No. 187 Radian Group Inc. (RDN)	676

Table 6.5: The selected risk factors ($CoVaR_L$)

Declaration of Authorship

I hereby confirm that I have authored this master thesis independently and without use of others than the indicated sources. Where I have consulted the published work of others, in any form (e.g. ideas, equations, figures, text, tables),this is always explicitly attributed.

Berlin, August 2, 2014

Lining Yu

Erklärung zur Urheberschaft

Hiermit erklare ich, Lining Yu, dass ich die vorliegende Arbeit allein und nur unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe. Die Prüfungsordnung ist mir bekannt. Ich habe in meinem Studienfach bisher keine Masterarbeit eingereicht bzw. diese nicht endgültig nicht bestanden.

Berlin, August 2, 2014

Lining Yu