Sparse ridge Sliced inverse quantile regression without quantile crossing

Assist. Dr. Ali Jawad Kadhim Alkenani
Department of Statistics, College of Administration and Economic
Al-Qadisiyah University

Abstract

Quantile regression provides a more complete statistical analysis of the stochastic relationships between random variables. While this technique has become very popular as a comprehensive extension of the classical mean regression it nonetheless suffers the problem of crossing of regression functions estimated at different orders of quantiles. Theoretically, the extension of conditional quantiles to higher dimension *p* of X is straight forward. However, its practical success suffers from the so-called 'curse of dimensionality'. In this article we propose a method of obtaining quantile regression estimates for high dimension data without the unfavourable quality of quantile crossing. The proposed method is a two step procedure that initially employs sparse ridge sliced inverse regression (SRSIR) to achieve dimension reduction when the predictors are possibly correlated and then followed by the usage of non-parametric method to estimate non-crossing quantile regression. For the second stage of our method we employ double kernel smoothing method (Yu and Jones,1998); monotone-based smoothing method based on the convolution of the distribution (Dette and Volgushev,2008) and joint non-crossing quantile smoothing spline method (Bondell et al., 2010) for estimating the conditional quantile without quantile crossing. Through a simulation and empirical study we compare our estimators with that of Gannon et al. (2004).

Keywords: Dimension reduction, Sliced inverse regression (SIR), Conditional quantiles, Non-crossing quantiles.

1. Introduction

Quantile regression provides a more complete statistical analysis of the stochastic relationships between random variables. While this technique has become very popular as a comprehensive extension of the classical mean regression it nonetheless suffers the problem of crossing of regression functions estimated at different orders of quantiles.

Standard nonparametric quantile regression estimation suffers quantile crossing phenomenon, particularly for the neighbouring quantile curves. This problem is partly due to sparse data in an area (or reflects a paucity of data in the region concerned) and partly due to the quantile curves are usually computed one level at a time. This phenomenon usually causes difficulty in interpretation for applicants, see Koenker (2005). There are a number of researchers attempt to find solutions for this problem. He (1997) proposed RRQ (restricted regression quantiles due to quantile smoothing spline optimization problem with boundary constraints) on regression models. Yu and Jones (1998) proposed a nonparametric quantile regression 'double-kernel' method. This method smoothes the data along both the X and Y-axes using kernel smoothing followed by local-linear weighting in the X -axis direction. The authors urge that their method give non-crossing quantiles curves. Takeuchi et al (2005) added a positive derivative constrain on quantile smoothing spline optimization problem. Dette and Volgushev (2008) proposed monotone-based smoothing method based on the convolution of the distribution to give non-crossing quantiles curves. Bondell et al. (2010) proposed joint non-crossing quantile smoothing spline method for estimating the conditional quantile without quantile crossing.

The nonparametric quantile regression estimation suffers from another problem is called the 'curse of dimensionality'. Theoretically, while the extension of nonparametric conditional quantiles from univariate to higher dimension case is quite clear its practical success is impeded by the 'curse of dimensionality'. The reason for the curse of dimensionality is due to the exponential increase in dimension associated with adding extra dimensions to a space.

A generic problem in non-parametric quantile regression estimation is the sparseness of high dimensional data. Furthermore, convergence rate of non-parametric estimators depends on the inverse relationship with the dimension p of predictor space. That is, when convergence decreases the dimension increases for details see Stone (1982). Therefore, in light of the preceding discussion, the challenge is to reduce the P -dimensional predictor X without loss of information on the conditional distribution of Y given X and without requiring a pre-specified parametric model.

There are a number of approaches that attempts to reduce the P -dimensional predictor X without loss of information and then estimate the conditional quantile. Chaudhuri (1991), Gooijer and Zerom (2003), Yu and Lu (2004), Horowitzand and Lee (2005), Dette and Scheder (2011), and Yebin et al. (2011) used variants of adaptive model in order to reduce the dimension and thereafter estimate the conditional quantiles. Wu et al. (2010) proposed single-index quantile regression. They introduced a practical algorithm where the unknown link function is estimated by local linear quantile regression and the parametric index is estimated through linear quantile regression.

Gannon et al. (2004) used sliced inverse regression (SIR) to reduce the dimensionality of the covariates in order to construct a more efficient estimator of conditional quantiles. Specifically, the authors employ (SIR) method as a pre-step to avoid the curse of dimensionality and then conditional quantile estimators are obtained by inverting the conditional distribution which is estimated using the Nadaraya-Watson method. However, the estimator proposed by Gannon et al. (2004) has some disadvantages. Notably, while the (SIR) method is proven to be an effective dimension reduction tool, it suffers from the drawback of using each dimension reduction component as a linear combination of all the original predictors. Consequently, this makes the interpretation of resulting estimates very difficult. Furthermore, in the presence of collinearity among the predictors the (SIR) method is documented to be inefficient (Li and Yin ,2008). Additionally, the usage of Nadarya-Watson method to estimate the conditional distribution may induce large bias and boundary effects as is widely acknowledge in the literature.

Existing estimation procedures for nonparametric quantile regression which are proposed to solve the curse of dimensionality problem fail in solving the quantile crossing problem. For example, Single index quantile regression (Wu et al.,2010) , Local linear additive quantile regression (Yu and Lu, 2004).On the other hand, the estimation procedures which are proposed to solve the quantiles crossing problem unable to work in high dimensional data because they are proposed to work with low dimensions. So, we have proposed our method to tackle these two problems together.

In this paper, we propose a semi-parametric quantile regression estimation method for high dimension data when the predictors are possibly correlated. For different order of quantiles, the proposed approach guarantees that the estimated quantile regression functions do not cross. In order to ensure the quantile functions are not overlapping our proposed two-step procedure first employs the sparse ridge sliced inverse regression (SRSIR) to overcome the curse of dimensionality and then followed by a conditional quantile estimation methods which are give non-crossing quantile curves.

The article is organized as follows. In sections 2 and 3 we respectively give brief reviews of sparse ridge sliced inverse regression (SRSIR) followed by non-parametric quantile estimation methods focusing on the double kernel quantile regression method (Yu and Jones, 1998); monotone-based smoothing method (Dette and Volgushev, 2008) and non-crossing quantile smoothing spline method (Bondell et al. , 2010). In section 4, we present our two-step proposed estimation procedure. we conduct simulations under different settings in section 5. We report the applications of our method on real data in section 6. Finally, the conclusions are summarized in Section 7.

2. Sparse ridge sliced inverse regression

One of the most important objectives in analyzing high dimension datasets is the reduction of dimension for visualizing patterns of data structure. The theory of sufficient dimension reduction (Li, 1991; Cook, 1998) has been developed to achieve this aim. In high-dimensional, for regression problems involving a univariate response Y and a p -dimensional predictor $X = (\mathrm{x}_1, \dots, \mathrm{x}_p)$, the sufficient dimension reduction aims to replace X with a lower-dimensional projection $P_S X$ without loss of information about the conditional distribution of Y |X, where P_S is the orthogonal projection in the usual inner product. In this setting no pre-specified model for Y |X is required. The central subspace $S_{\mathrm{Y}|\mathrm{X}}$ is the important object in the dimension reduction and in light of this there are number of dimension reduction methods proposed to estimate $S_{\mathrm{Y}|\mathrm{X}}$. Sliced Inverse Regression (SIR) approach proposed by (Li, 1991) is one of these methods. The author showed that an estimate of the central subspace $S_{\mathrm{Y}|\mathrm{X}}$ can be obtained by the first d eigenvectors $\upsilon_1, \dots, \upsilon_d$ for the eigenvalue problem of the form

$$M\, \upsilon_i = \rho_i \Sigma_x\, \upsilon_i \qquad (1)$$

where $\rho_1 \geq \cdots \geq \rho_d > 0$ are the corresponding positive eigenvalues, $\Sigma_x = Cov(X)$ and

$M = Cov\{E(\mathrm{X}\,|\mathrm{Y})\}$.

SIR suffers from the fact that each dimension reduction component is a linear combination of all the original predictors and no variable selection is achieved and thus making it difficult to interpret the resulting estimates. Moreover, (SIR) method suffers when there is high collinearity among the predictors (Li and Yin ,2008). In light of these drawbacks, in this article we shall make use of sparse ridge sliced inverse (SRSIR) method to remedy these problems.

Cook (2004) rewrite SIR in (1) as a least-squares minimization problem and SIR estimate can be obtain by minimizing

$$L(B, C) = \sum_{s=1}^{a} \hat{f}_s \, \|\bar{Z}_s - BC_s\|^2 \qquad (2)$$

Where $\hat{Z} = \hat{\Sigma}^{-\frac{1}{2}} \left(X - E(X)\right)$ with $\bar{Z}_s$ denoting the mean of $\hat{Z}$ in the sth slice, n_s is the number of observations within each slice and $\hat{f}_s = n_s/n$ is the observed fraction of observations falling in slice y. Over $B \in \mathbb{R}^{p\times d}$ and $C = (C_1, \dots., C_a) \in \mathbb{R}^{p\times a}$ the values of B which minimize $L(B, C)$ form an estimation of the central space $S_{Y\,|X}$.

Li and Yin (2008) derived another least-squares formulation of SIR, which is equivalent to (2) but in the original predictor X scale as follows:

$$\tilde{L}(B, C) = \sum_{s=1}^{a} \hat{f}_s \left\|(\bar{X}_s - \bar{X}) - \hat{\Sigma}_x\, BC_s\right\|^2 \qquad (3)$$

and then they proposed ridge sliced inverse regression (RSIR) estimator given by:

$$L_\eta(B, C) = \sum_{s=1}^{a} \hat{f}_s \left\|(\bar{X}_s - \bar{X}) - \hat{\Sigma}_x\, BC_s\right\|^2 + \eta\, vec(B)^T vec(B) \qquad (4)$$

Where η is a nonnegative constant and $vec(.)$ is a matrix operator that stacks all columns of the matrix to a single vector. For a fixed η , Li and Yin (2008) proposed an alternating least –squares algorithm to minimize (4). The mechanics of the algorithm is as follows:

Given B the solution of C can be obtained by:

$$\hat{C} = \left(\hat{C}_1, \dots, \hat{C}_a\right) \text{ where } \quad \hat{C}_s = \left(B^T \hat{\Sigma}_x^2 B\right)^{-1} B^T \hat{\Sigma}_x \,(\bar{X}_s - \bar{X}), \qquad (5)$$

$s = 1, \dots \dots, a.$

Thereafter, rewrite (4) in the form of least-squares regression,

$$L_\eta(B, C) = \left\|\tilde{A}^{1/2}\, \tilde{Y} - \tilde{A}^{1/2}\left(C^T \otimes \hat{\Sigma}_x\right) vec(B)\right\|^2 + \eta\, vec(B)^T vec(B) \qquad (6)$$

where $\otimes$ is the Kronecker product, $\tilde{Y} = vec(\bar{X}_1 - \bar{X}, \dots, \bar{X}_a - \bar{X})$, $\tilde{A}^{1/2} = D_f^{1/2} \otimes I_P$, and $D_f = diag\left(\hat{f}_1, \dots, \hat{f}_a\right)$. Given C, the solution of B in (6) is

$$vec\left(\hat{B}\right) = \left(CD_f C^T \otimes \hat{\Sigma}_x^2 + \eta\, I_{pd}\right)^{-1}\left(CD_f \otimes \hat{\Sigma}_x\right)\tilde{Y} \qquad (7)$$

and this procedure will continue between minimizing B and C until convergence.

Li and Yin (2008) derived a generalized cross-validation criterion (GCV) to select the ridge parameter η in (4) given by

$$GCV = \frac{\left\|\left(I_{pa} - \Lambda_\eta\right)\tilde{A}^{1/2}\,\tilde{Y}\right\|^2}{pa\{1 - trace\left(\Lambda_\eta\right)/pa\}^2}, \tag{8}$$

where,

$$\Lambda_\eta = \left(D_f^{1/2}\hat{C}^T \otimes \hat{\Sigma}_x\right)\left(\hat{C}D_f\hat{C}^T \otimes \hat{\Sigma}_x^2 + \eta\, I_{pd}\right)^{-1}\left(\hat{C}D_f^{1/2} \otimes \hat{\Sigma}_x\right) \tag{9}$$

Estimates of ridge SIR (RSIR) are linear combinations of all the predictors, and no variable selection is achieved. Li and Yin (2008) followed the least absolute shrinkage and selection operator (Lasso) idea to the RSIR estimator to induce sparsity in the estimated linear combinations.

Let $\left(\hat{B}, \hat{C}\right)$ denote the RSIR estimator. A sparse ridge sliced inverse regression (SRSIR) estimator of the central subspace $S_{Y|X}$ is defined as $Span\left(diag(\hat{\alpha})\hat{B}\right)$, where the shrinkage index vector $\hat{\alpha} = \left(\hat{\alpha}_1, \dots, \hat{\alpha}_p\right) \in \mathbb{R}^p$ is obtained by minimizing

$$L_\lambda(\alpha) = \sum_{s=1}^{a} \hat{f}_s \left\|(\bar{X}_s - \bar{X}) - \hat{\Sigma}_x\, diag(\alpha)\hat{B}\hat{C}_s\right\|^2 \tag{10}$$

over α, subject to $\sum_{j=1}^{p}\left|\alpha_j\right| \le \lambda$, for some non-negative constant λ. Because $g(\alpha)\hat{B}\hat{C}_s = diag\left(\hat{B}\hat{C}_s\right)\alpha$, we have

$$L_\lambda(\alpha) = \sum_{s=1}^{a} \hat{f}_s \left\|(\bar{X}_s - \bar{X}) - \hat{\Sigma}_x\, diag\left(\hat{B}\hat{C}_s\right)\alpha\right\|^2 \tag{11}$$

Let $\tilde{Y} = vec(\bar{X}_1 - \bar{X}, \dots, \bar{X}_a - \bar{X}) \in \mathbb{R}^{pa}$, $\tilde{X} = \left(diag\left(\hat{B}\hat{C}_1\right)\hat{\Sigma}_x, \dots, diag\left(\hat{B}\hat{C}_a\right)\hat{\Sigma}_x\right)^T \in \mathbb{R}^{pa\times p}$

Then the shrinkage vector α, is exactly the Lasso estimator for the regression $\tilde{Y}$ with pa observations on the p-dimensional data matrix $\tilde{X}$. Akaike information criterion (AIC), Baysian information criterion(BIC), and residual information criterion (RIC) can be used to select λ.

Li and Yin (2008) adopted a criterion proposed by Zhu et al. (2006) to estimate $d = \dim(S_{Y|X})$. Zhu et al.(2006) suggested that d can be estimated by:

$$\hat{d} = \arg \max_{0 \le m \le p-1} \left\{ \frac{n}{2} \sum_{i=1+min(k\varkappa,m)}^{p} (log(\hat{\gamma}_i) + 1 - \hat{\gamma}_i) - \frac{Л_n\, m(2p-m+1)}{2}, \right. \quad (12)$$

where, the matrix Ƕ$\Gamma = Cov\big(E(X/Y)\big) + I_p$ and $\hat{\gamma}_1, \dots, \hat{\gamma}_p$ denote the eigenvalues of the sample estimate $\hat{\Gamma}$ of Γ , $\varkappa$ is the number of $\hat{\gamma}_i > 1$, and $Л_n$ is a penalty constant taken to be $Л_n = (log(n)a/n)$.

3.Non-crossing nonparametric quantile regression methods

First of all, we need to explain a class of kernel smoothing quantile regression. A class of kernel smoothing quantile regression is defined as a solution to

$$\min_{\vartheta} \sum \rho_{\tau(yi-\vartheta)}\, K_h(x_i - x) \qquad (13)$$

Where $K_h(.) = \frac{1}{h} K\left(\frac{.}{h}\right)$ is a kernel function with bandwidth h, and ρ_τ is the check function given by:

$$\rho_\tau(u) = \tau\, u\, I_{[0,\infty)}(u) - (1-p)\, u\, I_{(-\infty,0)}\,(u) \qquad (14)$$

with τ indexes the conditional quantile of interest.

Another class of kernel quantile regression estimation is based on estimating conditional distribution function or conditional density function. For instance, the conditional $\tau\, th$ quantile $q_\tau(x)$ is the solution of

$$F(q_\tau(x)/x) = \tau \qquad (15)$$

Where the distribution function $F(y/x)$ is estimated by

$$\hat{F}\,(y/x) = \textstyle\sum_i w_i\,(x)\;\; I(Y_i < y) \qquad (16)$$

Then estimating the conditional quantile function by the inverse of $\hat{F}\,(y/x)$ where the weights $w_i(x)$ are either obtained using the Nadarya-Watson procedure

$$w_i(x) = \frac{K_{h(x_i - x)}}{\sum_{i=1}^{n} K_h\,(x_i - x)} \qquad i = 1, \dots, n \qquad (17)$$

or the local linear approach

$$w_i(x) = K\left(\frac{x - X_i}{h}\right)[1 - \hat{B}_x(x - X_i)] \qquad (18)$$

where $\hat{B}_x = \frac{S_{n,1}}{S_{n,2}}$,

$$S_{n,l} = \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)(x-X_i)^l \ , l = 1,2 \qquad (19)$$

with kernel function K and bandwidth h , and $I\left(Y_j < y\right)$ is indicator function.

Now, we will give view for the methods which we used to get non-crossing quantile regression. The first and second of the following methods are from nonparametric kernel smoothing quantile regression while the third method from nonparametric quantile regression smoothing spline.

3.1 Double kernel method

Yu and Jones (1998) proposed a nonparametric quantile regression 'double-kernel' method. This method smoothes the data along both the X and Y-axes using kernel smoothing followed by local-linear weighting in the X -axis direction. The double-kernel method requires two bandwidths, one for kernel smoothing along the X-axis and the other for kernel smoothing along the Y-axes. Yu and Jones (1998) outlined an automated method for determining the two bandwidths for each percentile curve (Li et al. , 2010). The idea behind this method is that replacing the indicator function $I(Y_i < y)$ in (16) by a kernel density function to get a smoother version of the estimated conditional distribution function in (16) as

$$\hat{F}_{DKLL}(y/x) = \frac{1}{w_i(x,h_1)} \sum_i w_i(x,h_1)\, \Omega\left(\frac{\hat{q}_\tau(x)-Y_i}{h_2}\right), \qquad (20)$$

where Ω is the distribution function associated with a kernel density function W and

$$\Omega\left(\frac{y-Y_i}{h_2}\right) = \int_{-\infty}^{y} W_{h_2}\,(Y_i - u)du \qquad (21)$$

$w_i(x,h_1)$ are the local linear weights .

This method focuses on choosing a second kernel function W and second bandwidth h_2 so that $\hat{q}_\tau(x)$ is monotonic function of τ for all x in its range to solve the problem of quantile crossing. If W is taken to be the uniform kernel density, $W(u) = 0.5(|u| \le 1)$ it can show that $\frac{d}{d\tau}\hat{q}_\tau(x) > 0$.

The choice of the bandwidth is critical in determining how smooth the resulting conditional mean/percentile curve will be. Yu and Jones (1998) proposed an automatic bandwidth selection strategy for estimating conditional percentiles using single and double-kernel methods. They summarise their automatic bandwidth selection strategy for smoothing conditional quantiles as follows:

a. Use ready methods to select h_{mean} ,where h_{mean} is the bandwidth for the smoothing estimation of the mean regression.

b. Use $$h_{1,\tau} = h_{mean}\left\{\tau(1-\tau)/\emptyset\left(\Phi^{-1}(\tau)\right)^2\right\}^{1/5}$$
(22)

to obtain all $h_\tau s$ from h_{mean}, where $h_{1,\tau}$ is the first bandwidth for the percentage and

$\emptyset$ and Φ^{-1} are the standard normal density and distribution functions, respectively.

Additionally, the authors proposed a method that depends only on τ and $h_{1,\tau}$ for obtaining $h_{2,\tau}$, the second bandwidth estimator for smoothing in the Y direction. This bandwidth is given by $h_{2,\tau} =$
$max\left(\frac{h^5{}_{1,1/2}}{h^3{}_{1,\tau}}, \frac{h_{1,\tau}}{10}\right)$ if $h_{1,1/2} < 1$ (23)

$= \frac{h^4{}_{1,1/2}}{h^3{}_{1,\tau}}$ Otherwise

Yu and Jones (1998) preferred the double-kernel to the single-kernel estimator because of its smoother appearance, improved mean-squared-error properties (Li et al., 2010) and its yielding of non-crossing quantiles.

3.2 Monotone-based smoothing

Dette and Volgushev (2008) proposed Monotone-based smoothing method to give non-crossing quantiles curves. The idea of this method is based on convolution of the distribution. Let G denote a strictly increasing (fix or known) distribution function, consider the convolution $H^{-1}(\tau) = \int I\{F(y) \leq \tau\}\, dG(y)$

$$= \int_0^1 I[F\{G^{-1}\} \leq \tau]dv \qquad (24)$$

or $H^{-1}(\tau) = G_0F^{-1}(\tau)$ *or* $F^{-1}(\tau) = G^{-1}{}_0H^{-1}(\tau)$

The function $\tau \rightarrow H^{-1}(\tau)$ is obviously always increasing but the function H is not smooth, so first smooth $H^{-1}(\tau)$ as

$$H_{h_d}^{-1}(\tau) = \frac{1}{h_d}\int_{-\infty}^{\tau}\int_0^1 K_d\left[\frac{F\{G^{-1}(v) - u\}}{h_d}\right] dvdu \qquad (25)$$

$\rightarrow_{h\rightarrow 0} H_h^{-1}(\tau)$, then estimate $F^{-1}(\tau)$ as $F_{x,G}^{\widehat{-1}}(\tau) = \left(G^{-1}{}_0 H_x^{\widehat{-1}}\right)(\tau)$ where function G is introduced here to enforce the existence of the integral if the support of the distribution function is unbounded and h_d denotes a bandwidth converging to 0.

For computational reasons the integration with respect to the variable dv is substituted by a summation and the function F is replaced by an appropriate estimate of the conditional distribution, say $\hat{F}_x$:

$$H_x^{\widehat{-1}}(\tau) = \frac{1}{Nh_d}\sum_{i=1}^{N}\int_{-\infty}^{\tau} K\left[\frac{\hat{F}_x\{G^{-1}(i/N)\} - u}{h_d}\right] du, \qquad (26)$$

where,
$$h_d = min\left(\tau, 1-\tau, h_{1,\tau}^{1.3}\right) \qquad (27)$$

3.3 The joint constrained quantile smoothing spline

Koenker et al. (1994) explored a class of quantile smoothing splines, defines as a solution to

$$\min_{g} \sum \rho_\tau\left(y_i - g(x_i)\right) + \lambda\, V(g') \ldots \ldots (28)$$

Where $V(g')$ is the total variation of the derivative of the function g and for twice continuously differentiable g $\quad V(g') = \int_0^1 |g''(x)| dx.$

Koenker et al. (1994) , He and Ng(1999) suggested the use of a Schwartz-type information criterion for choosing the regularization parameter in quantile smoothing splines. For each individual quantile curves, this criterion is

$$SIC\left(\lambda_{\tau_t}\right) = log\left[n^{-1}\sum_{i=1}^{n}\rho_{\tau_t}\left\{y_i - g_{\tau_t}^{\lambda_{\tau_t}}(x_i)\right\}\right] + (2n)^{-1} P_\lambda \;\; log(n) \quad (29)$$

Where $x \in [0,1]$, $t = 2, \ldots, q$, $g_\tau^{\lambda_\tau}$ denotes the estimated function for that choice of λ_τ and P_λ is the number of interpolated data points.

Bondell et al.(2010) proposed the constrained joint quantile smoothing spline as the set of functions $\hat{g}_{\tau 1}, \ldots, \hat{g}_{\tau q} \in \boldsymbol{g}$ that minimize

$$\sum_{t=1}^{q} w(\tau_t)\sum_{i=1}^{n}\rho_{\tau_t}\{y_i - g_{\tau_t}(x_i)\} + \sum_{t=1}^{q}\lambda_{\tau_t} V\left(g'_{\tau_t}\right) \qquad (30)$$

Subject to $g_{\tau_t}(x) \geq g_{\tau_{t-1}}(x)$ and where $w(\tau_t)$ denote some weight function such that $w(\tau_t) > 0$ for all $t = 2, \ldots, q$.

The full set of tuning parameters for the individual quantile curves minimizes the joint Schwartz-type criterion

$$SIC_J(\lambda) = \sum_{t=1}^{q} w_*(\tau_t)\, log\left[n^{-1}\sum_{i=1}^{n} \rho_\tau\left\{y_i - g_{\tau_t}^{\lambda_{\tau_t}}(x_i)\right\}\right]$$
$$+ (2n)^{-1} P_\lambda \;\; log(n) \sum_{t=1}^{q} P_{\lambda_{\tau_t}} \quad (31)$$

with the weights satisfying $w_*(\tau_t) > 0$.

Bondell et al.(2010) show that their method give non-crossing quantile curves.

4. Estimation procedure

In this section we describe the practical implementation of our proposed two step estimation procedure. The first step is the dimension reduction using SRSIR followed by the employment of non-crossing quantile regression methods for estimating the conditional quantiles.

4.1 SRSIR step

Let $\bar{X}$ and $\hat{\Sigma}$ be the sample variance matrix of the X_i's. Follow the algorithm described in section 2, according to equations (5), (6), (7) we will obtain the SRSIR estimates for effective dimension reduction (EDR) directions sample version $\hat{b}_i$.

4.2 Non-crossing quantile regression estimator step

Using the SRSIR estimates obtained in the previous subsection, we now give non-crossing quantile regression estimators for conditional quantile according to the methods which are described in section (3). For the sake of convenience, let $d = \dim\left(S_{\mathrm{Y}\,|\mathrm{X}}\right) = 1$. Now:

$$q_\tau(\mathrm{X}) = q_\tau(\mathrm{y}/\beta^{\mathrm{T}}\,\mathrm{X}) = q_\tau(\mathrm{y}/\mathrm{b}^{\mathrm{T}}\,\mathrm{X}) \qquad (32)$$

Let $\hat{b} = \hat{b}_1$, $\hat{b}$ is an estimated basis of $span(\beta)$, then $\left\{\hat{v}_i = \hat{b}^T \mathrm{X}_i\right\}_{i=1}^{n}$ and $\hat{v} = \hat{b}^T \mathrm{X}$.

From the data $\{(y_i, \hat{v}_i)\}_{i=1}^{n}$ we can define the non-crossing quantile regression estimators for the conditional quantile according to the methods outlined in section (3).

5. Simulation study

We now provide a simulation study to illustrate the numerical performance of the proposed method. We compare our method with

Gannon et al. (2004) method as well as comparing all estimators when SIR and SRSIR are used in the first step.

5.1 Estimation methods

In our simulation study we compare between these estimators of the $\tau\ th$ −conditional quantile:

a. $\hat{q}_\tau(\hat{v}) = \hat{q}_\tau(\hat{b}^T\mathrm{X})$, this estimator is the double kernel quantile regression estimator by numerically inversing the estimated conditional cumulative distribution function in (20). The direction $\hat{b}$ is estimated with SRSIR according to equations (5), (6) and (7). The standard normal $N(0,1)$ is used as the kernel density function. The first and second bandwidths are chosen according to the equations (22) and (23) respectively.

b. $\tilde{\overset{\frown}{q}}_\tau(\hat{v}) = \tilde{\overset{\frown}{q}}_\tau(\hat{b}^T\mathrm{X})$, this estimator is similar to estimator $\hat{q}_\tau(\hat{v})$ except the dimension reduction direction $\hat{b}$ is estimated using SIR according to equation (3).

c. $\check{q}_\tau(\hat{v}) = \check{q}_\tau(\hat{b}^T\mathrm{X})$, this estimator is the monotone-based quantile regression estimator as in the (26). The direction $\hat{b}$ is estimated with SRSIR according to equation (5), (6) and (7). The kernel density function is the density of standard normal $N(0,1)$ with the bandwidth chosen according to equation (27).

d- $\dddot{q}_\tau(\hat{v}) = \dddot{q}_\tau(\hat{b}^T\mathrm{X})$, this estimator is similar to the estimator $\check{q}_\tau(\hat{v})$ except the dimension reduction direction $\hat{b}$ is estimated using SIR according to equation (3).

e. $\breve{q}_\tau(\hat{v}) = \breve{q}_\tau(\hat{b}^T\mathrm{X})$, the joint constrained quantile smoothing spline estimator as in equation (30) and the full set of tuning parameters for the individual quantile curves selected to minimizes the joint Schwartz-type criterion according to (31). The direction $\hat{b}$ is estimated with SRSIR according to equation (5), (6) and (7).

f. $\ddot{q}_\tau(\hat{v}) = \ddot{q}_\tau(\hat{b}^T\mathrm{X})$, this estimator is similar to the estimator $\breve{q}_\tau(\hat{v})$ except the dimension reduction direction $\hat{b}$ is estimated using SIR according to equation (3).

g. $\tilde{q}_\tau(\hat{v}) = \tilde{q}_\tau(\hat{b}^T\mathrm{X})$, Gannoun et al. (2004) estimator , the direction $\hat{b}$ is estimated with SIR according to the equation (3) , and the conditional quantile by numerically inverting the estimated conditional c.d.f. in (16) with weights as in (17). The kernel density function is the density of standard normal $N(0,1)$, and the bandwidth is chosen by a cross-validation technique.

5.2 Simulated model

We tested the performance of our proposed estimators and Gannoun et al. (2004) estimator for correlated predictor, by considering the following model:

$$Y = 2(\beta^T X) + (\beta^T X)^2 + \varepsilon \qquad (33)$$

Where, $X = (x_1, \dots, x_{20})$ follows a multivariate normal distribution with mean 0, and the correlation between x_i and x_j is $\rho^{|i-j|}$, we take $\rho = (0.3, 0.5, 0.7)$. Additionally, we assume the error term $\varepsilon \sim N(0,1)$ and ε independent of X. We study the estimators in these cases:

Case (I) $\beta = (1,1,0.1,0.1,0, \dots, 0)$ *and* $\rho = (0.3)$.
Case (II) $\beta = (1,1,0.1,0.1,0, \dots, 0)$ *and* $\rho = (0.5)$.
Case (III) $\beta = (1,1,0.1,0.1,0, \dots, 0)$ *and* $\rho = (0.7)$.
Case (IV) $\beta = (1,1,1,1,0, \dots, 0)/\sqrt{3}$ *and* $\rho = (0.3)$.
Case (V) $\beta = (1,1,1,1,0, \dots, 0)/\sqrt{3}$ *and* $\rho = (0.5)$.
Case (VI) $\beta = (1,1,1,1,0, \dots, 0)/\sqrt{3}$ *and* $\rho = (0.7)$.

The true conditional $\tau\, th$-quantile for the model in (33) is $q_\tau(X) = f(\beta^T X) + N_\tau$, where N_τ is the $\tau\, th$-quantile of standard normal distribution.

For the purpose of comparison the estimators $\hat{q}_p(\hat{v}), \tilde{\breve{q}}_p(\hat{v}), \check{q}_p(\hat{v}), \ddot{q}_p(\hat{v}), \breve{q}_p(\hat{v}), \ddot{q}_p(\hat{v})$, and $\tilde{q}_p(\hat{v})$ to the true quantile for the previous different cases we generate data replications $R = 200$ of sample size n=300. Five quantile values $\tau = 0.1, 0.25, 0.5, 0.75, \mathit{and}\ 0.9$ were considered. The performance of the estimators can be assessed on each of the 200 replications by a mean square error criterion (and averaged over replications).

5.3 Results

All the experiments concerning the behavior of the estimators with respect to the different forms of β and different values of ρ according to the previous six cases have been conducted based on model (33) and the Table 1 and Figure 1 provide the mean squares errors for these estimators.

Table1. Mean squared error of the estimators $\hat{q}_p(\hat{v}), \overset{\frown}{\tilde{q}}_p(\hat{v}), \check{q}_p(\hat{v}), \dddot{q}_p(\hat{v}), \breve{q}_p(\hat{v}), \ddot{q}_p(\hat{v})$, and $\tilde{q}_p(\hat{v})$ for the different quantiles $\tau = (0.10,0.25,0.50,0.75,0.90)$.

			Case (I)		
	$p = 0.1$	$p = 0.25$	$p = 0.50$	$p = 0.75$	$p = 0.90$
$\tilde{q}_p(\hat{v})$	0.099409	0.071722	0.040051	0.089786	0.097308
$\overset{\frown}{\tilde{q}}_p(\hat{v})$	0.092989	0.049379	0.024830	0.077639	0.095013
$\dddot{q}_p(\hat{v})$	0.097168	0.069118	0.034219	0.060412	0.069323
$\ddot{q}_p(\hat{v})$	0.061525	0.044617	0.030515	0.064587	0.080406
$\hat{q}_p(\hat{v})$	**0.056112**	0.026635	**0.007754**	0.043927	**0.053793**
$\check{q}_p(\hat{v})$	0.062205	**0.018083**	0.013730	**0.043039**	0.055438
$\breve{q}_p(\hat{v})$	**0.05647**	**0.013983**	**0.005788**	**0.028570**	**0.045059**
			Case (II)		
	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\tilde{q}_p(\hat{v})$	0.106747	0.078177	0.043856	0.098405	0.109350
$\overset{\frown}{\tilde{q}}_p(\hat{v})$	0.106472	0.053823	0.029051	0.085092	0.107365
$\dddot{q}_p(\hat{v})$	0.105913	0.066212	0.037470	0.075339	0.079721
$\ddot{q}_p(\hat{v})$	0.071984	0.048633	0.033414	0.070787	0.088205
$\hat{q}_p(\hat{v})$	**0.061723**	0.029032	**0.008491**	0.048144	**0.059011**
$\check{q}_p(\hat{v})$	0.068426	**0.019710**	0.015034	**0.047171**	0.060815
$\breve{q}_p(\hat{v})$	**0.062117**	**0.015241**	**0.007467**	**0.031313**	**0.049430**
			Case (III)		
	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\tilde{q}_p(\hat{v})$	0.116770	0.085349	0.047621	0.106576	0.117899
$\overset{\frown}{\tilde{q}}_p(\hat{v})$	0.111587	0.058761	0.029523	0.092157	0.112685
$\dddot{q}_p(\hat{v})$	0.116602	0.071890	0.040686	0.082043	0.089427
$\ddot{q}_p(\hat{v})$	0.081213	0.053094	0.036282	0.076665	0.095362
$\hat{q}_p(\hat{v})$	**0.067334**	0.031696	**0.009220**	0.052141	**0.063798**
$\check{q}_p(\hat{v})$	0.074646	**0.021519**	0.016325	**0.051087**	0.065749
$\breve{q}_p(\hat{v})$	**0.067764**	**0.018877**	**0.007756**	**0.033913**	**0.059928**
			Case (IV)		
	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\tilde{q}_p(\hat{v})$	0.095930	0.069427	0.038769	0.086823	0.096335
$\overset{\frown}{\tilde{q}}_p(\hat{v})$	0.090199	0.047799	0.024035	0.075077	0.091973
$\dddot{q}_p(\hat{v})$	0.094253	0.066906	0.033124	0.058418	0.067105
$\ddot{q}_p(\hat{v})$	0.059679	0.043189	0.029539	0.062456	0.077833
$\hat{q}_p(\hat{v})$	**0.054429**	0.025783	**0.007506**	0.042477	**0.052072**
$\check{q}_p(\hat{v})$	0.060339	**0.017504**	0.013291	**0.041619**	0.053664
$\breve{q}_p(\hat{v})$	**0.054776**	**0.013536**	**0.005603**	**0.027627**	**0.043617**
			Case (V)		
	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\tilde{q}_p(\hat{v})$	0.103604	0.074981	0.041871	0.093769	0.104042
$\overset{\frown}{\tilde{q}}_p(\hat{v})$	0.097415	0.051623	0.025958	0.081083	0.099331
$\dddot{q}_p(\hat{v})$	0.101793	0.072258	0.035774	0.063091	0.072473
$\ddot{q}_p(\hat{v})$	0.064453	0.046644	0.031902	0.067452	0.08406
$\hat{q}_p(\hat{v})$	**0.058783**	0.027846	**0.008106**	0.045875	**0.056238**
$\check{q}_p(\hat{v})$	0.065166	**0.018904**	0.014354	**0.044949**	0.057957
$\breve{q}_p(\hat{v})$	**0.059158**	**0.014619**	**0.006051**	**0.029837**	**0.047106**
			Case (VI)		
	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\tilde{q}_p(\hat{v})$	0.107749	0.077980	0.043545	0.097520	0.108203
$\overset{\frown}{\tilde{q}}_p(\hat{v})$	0.101312	0.053688	0.026996	0.084326	0.103304
$\dddot{q}_p(\hat{v})$	0.105865	0.075149	0.037205	0.065615	0.075372
$\ddot{q}_p(\hat{v})$	0.067031	0.048510	0.033178	0.070151	0.087422
$\hat{q}_p(\hat{v})$	**0.061135**	0.028959	**0.008431**	0.047710	**0.058487**
$\check{q}_p(\hat{v})$	0.067773	**0.019660**	0.014928	**0.046746**	0.060275
$\breve{q}_p(\hat{v})$	**0.061524**	**0.015204**	**0.006293**	**0.031031**	**0.048991**

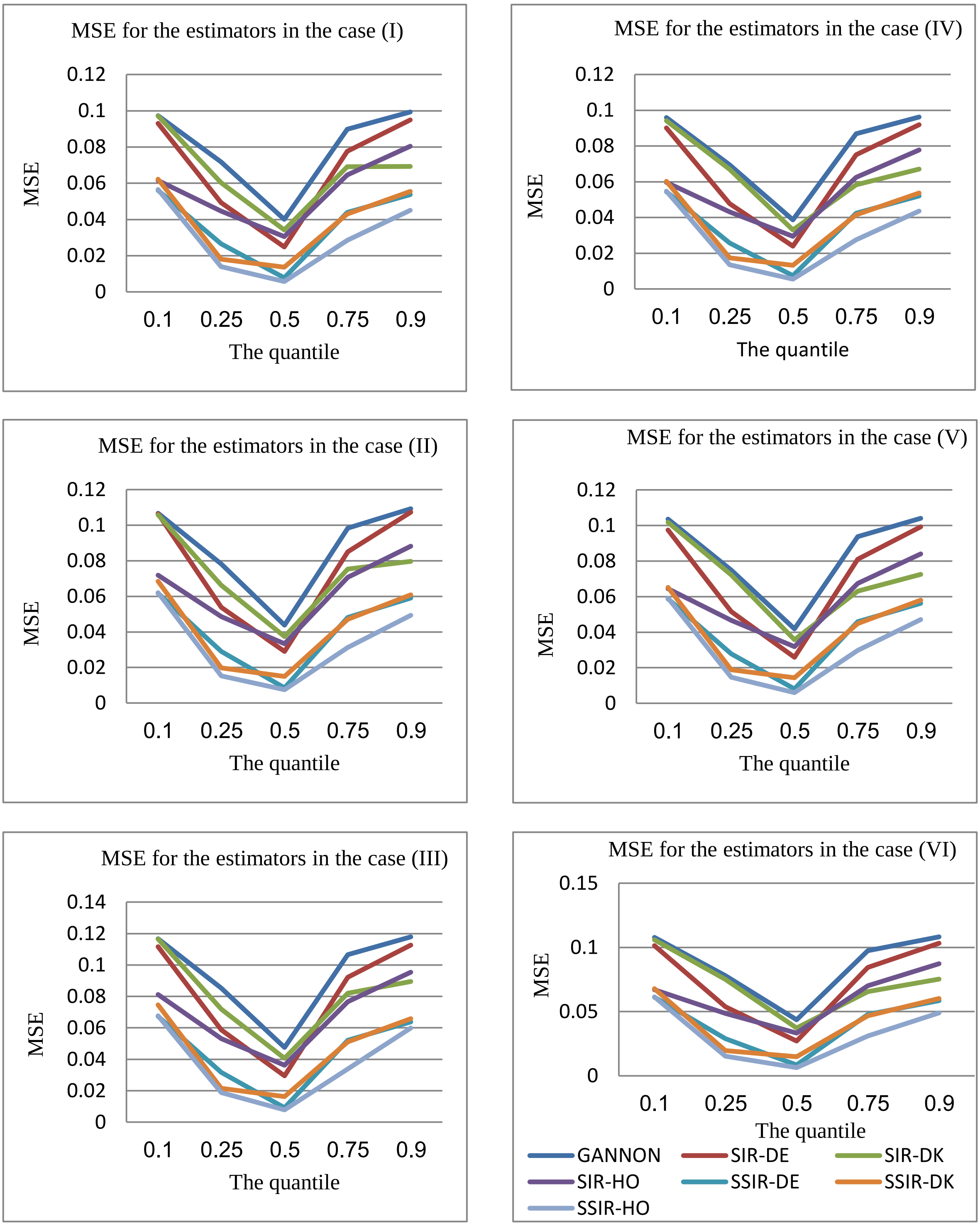

Figure.1. Mean squared error of the estimators $\hat{q}_p(\hat{v})$ (SRSIR-DK), $\ddot{\tilde{q}}_p(\hat{v})$ (SIR-DK), $\breve{q}_p(\hat{v})$ (SRSIR-DE), $\ddot{q}_p(\hat{v})$ (SIR-DE), $\check{q}_p(\hat{v})$ (SRSIR-HO), $\ddot{q}_p(\hat{v})$ (SIR-HO), and $\tilde{q}_p(\hat{v})$ (GANNO) for the different quantiles $p = (0.10, 0.25, 0.50, 0.75, 0.90)$

From Table 1 and Figure 1, it can be observed that for the all cases the estimators of sparse ride sliced inverse for non-crossing quantile regression were better than the estimators of the sliced inverse regression for non-crossing quantile . Further, we can also note that for the all cases the estimators of the sparse ride sliced inverse for non-crossing quantile and the estimators of the sliced inverse regression for non-crossing quantile are better than Gannoun et al. (2004) estimator, $\tilde{q}_p(\hat{v})$.

In addition, we can see that for the all estimators the MSE criterion in the cases (i), (ii) and (iii) are lower than the MSE criterion in cases (iv), (v) and (vi) for the same ρ.

The MSE criterion for the estimators of the sliced inverse regression for non-crossing quantile regression and $\tilde{q}_p(\hat{v})$(Gannoun et al. 2004) estimator increase significantly when ρ increases, while the MSE criterion for the estimators of the sparse ridge sliced inverse regression for non-crossing quantile regression increase in small amount when ρ increase.

For all cases, the performance of the all estimators when $\tau = 0.5$ were better than the performance of these estimators with other quantiles; while, the performance of all estimators when $\tau = 0.9$ were worse than the performance of these estimators with other quantiles.

For the all cases, the performance of the estimator $\hat{q}_\tau(\hat{v})$ was better than the performance of the other estimators when $\tau = (0.5, 0.25)$, while the performance of the estimator $\breve{q}_\tau(\hat{v})$ was better than the other estimators when $\tau = (0.1, 0.75, 0.9)$.

In general, the differences between the MSE for the estimators $\hat{q}_\tau(\hat{v})$,$\check{q}_\tau(\hat{v})$, and $\breve{q}_\tau(\hat{v})$ were small amounts.

6. Application to real data

6.1 *Description of the data set*

We applied our estimation procedure on body dimension datasets which is described in (Heinz et al. , 2003). This dataset provide body girth and skeletal diameter measurements, as well as age, weight, height and gender, are given for 507 physically active individuals - 247 men and 260 women.

We applied our estimation procedure on the body girth (circumference) measurements and the height as covariate variables against weight as a response variable for 247 men. The body girth (circumference) measurements contained twelve variables, namely, x_1 *(Shoulder Girth)*, x_2 *(Chest Girth)*, x_3 *(Waist Girth)*, x_4 *(Navel Girth)*, x_5 *(Hip Girth)*, x_6 *(Thigh Girth)*,
x_7 *(Flexed Bicep Girth)*, x_8 *(Forearm Girth)*, x_9 *(Knee Girth)*, x_{10} *(Maximum Girth)*, x_{11} *(Ankle Minimum Girth), and* x_{12} *(Wrist Minimum Girth)*.we refer to the height by x_{13}.

We computed the correlation matrix for the covariates variables, and we found the variables to be highly correlated. The maximum correlation is 0.88 between x_3 and x_4.

6.2 Application of estimation procedure

We can describe the estimation procedure by the following two steps. Assume X is the set of the p covariates ($p = 13$) in our research.

Step1. We apply the sparse ridge sliced inverse regression (SRSIR) method using the response variable Y (weight) and the covariates of X (body girth measurements and height). We determine the number $\hat{d}$ of EDR directions according to (12) and we obtain $\hat{b}_1, \hat{b}_2, \dots \dots, \hat{b}_d$ .where $\hat{b}_1, \hat{b}_2, \dots \dots, \hat{b}_d$ are the estimated EDR directions.

Step2. We estimate the non-crossing quantile regression using the methods described in section 3.

6.3The results:

We apply the estimation procedure to get the estimates to quantile curves ($\tau = 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95$) for the response variable (weight) as follow:

Step1. We find $\hat{d} = 1$ for our data and thus we obtain $\hat{b}_1$ where $\hat{b}_1$ is the first estimated EDR direction. Table2 describes the estimated direction for (SIR) and (SRSIR) as follow:

Table2. the estimated direction for sliced inverse regression and sparse ridge sliced inverse regression

	$\hat{b}_1$-SRSIR	$\hat{b}_1$-SIR
X_1	0.0000000	0.11669728
X_2	0.0000000	0.10181972
X_3	0.5672437	0.35600727
X_4	0.1059630	0.01758213
X_5	0.5228808	0.21574339
X_6	0.0000000	0.30413007
X_7	0.0000000	0.16641450
X_8	0.0000000	0.57422016
X_9	0.0000000	0.24255242
X_{10}	0.0000000	0.30291702
X_{11}	0.0000000	-0.10121251
X_{12}	0.0000000	-0.22163315
X_{13}	0.6273771	0.37755221

From our analysis the first (SRSIR) linear combination $\hat{v}$ will be:

$$\hat{v} = 0.5672437\, X_3 + 0.1059630\, X_4 + 0.5228808\, X_5 + 0.6273771\, X_{13}$$

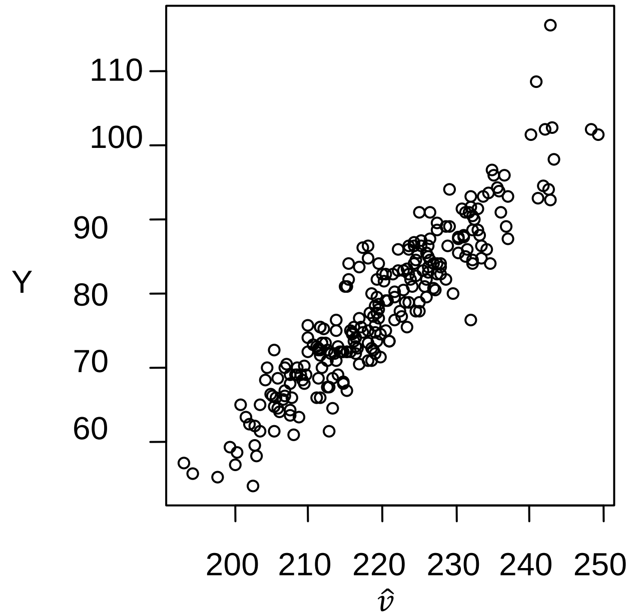

Figure. 2. Scatter plot of the response variable (weight) against the first SRSIR index. The first index has a strong structure in the scatter plot matrix.

<u>Step 2</u>: Figure3 describes the estimated quantiles regression curves

$\tau = (0.05,0.10,0.25,0.50,0.75,0.90,0.95)$ using the estimators $\hat{q}_\tau(\hat{v})$, $\check{q}_\tau(\hat{v})$, and $\breve{q}_\tau(\hat{v})$ respectively. We can see from the figure.3 the estimators $\hat{q}_\tau(\hat{v})$, $\check{q}_\tau(\hat{v})$, and $\breve{q}_\tau(\hat{v})$ are non-crossing quantile estimators.

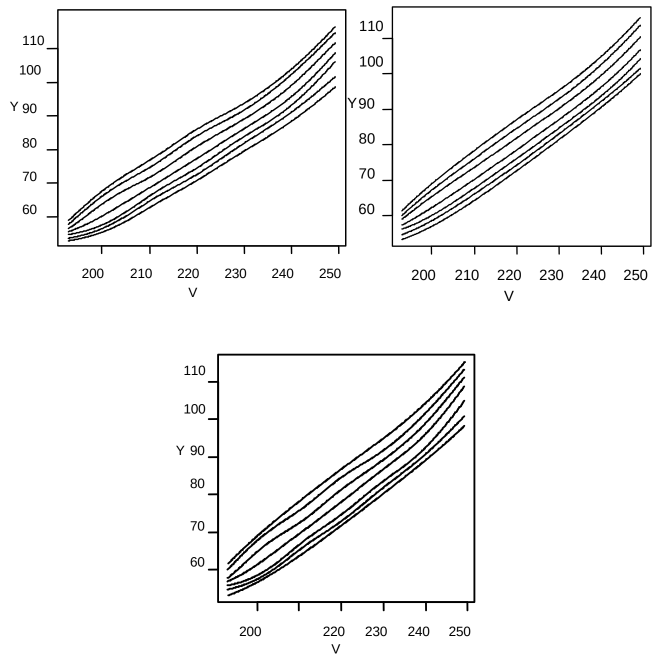

Figure. 3. The estimated quantile regression curves $p = (0.05,0.10,0.25,0.50,0.75,0.90,0.95)$ by using the estimators $\hat{q}_p(\hat{v})$, $\check{q}_p(\hat{v})$, and $\breve{q}_p(\hat{v})$ respectively from the left to the right.

References

Bondell, H. D., Reich, B. J. and Wang, H. (2010). Non-crossing quantile regression curve estimation. *Biometrika* **97**, 825-838.

Chaudhuri, P.(1991). Global nonparametric estimation of conditional quantile functions and their derivative. Journal of Multivariate Analysis 39, 246-269.

Chernozhukov, V., Fernandez-val, I. and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. Biometrika 96, 559-75.

Cook, R. (1998). Regression graphics: ideas for studying the regression through graphics. Wily, New York.

Cook, R.D. (2004). Testing predictor contributions in sufficient dimension reduction. The Annals of Statistics 32,1061-1092.

Dette, H. and Volgushev, S. (2008). Non-crossing non-parametric estimates of quantile curves. Journal of the Royal Statistical Society, Ser. B **70**, 609-627.

Dette, H. and Scheder., R. (2011). Estimation of additive quantile regression. Annals of the Institute of Statistical Mathematics 63,245-265.

Gannoun, A., Girard, S. and Saracco, J. (2004). Sliced inverse regression in reference curves estimation . Computational Statistics and Data Analysis 46,103-122.

Gooijer ,J.G. De and Zerom, D.(2003). On additive conditional quantiles with high-dimensional covariates. Journal of the American Statistical Association 98,135-146.

He, X.(1997). Quantile curves without crossing. The American Statistician 51,186-191.

He, X. and Ng, P. (1999). COBS: qualitatively constrained smoothing via linear programming. Computational Statistics 14, 315-337.

Heinz, G., Peterson, L., Johnson, R. and Kerk, C. (2003). Exploring Relationships in Body Dimensions. Journal of Statistics Education 11,2. Available at :
www.amstat.org/publications/jse/v11n2/datasets.heinz.html.

Horowitz, J.L and Lee S. (2005). Nonparametric Estimation of an Additive Quantile Regression Model. Journal of the American Statistical Association 100, 1238-1249.

Koenker, R., 2005. Quantile Regression. Cambridge: Cambridge University Press.

Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile smoothing splines. Biometrika 81,673-680.

Li, K. (1991). Sliced inverse regression for dimension reduction (with discussion). Journal of the American Statistical Association 86, 316–342.

Li ,L. and Yin, X. (2008). Sliced Inverse Regression with Regularizations. Biometrics 64, 124-131.

Li, Y., Graubard, B. and Korn, E. (2010). Application of nonparametric quantile regression to
body mass index percentile curves from survey data. Statistics in Medicine 29, 558-572.

Neocleous, T. and Portnoy, S. (2008). On monotonicity of regression quantile functions. Statistics and Probability Letters 78, 1226-9.

Ni, L. (2005). A note on shrinkage sliced inverse regression. Biometrika 92, 242-247.

Similä ,T. (2006). Self-organizing map visualizing conditional quantile functions with multidimensional covariates. Computational Statistics and Data Analysis 50 , 2097 - 2110.

Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. The Annals of Statistics 10,1040-1053.

Takeuchi, I., Le, Q. V., Sears, T. D. and Smola, A. J.(2006). Nonparametric quantile estimation. Journal of Machine Learning Research **7**, 1231–1264.

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

Yu, K., Lu, Z., and Stander, J. (2003). Quantile regression: applications and current research area. The Statistician 52,331-350.

Yu, K. and Lu, Z. (2004). Local linear additive quantile regression. Scandinavian Journal of Statistics 31 , 333-346.

Yu, K., 2010. Kernel conditional quantile estimation. Presentation.

Yebin, C., Gooijer , J.G. D. and Zerom , D.(2011). Efficient estimation of an additive quantile regression model. Scandinavian Journal of Statistics 38 , 46-62**.**

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

Wu, Y. and Liu (2009). Stepwise multiple quantile regression estimation using non-crossing constraints. Statistics and Its Interface 2 , 299-310.

Zhu, L., Miao, B. and Peng, H. (2006). On sliced inverse regression with large dimensional covariates. Journal of American Statistical Association 101, 630-643.