

# Penalized Single-Index Quantile Regression

Ali Alkenani<sup>1,2</sup> & Keming Yu<sup>1</sup>

<sup>1</sup> Mathematical Science Department, School of Information Systems, Computing and Mathematics, Brunel University, Uxbridge, UK

<sup>2</sup> Statistics Department, College of Administration and Economics, Al-Qadisiyah University, Al Diwaniyah, Iraq  
Correspondence: Ali Alkenani, Mathematical Science Department, School of Information Systems, Computing and Mathematics, Brunel University, Uxbridge, UK. E-mail: mapgaja@brunel.ac.uk

Received: March 10, 2013 Accepted: April 26, 2013 Online Published: May 27, 2013

doi:10.5539/ijsp.v2n3p12 URL: <http://dx.doi.org/10.5539/ijsp.v2n3p12>

## Abstract

The single-index (SI) regression and single-index quantile (SIQ) estimation methods product linear combinations of all the original predictors. However, it is possible that there are many unimportant predictors within the original predictors. Thus, the precision of parameter estimation as well as the accuracy of prediction will be effected by the existence of those unimportant predictors when the previous methods are used.

In this article, an extension of the SIQ method of Wu et al. (2010) has been proposed, which considers Lasso and Adaptive Lasso for estimation and variable selection. Computational algorithms have been developed in order to calculate the penalized SIQ estimates. A simulation study and a real data application have been used to assess the performance of the methods under consideration.

**Keywords:** dimension reduction, variable selection, adaptive lasso, lasso, single index model, quantile regression

## 1. Introduction

In many applications the linear relationship does not hold. So the use of linear regression to describe the relations in these cases is not suitable. The SI model is an extension of the linear regression to deal with nonlinear relationships. It is more elastic than the parametric models and retains their good properties. Besides its ability to reduce the risk of miss-pacifying the link function, it helps to overcome the “Curse of Dimensionality” (CD). Due to the index  $\mathbf{X}^T \gamma$  aggregates the high dimensionality of  $\mathbf{X}$ , many researchers have used the SI model to deal with the CD problem. The notion of the CD was reported by Richard Bellman (1961). It is caused by the exponential increase in volume associated with adding extra dimensions to an associated mathematical space. This means that the increasing of the sparsity will be exponential, given a fixed amount of data points. This problem causes the standard statistical tools to break down quickly in high dimensional data.

The single-index technique has been proven over the years to be an active and efficient method to deal with high-dimensional estimation problems in standard mean regression. It has gained much attention in recent years because of its usage in many fields. For example, discrete choice analysis in econometrics and dose–response models in biometrics (Härdle et al., 1993). It has the following form:

$$y = g(\mathbf{X}^T \gamma) + \varepsilon. \quad (1)$$

Where  $y$  is the univariate response variable and  $\mathbf{X}$  is a vector of  $p$ -dimensional covariates,  $g(\cdot)$  is an unknown univariable measurable function,  $\varepsilon$  is a random error satisfying  $E(\varepsilon/\mathbf{X}) = 0$ , and  $\gamma$  is the unknown SI vector coefficient satisfying  $\|\gamma\| = 1$  and the first component  $\gamma_1$  is positive for the sake of model identifiability. Here  $\|\cdot\|$  denotes the Euclidean norm.

There are three types of methods that have been suggested to estimate  $\gamma$  in the literature. The first type utilizes the truth that  $\gamma$  is proportional to the  $\frac{\partial g(x)}{\partial x} = \gamma g'(x \gamma)$ , which includes the average derivative estimation method (Härdle & Stoker, 1989), the structure adaptive method (Hristache et al., 2001) and the outer product of gradients (OPG) method (Xia et al., 2002). The second type contains methods that estimate  $g$  and  $\gamma$  in the same time. For example, the semiparametric least squares estimation method (Ichimura, 1993) and the MAVE method (Xia et al., 2002). The third type consist of methods that use regressing  $\mathbf{X}$  on  $y$  instead of regressing  $y$  on  $\mathbf{X}$  and were originally

proposed to deal with the sufficient dimension reduction (SDR).

For the sake of reducing the predictor dimension without losing any regression information, the SDR theory (Cook, 1998) has been proposed. Many methods have been developed to estimate the SDR space. Some of these methods focusing on the central subspace which is denoted by  $S_{y|X}$ . These methods try to find an answer to the question, “how does the conditional distribution of  $y|X$  change when the value assumed by  $X$  changes?” For example, the sliced inverse regression method (SIR) (Li, 1991), the sliced average variance estimation method (SAVE) (Cook & Weisberg, 1991) and the directional regression (Li & Wang, 2007).

The majority of known estimation approaches for model (1) were constructed on either least squares or likelihood based methods. Thus, these approaches are expected to be sensitive to outliers. In contrast to the stated approaches, quantile regression (QR) (Koenker & Bassett, 1978) provides a robust alternative. It supplies us with a full statistical analysis of the stochastic relationships among the predictors and the response variable. QR has been applied in different fields such as econometrics, finance, microarrays, medical and agricultural studies—see Koenker (2005) and Yu et al. (2003) for more details. Many researchers have studied the QR methods in the literature; see for example, He and Shi (1996), He et al. (2002), Lee (2003), Cai and Xu (2009), Wang et al. (2010), Kai et al. (2011), among others.

A lot of work exists on nonparametric standard mean regression, however, very little exists on nonparametric QR. Nonparametric QR includes local linear methods and the spline methods. The local linear QR method for univariate QR is proposed by Yu and Jones (1998), see Koenker (2005) and Koenker et al. (1994) for more details. Theoretically, while the extension of nonparametric conditional quantiles from univariate to higher dimension cases is quite clear, its practical success is impeded by the “curse of dimensionality”. Therefore, the challenge is to reduce the  $p$ -dimensional predictor  $X$  without the loss of any information on the conditional distribution of  $y|X$  and without needing a pre-specified parametric model.

Recently, dimension reduction methods for nonparametric QR models have received a lot of attention in the statistical literature. Many approaches attempt to reduce the  $p$ -dimensional predictor  $X$  without losing information and then estimate the conditional quantile. Chaudhuri (1991), Gooijer and Zerom (2003), Yu and Lu (2004), Horowitz and Lee (2005), Dette and Scheder (2011), and Yebin et al. (2011) used variants of the adaptive model in order to reduce the dimension and thereafter estimate the conditional quantiles. To introduce a more efficient estimator of conditional quantiles, Gannon et al. (2004) used the SIR to reduce the dimensionality of the covariates. Recently, Wu et al. (2010) proposed the SIQ. A practical algorithm is introduced where the authors used the local linear QR to estimate the unknown link function and linear QR to estimate the parametric index. Jiang et al. (2012) proposed the local linear composite QR estimator for a single-index model. Hua et al. (2012) developed a Bayesian method for fitting models with a single-index using conditional QR.

The selection of predictors plays a crucial role in building a multiple regression model. The choice of a suitable subset of predictors can help to improve prediction accuracy. Also, in practice, the interpretation of a smaller subset of predictors is often easier to understand and interpret (Li et al., 2010). Variable selection by penalizing the classical least squares has attracted significant research interest. See for example, least absolute shrinkage and selection operator Lasso (Tibshirani, 1996), smoothly clipped absolute deviation SCAD (Fan & Li, 2001) and Adaptive Lasso (Zou, 2006).

Because SI methods produce linear combinations of all of the original predictors, the variable selection approaches become very necessary for SI modelling when the number of predictor variables is large and when there are unimportant predictors. Some researchers suggested to generalize a number of classical variable selection procedures from linear regression to the SI model such as the Akaike information criterion and cross-validation (AIC) and others, see for example, Naik and Tsai (2001) and Kong and Xia (2007). These procedures are computationally intensive and unstable.

Some research has proposed to generalize the Lasso (Tibshirani, 1996) under the SI model assumptions. Under the framework of sufficient dimension reduction, Li and Yin (2008) combined the idea of Lasso with the SIR. Recently, Wang and Yin (2008) suggested the sparse MAVE (sMAVE). The authors proposed to add an  $l_1$  penalty term  $\lambda \sum_{k=1}^p |\gamma_k|$  to the MAVE loss function to obtain the sMAVE. The idea of combining MAVE and Lasso, which is proposed in Wang and Yin (2008), was exploited by Zeng et al. (2012) by proposing an  $l_1$  penalty function that penalizes the  $\gamma$  and the norm of the  $\frac{\partial g(x)}{\partial x}$  together.

Koenker (2004) proposed to use the regularization in QR. In order to shrink individual effects towards a common value, the author put an  $l_1$  penalty onto the random effects in a mixed-effect QR model. Li and Zhu (2008) evolved

a piecewise linear solution path for the  $l_1$  penalized QR. Moreover, Wu and Liu (2009) proposed penalized QR with the SCAD and the Adaptive Lasso penalties. Yuan and Yin (2010) proposed a Bayesian approach to shrink the random effects towards a common value by introducing an  $l_2$  penalty to the usual QR check function. Li et al. (2010) suggested Bayesian regularized QR. They proposed different penalties such as Lasso, group Lasso and elastic net penalties. Alhamzawi et al. (2012) extended the Bayesian Lasso QR reported in Li et al. (2010) to Bayesian Adaptive Lasso QR by using different penalization parameters for different regression coefficients.

In this article, we propose an extension of the SIQ model of Wu et al. (2010) by considering Lasso and Adaptive Lasso for estimation and variable selection. Computational algorithms have been developed in order to calculate the penalized SIQ estimates. Our motivating example is an analysis of the Boston housing data which is previously analyzed by many researchers and available in the package ('MASS') in R. The objective of this study is to investigate the relationship between the median value of owner-occupied homes in \$1000s and 13 statistical measurements on the 506 census tracts in suburban Boston from the 1970 census. In this paper we are interested in selecting the most significant statistical measurements of the 13 statistical measurements for the SIQ model, relating to the median value of owner-occupied homes in \$1000s. A certain correlation is present between the predictors in the Boston housing data. For example, the correlation coefficient is (-0.7692) between the nitric oxides concentration and the weighted mean of distances to five Boston employment centres, (0.7636) between the nitric oxides concentration and the proportion of non-retail business acres per town, (-0.7478) between the weighted mean of distances to five Boston employment centres and proportion of owner-occupied units built prior to 1940 and so on. The selection of variables is important in this application, in order to know which predictors have coefficients that vary among subjects. The high correlation between the predictors is an argument to use the Adaptive Lasso because the procedure deals with correlated predictors by using adaptive weights for the different predictors.

The remainder of the paper is organized as follows. A brief review of the SIQ method is given in section 2. Penalized SIQ with Lasso and Adaptive Lasso are introduced in section 3 and section 4 respectively. Simulation studies are conducted under different settings in section 5. The applications of the methods using real data are reported in section 6. Lastly, the conclusions are summarized in section 7.

## 2. Single-Index Quantile Regression (SIQ)

Given  $\tau \in (0, 1)$ , Wu et al. (2010) proposed the SIQ for the  $\tau$ th conditional quantile  $\theta_\tau(\mathbf{X})$  of  $y$  given  $\mathbf{X}$  as follows

$$\theta_\tau(\mathbf{X}) = g(\mathbf{X}^T \boldsymbol{\gamma}) \quad (2)$$

Where  $y$  is a real valued response variable and  $\mathbf{X}$  is a vector of  $d$ -dimensional covariates,  $g(\cdot)$  is an unknown univariable measurable function,  $\boldsymbol{\gamma}$  is the unknown SI vector coefficient satisfying  $\|\boldsymbol{\gamma}\| = 1$  and the first component  $\gamma_1$  is positive for the sake of model identifiability.

By replacing the nonparametric counterpart  $g(\mathbf{X}^T \boldsymbol{\gamma})$  in model (2) with  $\mathbf{X}^T \boldsymbol{\gamma}$ , we obtain the linear QR of Koenker and Basset (1978). For the SIQ model (2), note  $g(\cdot)$  should be  $g_\tau(\cdot)$  and  $\boldsymbol{\gamma}$  should be  $\boldsymbol{\gamma}_\tau$ . For notational convenience the subscript  $\tau$  was omitted.

Let  $\{\mathbf{X}_i, y_i\}$  be an independent identically distributed (i.i.d) sample from  $(\mathbf{X}, y)$ . For  $\mathbf{X}_i^T \boldsymbol{\gamma}$  close to  $u$ , the  $\tau$ th conditional quantile at  $\mathbf{X}_i^T \boldsymbol{\gamma}$  can be approximated by

$$g(\mathbf{X}_i^T \boldsymbol{\gamma}) \approx g(u) + g'(u)(\mathbf{X}_i^T \boldsymbol{\gamma} - u) = a + b(\mathbf{X}_i^T \boldsymbol{\gamma} - u). \quad (3)$$

Where

$$a \stackrel{\text{def}}{=} g(u) \text{ and } b \stackrel{\text{def}}{=} g'(u).$$

Wu et al. (2010) proposed an estimation procedure for estimating  $\boldsymbol{\gamma}$  and  $g(\cdot)$  as follows:

Step 0. Obtain the initial  $\widehat{\boldsymbol{\gamma}}^{(0)}$  from the average derivative estimate (ADE) of Chaudhuri et al. (1997). Standardize the  $\widehat{\boldsymbol{\gamma}}^{(0)}$  such that  $\|\boldsymbol{\gamma}\| = 1$  and  $\widehat{\gamma}_1 > 0$ .

Step 1. Given  $\widehat{\boldsymbol{\gamma}}$ , obtain  $\{\widehat{a}_j, \widehat{b}_j\}_{j=1}^n$  by solving the following

$$\min_{a_j, b_j} \sum_{i=1}^n \rho_\tau(y_i - a_j - b_j(\mathbf{X}_i - \mathbf{X}_j)^T \widehat{\boldsymbol{\gamma}}) \omega_{ij} \quad (4)$$

Where  $\rho_\tau(\cdot)$  is the check function defined by  $\rho_\tau(u) = \tau u I_{(0,\infty)}(u) - (1-\tau)u I_{(-\infty,0)}(u)$ , the weight function  $\omega_{ij} = K\left(\frac{\mathbf{X}_i^T \widehat{\gamma} - \mathbf{X}_j^T \widehat{\gamma}}{h}\right) / \sum_{i=1}^n K\left(\frac{\mathbf{X}_i^T \widehat{\gamma} - \mathbf{X}_j^T \widehat{\gamma}}{h}\right)$ ,  $K(\cdot)$  is a kernel function with the bandwidth  $h$  chosen to be optimal.

Step 2. Given  $\{\hat{a}_j, \hat{b}_j\}_{j=1}^n$ , obtain  $\widehat{\gamma}$  by solving

$$\min_{\gamma} \sum_{j=1}^n \sum_{i=1}^n \rho_\tau\left(y_i - \hat{a}_j - \hat{b}_j (\mathbf{X}_i - \mathbf{X}_j)^T \gamma\right) \omega_{ij} = \arg \min_{\gamma} \sum_{j=1}^n \sum_{i=1}^n \rho_\tau\left(y_{ij}^* - \mathbf{X}_{ij}^{*T} \gamma\right) \omega_{ij}^* . \quad (5)$$

Where

$$y_{ij}^* = y_i - \hat{a}_j, \mathbf{X}_{ij}^* = \hat{b}_j (\mathbf{X}_i - \mathbf{X}_j), \text{ and } \omega_{ij}^* = \omega_{ij} \text{ evaluated at the current estimate of } \gamma.$$

In step 2,  $\gamma$  is estimated through the linear QR without intercept on  $n^2$  observations  $\{y_{ij}^*, \mathbf{X}_{ij}^*\}_{i,j=1}^n$  with known weights  $\{\omega_{ij}^*\}_{i,j=1}^n$  evaluated at the estimate of  $\gamma$  from the previous iteration.

Step 3. Continue repeating the steps 1 and 2 until convergence.

The standardization of  $\widehat{\gamma}$  is done as  $\gamma = \text{sign}_{1\gamma} / \|\gamma\|$ , where  $\text{sign}_{1\gamma}$  is the sign of the first component of  $\gamma$ . The final estimate of  $g(\cdot)$  is  $\hat{g}(u; h, \widehat{\gamma}) = \hat{a}$  where

$$(\hat{a}, \hat{b}) = \arg \min_{(a,b)} \sum_{i=1}^n \rho_\tau\left(y_i - a - b (\mathbf{X}_i^T \widehat{\gamma} - u)\right) K\left(\frac{\mathbf{X}_i^T \widehat{\gamma} - u}{h}\right) \quad (6)$$

### 3. Single-Index Quantile Regression with Lasso Penalty (LSIQ)

The Lasso is proposed by Tibshirani (1996) for simultaneous variable selection and parameter estimation. According to the Lasso, the residual sum of squares is minimized subject to the  $\sum_{k=1}^p |\gamma_k|$  being less than a constant. By assuming this constraint, the Lasso shrinks some coefficients and set other to 0. As an extension to Lasso Tibshirani (1996), Li and Zhu (2008) suggested Lasso QR for simultaneous estimation and variable selection in QR models and it is given by:

$$\min_{\gamma} \sum_{i=1}^n \rho_\tau\left(y_i - \mathbf{X}_i^T \gamma\right) + \lambda \sum_{k=1}^p |\gamma_k| \quad (7)$$

Where  $\lambda > 0$  is the parameter controlling the value of penalty given. The  $\lambda \sum_{k=1}^p |\gamma_k|$  in (7) is the  $l_1$  penalty QR, which is important for the success of the Lasso.

The LSIQ is proposed here according to an algorithm similar to the algorithm in section 2, except in the initial step where we obtain the  $\widehat{\gamma}^{(0)}$  from the Lasso linear QR from Li and Zhu (2008). Also, in step 2, given  $\{\hat{a}_j, \hat{b}_j\}_{j=1}^n$ , we obtain  $\widehat{\gamma}_{Lasso}$  by solving

$$\begin{aligned} & \min_{\gamma} \sum_{j=1}^n \sum_{i=1}^n \rho_\tau\left(y_i - \hat{a}_j - \hat{b}_j (\mathbf{X}_i - \mathbf{X}_j)^T \gamma\right) \omega_{ij} + \lambda \sum_{k=1}^p |\gamma_k| \\ & = \arg \min_{\gamma} \sum_{j=1}^n \sum_{i=1}^n \rho_\tau\left(y_{ij}^* - \mathbf{X}_{ij}^{*T} \gamma\right) \omega_{ij}^* + \lambda \sum_{k=1}^p |\gamma_k| \end{aligned} \quad (8)$$

The final estimate of  $g(\cdot)$  is  $\hat{g}(u; h, \widehat{\gamma}_{Lasso}) = \hat{a}$ , where

$$(\hat{a}, \hat{b}) = \arg \min_{(a,b)} \sum_{i=1}^n \rho_\tau\left(y_i - a - b (\mathbf{X}_i^T \widehat{\gamma}_{Lasso} - u)\right) K\left(\frac{\mathbf{X}_i^T \widehat{\gamma}_{Lasso} - u}{h}\right) \quad (9)$$

### 4. Single-Index Quantile Regression with Adaptive Lasso Penalty (ALSIQ)

Under specific situations, Lasso has been shown to be consistent. Zou (2006) derived a necessary condition for the Lasso to be consistent. Consequently, the Lasso is inconsistent in other certain conditions. The Adaptive Lasso, where adaptive weights are used for penalizing different coefficients, is suggested by Zou (2006). The author showed that the major advantage of the Adaptive Lasso estimator, compared to the Lasso estimator is that the Adaptive Lasso estimator has the oracle property. Zou (2006) stated that the LARS algorithm can be used for solving the Adaptive Lasso. Wu and Liu (2009) proposed the Adaptive Lasso QR method, which solves the following minimization problem

$$\min_{\gamma} \sum_{i=1}^n \rho_\tau\left(y_i - \mathbf{X}_i^T \gamma\right) + \lambda_n \sum_{k=1}^p \widetilde{\omega}_k |\gamma_k| \quad (10)$$

Where the weights are set to be  $\tilde{\omega}_k = 1/|\tilde{\gamma}_k|^\delta, k = 1, \dots, p, \tilde{\gamma}$  is the non-penalized QR estimate and  $\delta > 0$ .

The ALSIQ has been proposed according to the algorithm similar to the algorithms in section 2 and 3, except in the initial step we obtained the  $\tilde{\gamma}^{(0)}$  from the Adaptive Lasso linear QR of Wu and Liu (2009). Also, in step 2, given  $\{\hat{a}_j, \hat{b}_j\}_{j=1}^n$ , we obtained  $\widehat{\gamma}_{ALasso}$  by solving

$$\begin{aligned} & \min_{\gamma} \sum_{j=1}^n \sum_{i=1}^n \rho_{\tau} \left( y_i - \hat{a}_j - \hat{b}_j (\mathbf{X}_i - \mathbf{X}_j)^T \gamma \right) \omega_{ij} + \lambda_n \sum_{k=1}^p \tilde{\omega}_k |\gamma_k| \\ & = \arg \min_{\gamma} \sum_{j=1}^n \sum_{i=1}^n \rho_{\tau} \left( y_{ij}^* - \mathbf{X}_{ij}^{*T} \gamma \right) \omega_{ij}^* + \lambda_n \sum_{k=1}^p \tilde{\omega}_k |\gamma_k| \end{aligned} \tag{11}$$

So we can obtain  $\widehat{\gamma}_{ALasso}$  by solving the minimization problem in (11) with Adaptive Lasso linear QR by using LARS algorithm. See Wu and Liu (2009).

The final estimate of  $g(\cdot)$  is  $\hat{g}(u; h, \widehat{\gamma}_{ALasso}) = \hat{a}$  where

$$(\hat{a}, \hat{b}) = \arg \min_{(a,b)} \sum_{i=1}^n \rho_{\tau} \left( y_i - a - b (\mathbf{X}_i^T \widehat{\gamma}_{ALasso} - u) \right) K \left( \frac{\mathbf{X}_i^T \widehat{\gamma}_{ALasso} - u}{h} \right) \tag{12}$$

The R codes for the proposed methods are available from the authors.

### 5. A Simulation Study

Many simulations have been implemented in order to check the performance of the suggested methods and some examples are reported below:

**Example 1**  $R = 200$  data-sets were generated with size  $n = 300$  observations from the following model where the error term follows an asymmetric (exponential) distribution:

$$y = 5 \cos(u) + \exp(-u^2) + \varepsilon$$

Where  $u = \mathbf{X}^T \gamma, \mathbf{X} = (x_1, \dots, x_5)^T, \gamma = (1, 2, 0, 0, 0)^T / \sqrt{5}, x_i$  i.i.d.  $\sim \text{Unif}(0, 1); i = 1, 2, \dots, 5$ , the error term  $\varepsilon \sim \text{Exp}(0.5), x_i$ 's and  $\varepsilon$  are mutually independent. The  $\gamma$  is estimated for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$ .

Table 1. The mean and standard division of MSE (mean squared error) for  $\mathbf{X}^T \widehat{\gamma}$  which is estimated by the ALSIQ, LSIQ and SIQ based on the model in example 1 and for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

		SIQ	LSIQ	ALSIQ
$\tau = 0.10$	M.MSE	0.0014	0.0006	0.0005
	SD.MSE	0.0011	0.0005	0.0004
$\tau = 0.25$	M.MSE	0.0046	0.0022	0.0020
	SD.MSE	0.0049	0.0026	0.0022
$\tau = 0.5$	M.MSE	0.0138	0.0046	0.0046
	SD.MSE	0.0128	0.0064	0.0065
$\tau = 0.75$	M.MSE	0.0467	0.0335	0.0311
	SD.MSE	0.0593	0.0454	0.0443
$\tau = 0.90$	M.MSE	0.0661	0.0581	0.0509
	SD.MSE	0.0857	0.0734	0.0702

**Example 2**  $R = 200$  data-sets were generated with size  $n = 300$  observations from the following model with homoscedastic errors.

$$y = \sin \left\{ \frac{\pi (u - A)}{C - A} \right\} + 0.5 \varepsilon$$

where  $u = \mathbf{X}^T \gamma, \mathbf{X} = (x_1, \dots, x_6)^T, \gamma = (1, 1, 0, 0, 0, 0)^T / \sqrt{3}, A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$  and  $C = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$ .  $x_i$  i.i.d.  $\sim \text{Unif}(0, 1) i = 1, 2, \dots, 6; \varepsilon \sim N(0, 1); x_i$ 's and  $\varepsilon$  are mutually independent. The  $\gamma$  is estimated for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$ .

Table 2. The mean and standard division of MSE for  $\mathbf{X}^T \hat{\gamma}$  which is estimated by the ALSIQ, LSIQ and SIQ based on the model in example 2 and for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

		SIQ	LSIQ	ALSIQ
$\tau = 0.10$	M.MSE	0.0294	0.0372	0.0136
	SD.MSE	0.1025	0.0396	0.0220
$\tau = 0.25$	M.MSE	0.0077	0.0067	0.0047
	SD.MSE	0.0086	0.0070	0.0045
$\tau = 0.5$	M.MSE	0.0044	0.0043	0.0042
	SD.MSE	0.0048	0.0048	0.0048
$\tau = 0.75$	M.MSE	0.0169	0.0072	0.0031
	SD.MSE	0.0198	0.0108	0.0048
$\tau = 0.90$	M.MSE	0.0197	0.0070	0.0018
	SD.MSE	0.0230	0.0080	0.0025

**Example 3**  $R = 200$  data-sets were generated with size  $n = 300$  observations from the model  $y = \exp(\mathbf{X}^T \gamma) + \varepsilon$ , where  $\mathbf{X} = (x_1, \dots, x_{10})^T$  are generated as i.i.d standard normals. The error term is assumed to be  $\varepsilon \sim N(0, 1)$  and that it is independent of  $\mathbf{X}$ .  $\gamma = (1, 1, 1, 0, 0, 0, 0, 0, 0, 0)^T / \sqrt{3}$  is used. The  $\gamma$  is estimated for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$ .

Table 3. The mean and standard division of MSE for  $\mathbf{X}^T \hat{\gamma}$  which is estimated by the ALSIQ, LSIQ and SIQ based on the model in example 3 and for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

		SIQ	LSIQ	ALSIQ
$\tau = 0.10$	M.MSE	0.0688	0.0565	0.0412
	SD.MSE	0.0434	0.0479	0.0343
$\tau = 0.25$	M.MSE	0.0494	0.0452	0.0367
	SD.MSE	0.0325	0.0278	0.0197
$\tau = 0.5$	M.MSE	0.0403	0.0336	0.0330
	SD.MSE	0.0455	0.0300	0.0206
$\tau = 0.75$	M.MSE	0.0495	0.0370	0.0360
	SD.MSE	0.0747	0.0298	0.0272
$\tau = 0.90$	M.MSE	0.0489	0.0453	0.0406
	SD.MSE	0.0298	0.0345	0.0285

We analyzed each simulated data set using three methods. The LSIQ and ALSIQ methods, which are described in sections 3 and 4 respectively, are compared with the SIQ. The  $rq(y^* \sim X^*, \text{tau}, \text{method} = \text{"lasso"})$  function in the quantreg package is used to obtain  $\hat{\gamma}_{ALasso}$  in Equation (8). The ALassoQR function from the code of Wu and Liu (2009) (Personal communication with Wu) is used to obtain  $\hat{\gamma}_{ALasso}$  in Equation (11). Similar to Wu and Liu (2009), the  $\lambda$  was chosen via a grid search based on the tuning error in terms of the mean squared error evaluated on the data. This means that the  $\lambda$  value has been chosen to minimize the mean squared error.

According to the mean and the standard deviation of the MSE for  $\mathbf{X}^T \hat{\gamma}$ , from Tables 1, 2 and 3 and Figure 1, it can be seen that the proposed methods (ALSIQ and LSIQ) perform better than the SIQ method described in Wu et al. (2010) for all the models under consideration. This indicates that the proposed methods give precise estimates even when the error distribution is asymmetric. Most noticeably, when  $\tau = 0.10$  and  $\tau = 0.90$  the ALSIQ and LSIQ are significantly more efficient than the SIQ method.

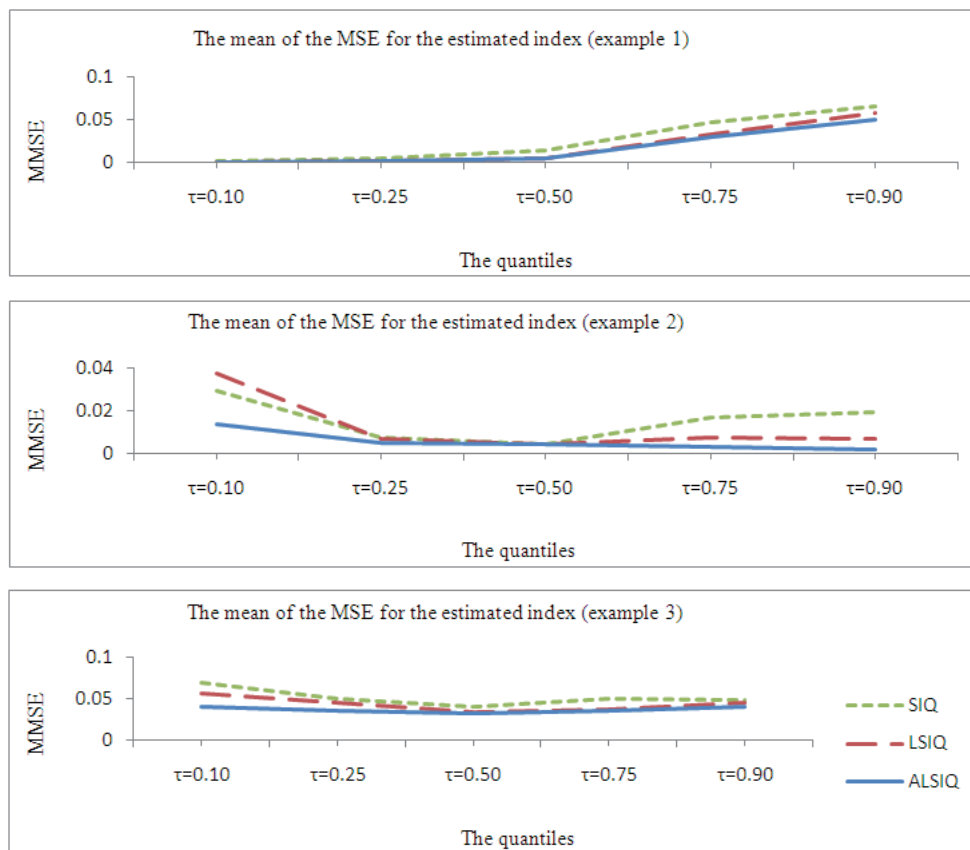


Figure 1. Plots explain the mean of MSE for  $\mathbf{X}^T \hat{\gamma}$  which is estimated by the ALSIQ, LSIQ and SIQ based on the model in examples 1, 2 and 3 respectively and for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

Table 4. The mean and MSE for the coefficients  $\widehat{\gamma}$  which are estimated by the ALSIQ, LSIQ and SIQ based on the model in example 1 and for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$ 

			$\widehat{\gamma}_1$	$\widehat{\gamma}_2$	$\widehat{\gamma}_3$	$\widehat{\gamma}_4$	$\widehat{\gamma}_5$
$\tau = 0.10$	SIQ	Mean	0.4750	0.8776	0.0081	0.0049	0.0122
		MSE	0.0017	0.0006	0.0005	0.0016	0.0012
	LSIQ	Mean	0.4609	0.8859	0.0037	0.0021	0.0037
		MSE	0.0016	0.0005	0.0003	0.0006	0.0003
	ALSIQ	Mean	0.4667	0.8829	0.0021	0.0050	0.0010
		MSE	0.0021	0.0006	0.0002	0.0002	0.0002
$\tau = 0.25$	SIQ	Mean	0.4865	0.8651	-0.0019	-0.0117	0.0097
		MSE	0.0038	0.0016	0.0018	0.0042	0.0069
	LSIQ	Mean	0.4737	0.8759	-0.0009	-0.0050	0.0010
		MSE	0.0039	0.0013	0.0010	0.0016	0.0021
	ALSIQ	Mean	0.4727	0.8766	0.0013	-0.0020	0.0029
		MSE	0.0042	0.0013	0.0009	0.0013	0.0020
$\tau = 0.50$	SIQ	Mean	0.4482	0.8664	-0.0149	0.0029	-0.0083
		MSE	0.0098	0.0032	0.0114	0.0111	0.0154
	LSIQ	Mean	0.4658	0.8725	-0.0026	0.0043	0.0016
		MSE	0.0089	0.0027	0.0049	0.0027	0.0044
	ALSIQ	Mean	0.4654	0.8727	-0.0028	0.0060	-0.0008
		MSE	0.0088	0.0027	0.0048	0.0026	0.0045
$\tau = 0.75$	SIQ	Mean	0.5429	0.7331	0.0158	0.0058	-0.0230
		MSE	0.0247	0.0531	0.0426	0.0592	0.0323
	LSIQ	Mean	0.5053	0.7881	0.0308	-0.0265	-0.0020
		MSE	0.0359	0.0290	0.0303	0.0268	0.0227
	ALSIQ	Mean	0.5486	0.7660	0.0407	-0.0433	-0.0059
		MSE	0.0412	0.0339	0.0305	0.0211	0.0181
$\tau = 0.90$	SIQ	Mean	0.5659	0.6591	-0.0229	-0.0099	-0.0299
		MSE	0.1015	0.0824	0.0456	0.0401	0.0202
	LSIQ	Mean	0.5943	0.6474	-0.0016	-0.0078	-0.0418
		MSE	0.1012	0.0923	0.0332	0.0385	0.0114
	ALSIQ	Mean	0.6029	0.6443	0.0140	0.0017	-0.0420
		MSE	0.0988	0.1017	0.0288	0.0324	0.0123



Table 5. The mean and MSE for the coefficients  $\hat{\gamma}$  which are estimated by the ALSIQ, LSIQ and SIQ based on the model in example 2 and for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

			$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$
$\tau = 0.10$	SIQ	Mean	0.6702	0.6924	-0.0253	0.0163	0.0132	-0.0421
		MSE	0.0217	0.0175	0.0136	0.0053	0.0189	0.0192
	LSIQ	Mean	0.7686	0.4655	-0.0068	0.0072	-0.0077	0.0054
		MSE	0.1217	0.1222	0.0011	0.0011	0.0016	0.0027
	ALSIQ	Mean	0.8211	0.5174	-0.0014	-0.0007	0.0086	0.0143
		MSE	0.0723	0.0447	0.0001	0.0003	0.0012	0.0052
$\tau = 0.25$	SIQ	Mean	0.7753	0.5937	0.0008	0.0238	0.0108	-0.0021
		MSE	0.0517	0.0172	0.0031	0.0054	0.0021	0.0076
	LSIQ	Mean	0.7771	0.5915	-0.0040	0.0160	0.0031	0.0008
		MSE	0.0551	0.0211	0.0017	0.0040	0.0014	0.0042
	ASIQ	Mean	0.6971	0.7016	0.0074	0.0165	0.0228	-0.0032
		MSE	0.0168	0.018	0.0031	0.0043	0.0054	0.0045
$\tau = 0.50$	SIQ	Mean	0.6884	0.7125	0.0059	0.0125	0.0211	-0.0102
		MSE	0.0143	0.0198	0.0028	0.0045	0.0056	0.0024
	LSIQ	Mean	0.7750	0.6099	0.0010	0.0197	0.0114	0.0062
		MSE	0.0472	0.0118	0.0003	0.0035	0.0022	0.0032
	ALSIQ	Mean	0.7738	0.6115	0.0015	0.0202	0.0115	0.0061
		MSE	0.04667	0.01183	0.0003	0.0035	0.0022	0.0032
$\tau = 0.75$	SIQ	Mean	0.7154	0.6355	0.0187	0.0330	0.0114	0.0232
		MSE	0.0368	0.0247	0.0058	0.0121	0.0110	0.0182
	LSIQ	Mean	0.6981	0.7016	0.0193	0.0103	0.0137	-0.0042
		MSE	0.0163	0.0172	0.0035	0.0044	0.0043	0.0052
	ALSIQ	Mean	0.7343	0.6680	0.0056	0.0115	0.0060	0.0074
		MSE	0.0274	0.0113	0.0009	0.0028	0.0017	0.0035
$\tau = 0.90$	SIQ	Mean	0.7045	0.6420	0.0264	0.0447	0.0246	0.0247
		MSE	0.0819	0.0843	0.0079	0.0160	0.0143	0.0189
	LSIQ	Mean	0.7019	0.6936	0.0087	-0.0038	0.0072	-0.0045
		MSE	0.0176	0.0155	0.0061	0.0057	0.0053	0.0057
	ALSIQ	Mean	0.7606	0.6430	0.0015	0.0068	-0.0023	0.0024
		MSE	0.0357	0.0075	0.0003	0.0015	0.0004	0.0008

Table 6. The mean and MSE for the coefficients  $\widehat{\gamma}$  which are estimated by the ALSIQ, LSIQ and SIQ based on the model in example 3 and for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

			$\widehat{\gamma}_1$	$\widehat{\gamma}_2$	$\widehat{\gamma}_3$	$\widehat{\gamma}_4$	$\widehat{\gamma}_5$	$\widehat{\gamma}_6$	$\widehat{\gamma}_7$	$\widehat{\gamma}_8$	$\widehat{\gamma}_9$	$\widehat{\gamma}_{10}$
$\tau = 0.10$	SIQ	Mean	0.5775	0.5570	0.5365	-0.0112	-0.0030	0.0008	0.0208	0.0047	0.0092	0.0023
		MSE	0.0078	0.0070	0.0099	0.0042	0.0056	0.0049	0.0069	0.0072	0.0119	0.0075
	LSIQ	Mean	0.5564	0.5784	0.5483	-0.0152	-0.0236	0.0188	0.0226	-0.0175	0.0221	-0.0275
		MSE	0.0051	0.002	0.0027	0.0088	0.0058	0.0068	0.0078	0.0074	0.0061	0.0061
	ALSIQ	Mean	0.5939	0.5380	0.5650	0.0013	-0.0068	-0.0168	0.0006	0.0033	0.0015	0.0127
		MSE	0.0089	0.0050	0.0062	0.0031	0.0019	0.0027	0.0032	0.0027	0.0032	0.0050
$\tau = 0.25$	SIQ	Mean	0.5716	0.5540	0.5633	0.0058	0.0008	0.0171	-0.0108	0.0130	-0.0035	0.0207
		MSE	0.0042	0.0038	0.0032	0.0045	0.0040	0.0084	0.0057	0.0038	0.0069	0.0071
	LSIQ	Mean	0.5698	0.5716	0.5519	-0.0272	-0.0142	-0.0234	0.0187	0.0105	-0.0089	-0.0215
		MSE	0.0032	0.0024	0.0033	0.0055	0.0039	0.006	0.0048	0.006	0.0072	0.0042
	ALSIQ	Mean	0.5685	0.5506	0.5810	0.0025	-0.0051	-0.0170	0.0061	0.0114	-0.0125	0.0025
		MSE	0.0050	0.0066	0.0042	0.0023	0.0034	0.0026	0.0047	0.0041	0.0032	0.0023
$\tau = 0.50$	SIQ	Mean	0.5574	0.5779	0.5627	-0.0073	0.0013	0.0236	0.0041	-0.0028	-0.0054	-0.0058
		MSE	0.0033	0.0012	0.0077	0.0026	0.0046	0.0042	0.0018	0.0042	0.0048	0.0072
	LSIQ	Mean	0.5958	0.5735	0.5330	0.0045	0.0053	-0.0137	0.0351	0.0003	-0.0157	-0.0112
		MSE	0.0027	0.0012	0.0038	0.0019	0.0014	0.0029	0.0079	0.0040	0.0057	0.0046
	ALSIQ	Mean	0.5737	0.5512	0.5770	-0.0047	-0.0084	0.0094	-0.0025	0.0023	-0.0095	0.0152
		MSE	0.0018	0.0036	0.0021	0.0027	0.0037	0.0032	0.0049	0.0056	0.0051	0.0033
$\tau = 0.75$	SIQ	Mean	0.5788	0.5446	0.5645	-0.0036	-0.0162	-0.0020	-0.0001	-0.0056	-0.0004	-0.0006
		MSE	0.0020	0.0083	0.0026	0.0102	0.0034	0.0032	0.0036	0.0049	0.0061	0.0084
	LSIQ	Mean	0.5833	0.5600	0.5558	0.0166	-0.0081	0.0100	0.0106	-0.0140	-0.0128	0.0046
		MSE	0.0026	0.0042	0.0026	0.0022	0.0050	0.0058	0.0046	0.0065	0.0052	0.0021
	ASIQ	Mean	0.5747	0.5514	0.5742	-0.0110	0.0069	0.0125	0.0195	-0.0015	-0.0093	0.0071
		MSE	0.0019	0.0071	0.0041	0.0033	0.0039	0.0054	0.0029	0.0028	0.0045	0.0022
$\tau = 0.90$	SIQ	Mean	0.5577	0.5646	0.5671	0.0043	-0.0235	-0.0154	0.0265	0.0087	0.0116	-0.0235
		MSE	0.0041	0.0054	0.0031	0.0041	0.0067	0.0052	0.0066	0.0033	0.0051	0.0074
	LSIQ	Mean	0.5866	0.5553	0.5512	0.0099	0.0060	0.0074	0.0072	-0.0184	-0.0169	0.0092
		MSE	0.0040	0.0054	0.0045	0.0039	0.0037	0.0039	0.0057	0.0041	0.0089	0.0025
	ALSIQ	Mean	0.5939	0.5559	0.5457	-0.0177	0.0034	-0.0067	-0.0043	-0.0097	-0.0111	0.0020
		MSE	0.0065	0.0064	0.0090	0.0039	0.0018	0.0017	0.0031	0.0048	0.0028	0.0036

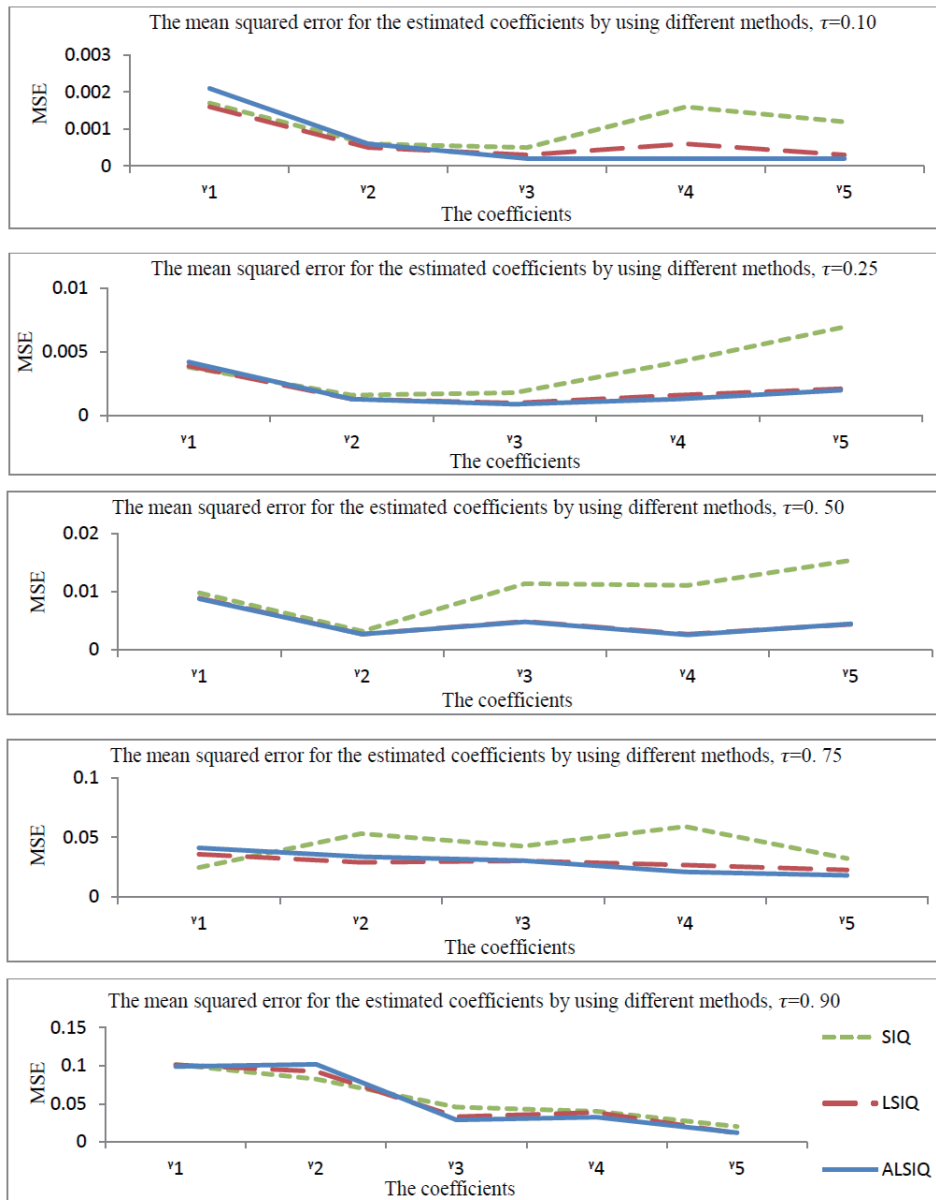


Figure 2. Plots explain the MSE for the coefficients  $\hat{\gamma}$  which are estimated by the ALSIQ, LSIQ, and SIQ based on the model in example 1 and for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

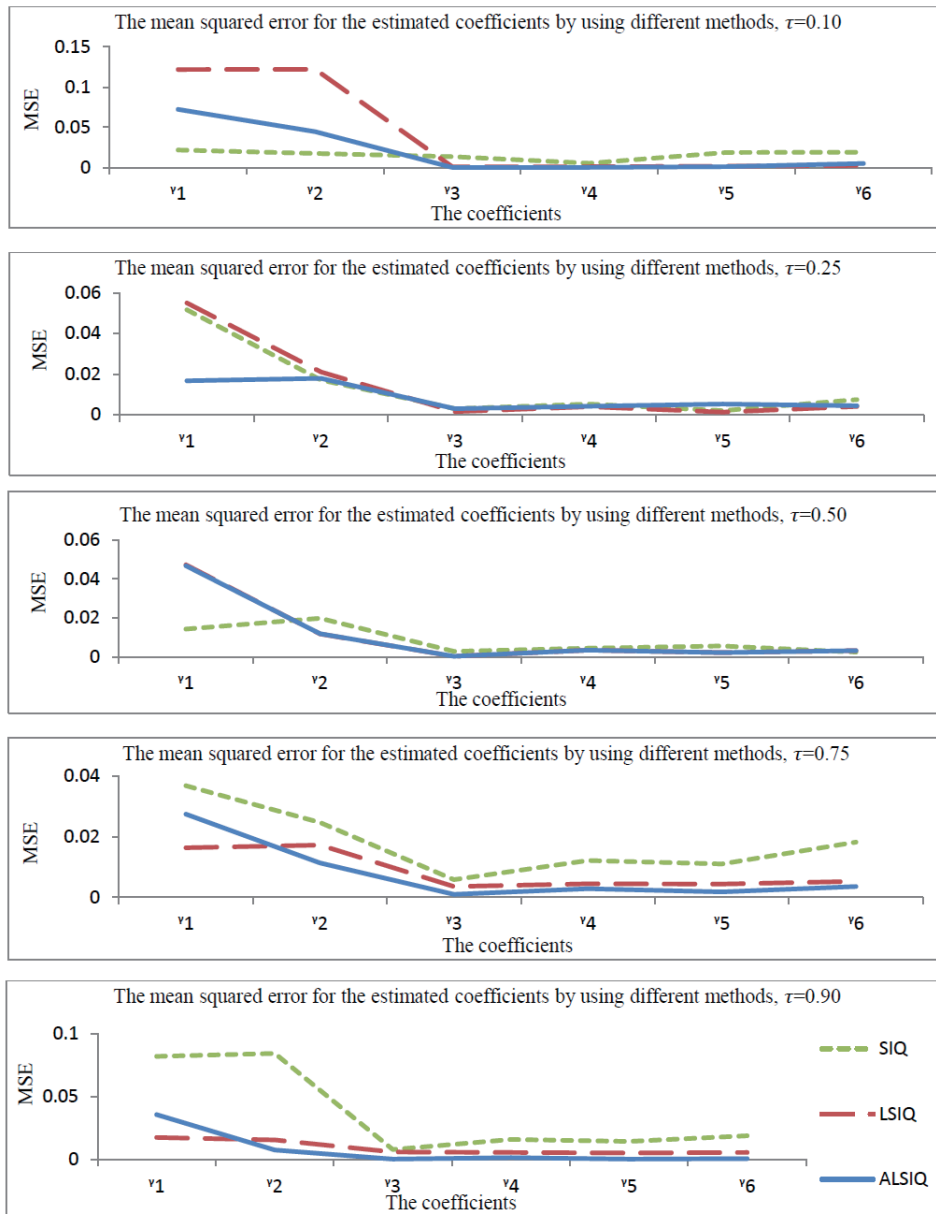


Figure 3. Plots explain the MSE for the coefficients  $\hat{\gamma}$  which are estimated by the ALSIQ, LSIQ, and SIQ based on the model in example 2 and for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

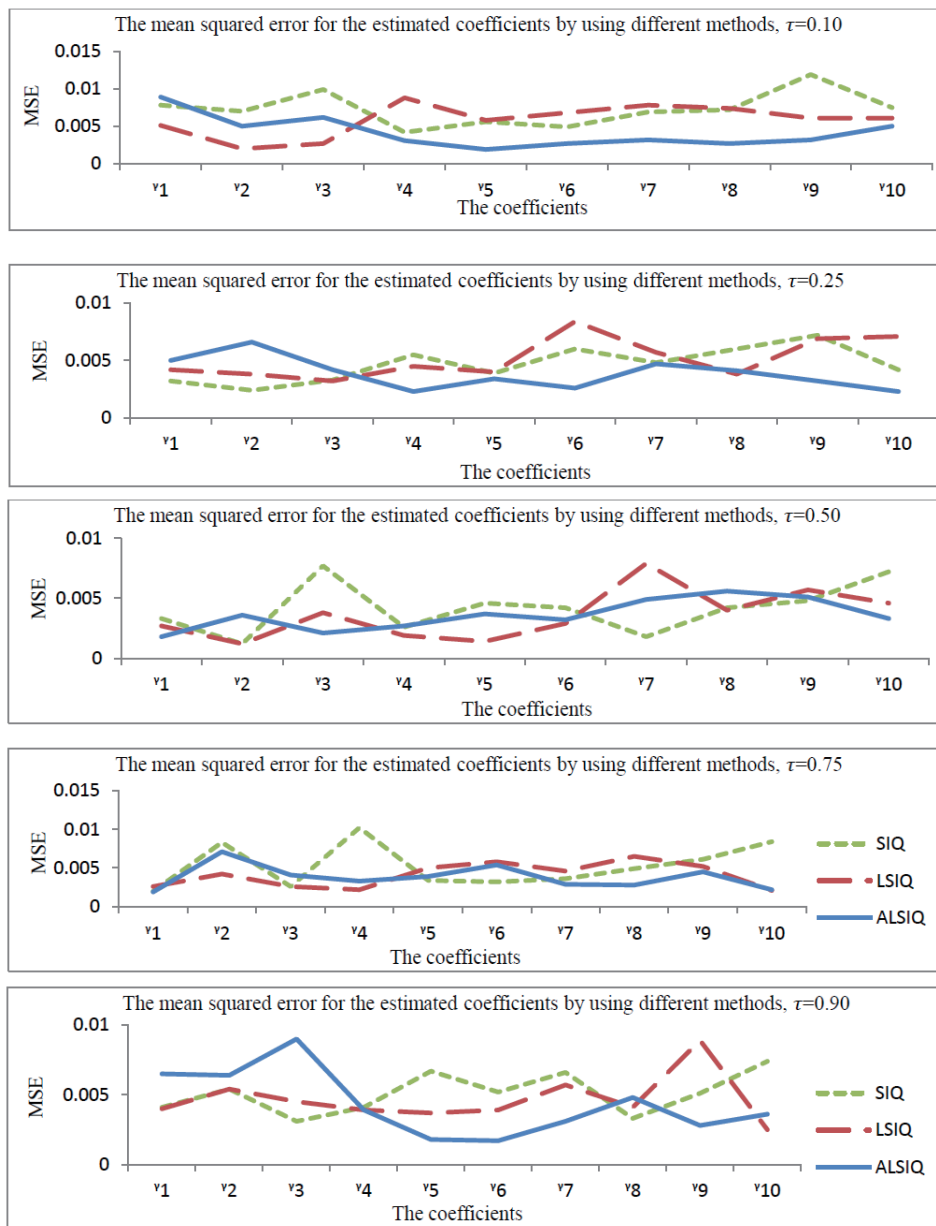


Figure 4. Plots explain the MSE for the coefficients  $\hat{\gamma}$  which are estimated by the ALSIQ, LSIQ, and SIQ based on the model in example 3 and for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

According to the MSE for the coefficients  $\hat{\gamma}$ , from Tables 4, 5 and 6 and Figures 2, 3 and 4, it can be observed that in the majority of the estimated coefficients, the proposed methods produce a lower mean squared error than the SIQ method. Furthermore, one can see that the coefficients estimators of the proposed methods are close to the true values.

The variations in the ALSIQ and LSIQ estimates are similar in the majority of cases and less than the variations in the estimate of the SIQ method.

### 6. Boston Housing Data

In this section, the methods are illustrated through an analysis of the Boston housing data. The data consist of  $n = 506$  observations on 14 variables; medv is the median value of owner-occupied homes and it refers to the response variable. The dataset consist of 13 predictors on the 506 census, which is available in the package ('MASS') in R. In our analysis, the dummy variable (chas) and the categorical variable (rad) were excluded. The predictors under consideration are crime average ( $x_1$ ), ratio of residential land zoned for lots over 25,000 sq.ft.

( $x_2$ ), ratio of non-retail business acres per town ( $x_3$ ), nitric oxides concentration ( $x_4$ ), rate number of rooms per dwelling ( $x_5$ ), ratio of owner-occupied units built prior to 1940 ( $x_6$ ), weighted mean of distances to five Boston employment centres ( $x_7$ ), tax average of the property ( $x_8$ ), pupil-teacher proportion by town ( $x_9$ ), black population ratio town ( $x_{10}$ ), and lower status of the population ( $x_{11}$ ). The response variable medv and the predictor variables were also standardised.

Table 7. The estimated coefficients  $\hat{\gamma}$  which are estimated by the ALSIQ, LSIQ and SIQ based on Boston housing data for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

		$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{11}$
$\tau = 0.10$	SIQ	0.351	0.012	-0.104	0.169	-0.494	0.191	0.228	0.311	0.220	-0.139	0.584
	LSIQ	0.342	-0.022	-0.059	0.296	-0.372	0.079	0.279	0.250	0.227	-0.202	0.644
	ALSIQ	0.446	0	0	0	-0.354	0	0.013	0.181	0.175	-0.158	0.767
$\tau = 0.25$	SIQ	0.647	-0.028	-0.030	0.031	-0.489	0.177	0.213	0.067	0.166	-0.166	0.451
	LSIQ	0.153	0	0	0.243	-0.513	0.042	0.228	0.266	0.314	-0.246	0.609
	ALSIQ	0.123	0	0	-0.252	0.659	-0.146	-0.254	-0.325	-0.328	0.250	-0.354
$\tau = 0.50$	SIQ	0.335	-0.009	-0.026	0.055	-0.500	0.130	0.217	0.059	0.206	-0.246	0.681
	LSIQ	0.110	-0.014	0	0.198	-0.597	0.092	0.246	0.165	0.325	-0.225	0.583
	ALSIQ	0.108	-0.014	0	0.198	-0.597	0.093	0.247	0.165	0.325	-0.224	0.583
$\tau = 0.75$	SIQ	0.234	-0.032	-0.006	0.085	-0.585	0.109	0.283	-0.002	0.214	-0.308	0.601
	LSIQ	0.084	-0.046	0	0.155	-0.715	0.090	0.282	0.063	0.295	-0.192	0.490
	ALSIQ	0.112	-0.003	0	0.190	-0.656	0.069	0.235	0.009	0.338	-0.217	0.547
$\tau = 0.90$	SIQ	0.174	-0.042	0.065	0.165	-0.461	-0.029	0.302	-0.090	0.204	-0.235	0.726
	LSIQ	0.033	-0.016	0.045	0.155	-0.722	0	0.187	0	0.379	-0.132	0.505
	ALSIQ	0.001	-0.057	0.053	0.069	-0.781	0	0.219	0	0.355	-0.135	0.432

Table 8. The MSE for estimated quantiles curves  $\hat{g}(\mathbf{X}^T \hat{\gamma})$  which are estimated by the ALSIQ, LSIQ and SIQ based on Boston housing data for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
SIQ	0.508	0.378	0.090	0.014	0.312
LASSO-SIQ	0.451	0.296	0.041	0.014	0.283
ALASSO-SIQ	0.440	0.261	0.039	0.013	0.292

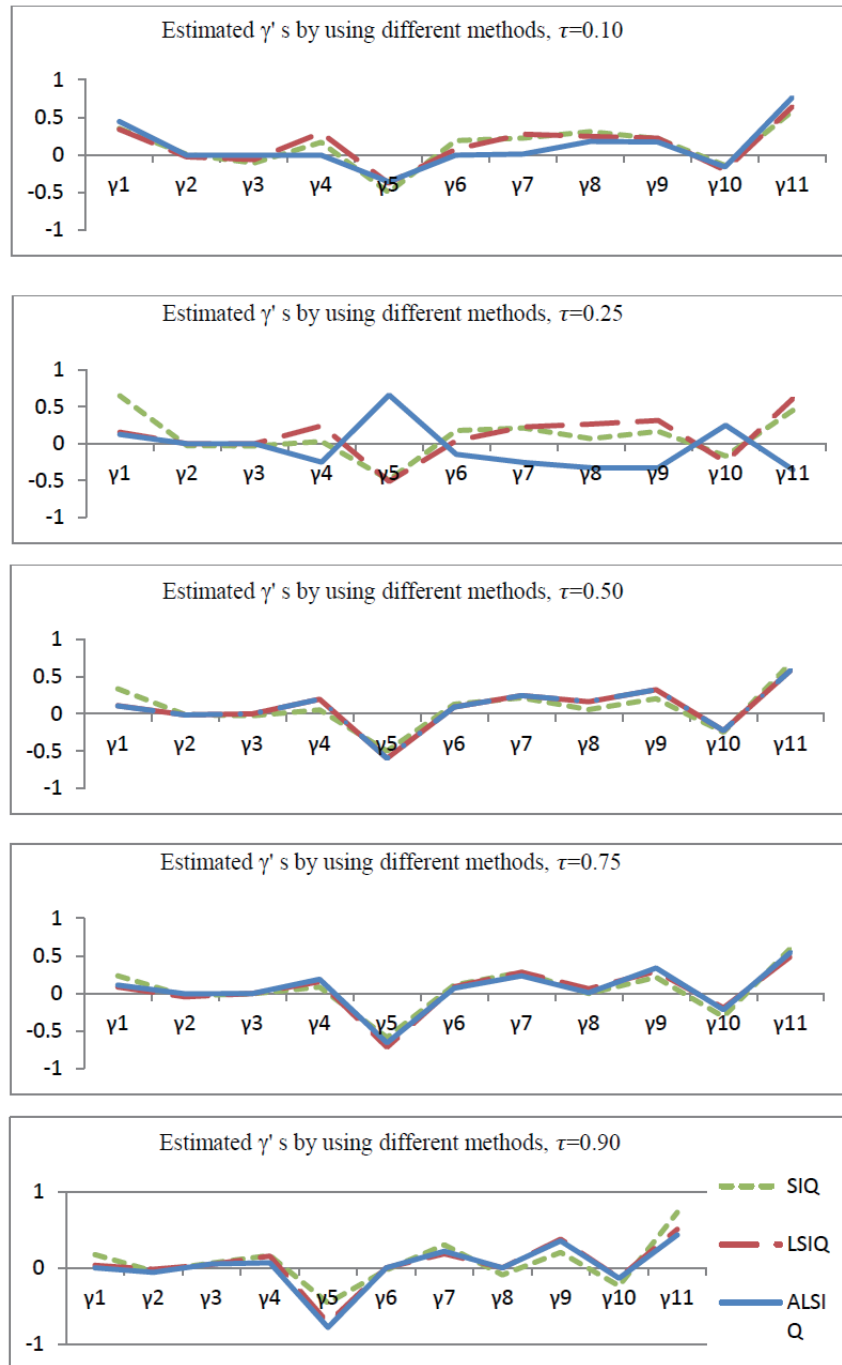


Figure 5. Plots explain the estimated coefficients  $\hat{\gamma}$  which are estimated by the ALSIQ, LSIQ and SIQ based on Boston housing data for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

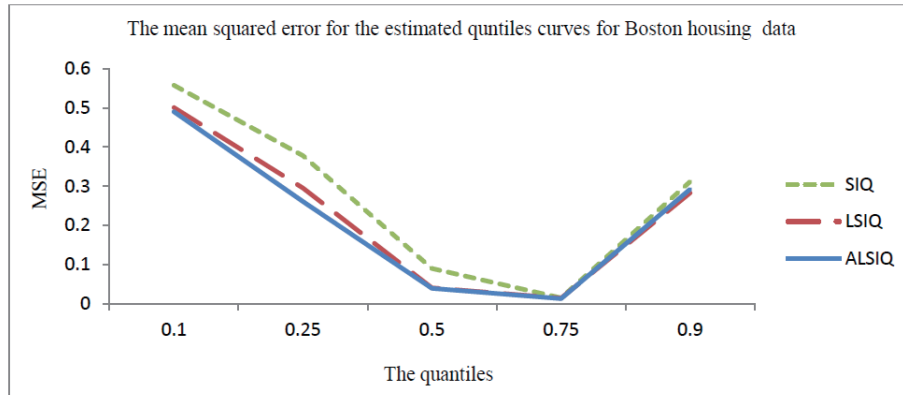


Figure 6. The MSE for the smooth estimated quantiles curves  $\hat{g}(\mathbf{X}^T\hat{\gamma})$  which are estimated by the ALSIQ, LSIQ and SIQ based on Boston housing data for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

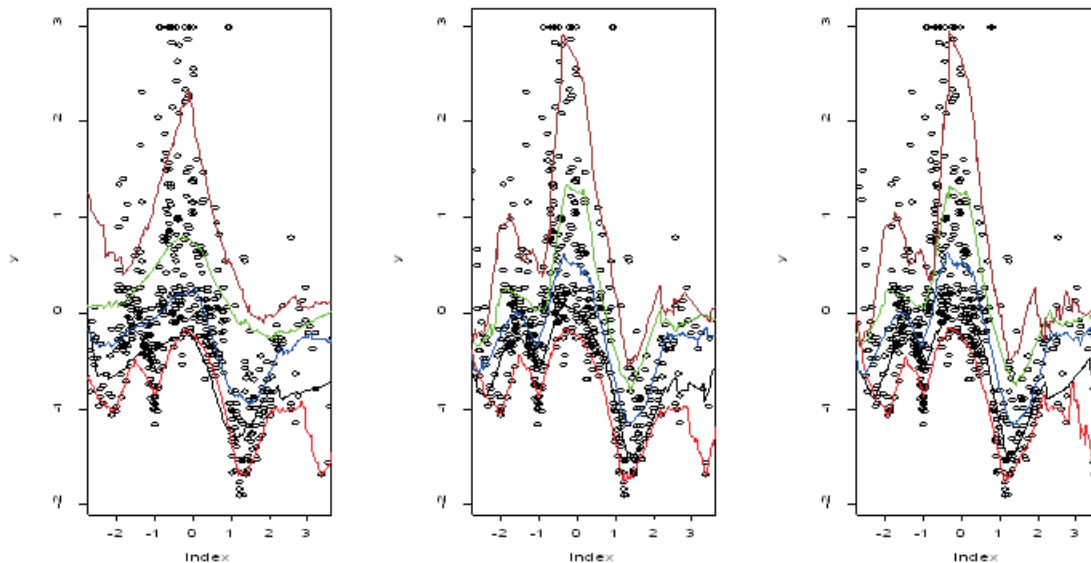


Figure 7. Plots for the smooth estimated quantiles curves  $\hat{g}(\mathbf{X}^T\hat{\gamma})$  which are estimated by the ALSIQ, LSIQ and SIQ respectively from the right to the left based on Boston housing data for  $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$

The estimated  $\hat{\gamma}$  using all the methods under consideration based on the Boston housing data are given in Table 7 and explained in Figure 5. The estimated coefficient is treated as zero if its absolute value is smaller than  $10^{-12}$ .

Table 8 and Figure 6 present the MSEs for estimated quantile curves  $\hat{g}(\mathbf{X}^T\hat{\gamma})$  which are estimated by the proposed methods and the SIQ method based on the Boston housing data for different quantile values. From Table 8 and Figure 6, it is clear that the proposed methods outperform the SIQ method in fitting the Boston housing data set. Again, it can be seen that when  $\tau = 0.10$  and  $\tau = 0.90$  the proposed methods are significantly more efficient than the other methods.

Figure 7 shows the smooth estimated quantile curves  $\hat{g}(\mathbf{X}^T\hat{\gamma})$  which are estimated by all the methods under consideration based on the Boston housing data for different quantile values.

Similar to Wu et al. (2010) possible quantile curves crossing at both tails can be seen, which due to the sparsity of data in the region concerned.

The results of the real data example confirm the results of the simulation studies that the proposed methods perform well.



## 7. Conclusions

In this study, the LSIQ and ALSIQ methods have been proposed. The effectiveness of the proposed methods is explained via many simulation examples, as well as a real data analysis. From the simulation study and the real data example, it can be concluded that the proposed methods perform well in comparison to the SIQ method. The authors believe that the proposed methods would supply helpful dimension reduction tools. Also, it would support the applicability of shrinkage methods to the SIQ models.

## Acknowledgments

The authors wish to thank the Editor, an Associate Editor and two anonymous referees for their helpful comments on an earlier version of the manuscript and suggestions which have led to an improvement of this paper. Also, we thank Prof. Yan Yu for sending us the code for the SIQ method in (Wu et al., 2010).

## References

- Alhamzawi, R., Yu, K., & Benoit, D. (2012). Bayesian adaptive LASSO quantile regression. *Statistical Modelling*, 12, 279-297. <http://dx.doi.org/10.1177/1471082X1101200304>
- Cai, Z., & Xu, X. (2009). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 104, 371-383. <http://dx.doi.org/10.1198/016214508000000977>
- Chaudhuri, P. (1991). Global nonparametric estimation of conditional quantile functions and their derivative. *Journal of Multivariate Analysis*, 39, 246-269.
- Chaudhuri, Doksum, P. K., & Samarov, A. (1997). On average derivative quantile regression. *Annals of Statistics*, 25, 715-744. <http://dx.doi.org/10.1214/aos/1031833670>
- Cook, R. D., & Weisberg, S. (1991). Comment on "Sliced Inverse Regression for Dimension Reduction," by K.-C. Li. *Journal of the American Statistical Association*, 86, 328-332.
- Dette, H., & Scheder, R. (2011). Estimation of additive quantile regression. *Annals of the Institute of Statistical Mathematics*, 63, 245-265. <http://dx.doi.org/10.1007/s10463-009-0225-5>
- Fan, J., & Li, R. Z. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348-1360. <http://dx.doi.org/10.1198/016214501753382273>
- Gannoun, A., Girard, S., & Saracco, J. (2004). Sliced inverse regression in reference curves estimation. *Computational Statistics and Data Analysis*, 46, 103-122. [http://dx.doi.org/10.1016/S0167-9473\(03\)00141-5](http://dx.doi.org/10.1016/S0167-9473(03)00141-5)
- De Gooijer, J. G., & Zerom, D. (2003). On additive conditional quantiles with high-dimensional covariates. *Journal of the American Statistical Association*, 98, 135-146. <http://dx.doi.org/10.1198/016214503388619166>
- Härdle, W., Hall, P., & Ichimura, H. (1993). Optimal smoothing in single-index models. *Annals of Statistics*, 21, 157-178.
- Härdle, W., & Stoker, T. (1989). Investing smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84, 986-995.
- He, X., & Shi, P. (1996). Bivariate tensor-product B-splines in a partly linear model. *Journal of Multivariate Analysis*, 58, 162-181. <http://dx.doi.org/10.1006/jmva.1996.0045>
- He, X., Zhu, Z., & Fung, W. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, 89, 579-590. <http://dx.doi.org/10.1093/biomet/89.3.579>
- Hristache, M., Juditski, A., & Spokoiny, V. (2001). Direct estimation of the index coefficients in a single-index model. *Annals of Statistics*, 29, 595-623. <http://dx.doi.org/10.1214/aos/1009210682>
- Horowitz, J. L., & Lee S. (2005). Nonparametric Estimation of an Additive Quantile Regression Model. *Journal of the American Statistical Association*, 100, 1238-1249. <http://dx.doi.org/10.1198/016214505000000583>
- Hua, Y., Gramacy, R. B., & Lian, H. (2012). Bayesian quantile regression for single-index models. *Statistics and Computing*. <http://dx.doi.org/10.1007/s11222-012-9321-0>
- Ichimura, H. (1993). Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models. *Journal of Econometrics*, 58, 71-120.
- Jiang, R., Zhou, Z. G., Qian, W. M., & Shao, W. Q. (2012). Single-index composite quantile regression. *Journal*

- of the Korean Statistical Society*, 3, 323-332. <http://dx.doi.org/10.1016/j.jkss.2011.11.001>
- Kai, B., Li, R., & Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Annals of Statistics*, 39, 305-332. <http://dx.doi.org/10.1214/10-AOS842>
- Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91, 74-89. <http://dx.doi.org/10.1016/j.jmva.2004.05.006>
- Koenker, R. (2005). *Quantile Regression*. Cambridge, U.K.: Cambridge University Press.
- Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33-50.
- Koenker, R., Ng, P., & Portnoys, S. (1994). Quantile smoothing splines. *Biometrika*, 81, 673-680.
- Kong, E., & Xia, Y. (2007). Variable Selection for the Single-Index Model. *Biometrika*, 94, 217-229. <http://dx.doi.org/10.1093/biomet/asm008>
- Lee, S. (2003). Efficient semi parametric estimation of a partially linear quantile regression model. *Econometric Theory*, 19, 1-31.
- Li, B., & Wang, S. L. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association*, 102, 997-1008. <http://dx.doi.org/10.1198/016214507000000536>
- Li, K. C. (1991). Sliced Inverse Regression for Dimension Reduction (with discussion). *Journal of the American Statistical Association*, 86, 316-342.
- Li, L., & Yin, X. (2008). Sliced Inverse Regression with Regularizations. *Biometrics*, 64, 124-131. <http://dx.doi.org/10.1111/j.1541-0420.2007.00836.x>
- Li, Q., Xi, R., & Lin, N. (2010). Bayesian Regularized Quantile Regression. *Bayesian Analysis*, 5, 1-24. <http://dx.doi.org/10.1214/10-BA521>
- Li, Y., & Zhu, J. (2008).  $l_1$ -norm quantile regressions. *Journal of Computational and Graphical Statistics*, 17, 163-185. <http://dx.doi.org/10.1198/106186008X289155>
- Naik, P. A., & Tsai, C.-L. (2001). Single-Index Model Selections. *Biometrika*, 88, 821-832.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Ser. B*, 58, 267-288.
- Wang, J. L., Xue, L. G., Zhu, L. X., & Chong, Y. S. (2010). Estimation for a partial-linear single index model. *Annals of Statistics*, 38, 246-274. <http://dx.doi.org/10.1214/09-AOS712>
- Wang, Q., & Yin, X. (2008). A nonlinear multi-dimensional variable selection method for high dimensional data: sparse MAVE. *Computational Statistics and Data Analysis*, 52, 4512-4520. <http://dx.doi.org/10.1016/j.csda.2008.03.003>
- Wu, T. Z., Yu, K., & Yu, Y. (2010). Single index quantile regression. *Journal of Multivariate Analysis*, 101, 1607-1621. <http://dx.doi.org/10.1016/j.jmva.2010.02.003>
- Wu, Y., & Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 19, 801-817.
- Xia, Y., Tong, H., Li, W. K., & Zhu, L. X. (2002). An Adaptive Estimation of Dimension Reduction Space (with discussion). *Journal of the Royal Statistical Society, Ser. B*, 64, 363-410. <http://dx.doi.org/10.1111/1467-9868.03411>
- Yebin, C., Gooijer, J. G. D., & Zerom, D. (2011). Efficient estimation of an additive quantile regression model. *Scandinavian Journal of Statistics*, 38, 46-62. <http://dx.doi.org/10.1111/j.1467-9469.2010.00706.x>
- Yu, K., & Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, 93, 228-237.
- Yu, K., & Lu, Z. (2004). Local linear additive quantile regression. *Scandinavian Journal of Statistics*, 31, 333-346. [http://dx.doi.org/10.1111/j.1467-9469.2004.03\\_035.x](http://dx.doi.org/10.1111/j.1467-9469.2004.03_035.x)
- Yu, K., Lu, Z., & Stander, J. (2003). Quantile regression: Applications and current research areas. *The Statistician*, 52, 331-350. <http://dx.doi.org/10.1111/1467-9884.00363>

- Yuan, Y., & Yin, G. (2010). Bayesian quantile regression for longitudinal studies with non-ignorable missing data. *Biometrics*, *66*, 105-114. <http://dx.doi.org/10.1111/j.1541-0420.2009.01269.x>
- Zeng, P., He, T., & Zhu, Y. (2012). A Lasso-type approach for estimation and variable selection in single index models. *Journal of Computational and Graphical Statistics*, *21*, 92-109. <http://dx.doi.org/10.1198/jcgs.2011.09156>
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*, 1418-1429. <http://dx.doi.org/10.1198/016214506000000735>

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).